



**You have downloaded a document from  
RE-BUŚ  
repository of the University of Silesia in Katowice**

**Title:** Język informacyjno-wyszukiwawczy jako narzędzie organizacji informacji w dziedzinowych systemach hipertekstowych

**Author:** Marcin Roszkowski

**Citation style:** Roszkowski Marcin. (2009). Język informacyjno-wyszukiwawczy jako narzędzie organizacji informacji w dziedzinowych systemach hipertekstowych. Praca doktorska. Katowice : Uniwersytet Śląski

© Korzystanie z tego materiału jest możliwe zgodnie z właściwymi przepisami o dozwolonym użytku lub o innych wyjątkach przewidzianych w przepisach prawa, a korzystanie w szerszym zakresie wymaga uzyskania zgody uprawnionego.



UNIWERSYTET ŚLĄSKI  
W KATOWICACH



Biblioteka  
Uniwersytetu Śląskiego



Ministerstwo Nauki  
i Szkolnictwa Wyższego

UNIwersytet Śląski  
Wydział Filologiczny

Marcin Roszkowski

Język informacyjno-wyszukiwawczy  
jako narzędzie organizacji informacji  
w dziedzinowych systemach hipertekstowych

ROZPRAWA DOKTORSKA  
PRZYGOTOWANA POD KIERUNKIEM  
DR. HAB. WIESŁAWA BABIKA

Katowice 2009

## **Streszczenie**

W rozprawie przedstawiono studium analityczne z zakresu wykorzystania języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych. Dokonano charakterystyki modelu dziedzinowego systemu hipertekstowego oraz wskazano na jego cechy dystynktywne. W rozprawie przyjęto koncepcję organizacji informacji w systemach informacyjnych przetwarzających metainformacje, na którą składają się procesy reprezentacji informacji, identyfikacji cech wyszukiwawczych oraz organizacja punktów dostępu. W rozprawie wskazano udział języka informacyjno-wyszukiwawczego w każdym z tych etapów w odniesieniu do dziedzinowych systemów hipertekstowych. Grupę reprezentatywną stanowiło 30 systemów. Wyniki badań umożliwiły opracowanie trzech wariantów budowy narzędzi dostępu do zasobów omawianego typu systemów informacyjnych, w których wykorzystuje się język informacyjno-wyszukiwawczy.

## **Słowa kluczowe**

DZIEDZINOWE SYSTEMY HIPERTEKSTOWE, ORGANIZACJA INFORMACJI,  
ORGANIZACJA WIEDZY, SIECIOWE SYSTEMY ORGANIZACJI WIEDZY,  
KATALOGOWANIE ZASOBÓW WWW, JĘZYKI INFORMACYJNO-  
-WYSZUKIWAWCZE

<b>WSTĘP .....</b>	<b>7</b>
<b>CZĘŚĆ PIERWSZA .....</b>	<b>15</b>
<b>1. ORGANIZACJA INFORMACJI – WPROWADZENIE DO ZAGADNIENIA .....</b>	<b>15</b>
1.1. PRZEDMIOT .....	15
1.2. JEDNOSTKI.....	19
1.3. PROCESY .....	24
1.3.1. FORMATOWANIE JEDNOSTKI OPISU .....	24
1.3.2. KODOWANIE META INFORMACJI .....	27
1.3.3. ORGANIZACJA DOSTĘPU. PUNKTY DOSTĘPU .....	27
1.4. PODSUMOWANIE .....	35
<b>2. ORGANIZACJA ZASOBÓW WORLD WIDE WEB .....</b>	<b>36</b>
2.1. HIPERTEKSTOWY MODEL ORGANIZACJI INFORMACJI .....	37
2.1.1. GENEZA HIPERTEKSTU .....	38
2.1.2. STRUKTURA HIPERTEKSTU.....	40
2.2. ARCHITEKTURA INFORMACJI.....	44
2.3. METADANE .....	47
2.4. REPREZENTACJA META INFORMACJI.....	49
2.5. NARZĘDZIA DOSTĘPU .....	51
2.5.1. WYSZUKIWARKI INTERNETOWE.....	51
2.5.2. KATALOGI INTERNETOWE.....	55
2.5.3. BIBLIOTEKI CYFROWE, REPOZYTORIA.....	56
2.5.4. FOLKSONOMIE .....	59
2.6. TRYBY DOSTĘPU DO ZASOBÓW W INTERNETOWYCH SYSTEMACH INFORMACYJNYCH.....	63
2.7. ROLA JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO W ORGANIZACJI INFORMACJI W WORLD WIDE WEB.....	70
2.7.1. IDENTYFIKACJA PUNKTÓW DOSTĘPU .....	70
2.7.2. KODOWANIE .....	73
2.7.3. ORGANIZACJA PUNKTÓW DOSTĘPU .....	74
2.7.3.1. MECHANIZMY WYSZUKIWAWCZE.....	75
2.7.3.2. NARZĘDZIA DOSTĘPU PRZEDMIOTOWEGO.....	78
2.7.3.2.1. STRUKTURY RÓWNOWAŻNOŚCIOWE.....	78
2.7.3.2.2. STRUKTURY HIERARCHICZNE.....	81
2.7.4. SIECIOWE SYSTEMY ORGANIZACJI WIEDZY .....	93
2.8. PODSUMOWANIE .....	95
<b>3. CHARAKTERYSTYKA I BUDOWA DZIEDZINOWEGO SYSTEMU HIPERTEKSTOWEGO..</b>	<b>96</b>
3.1. DZIEDZINOWY SYSTEM HIPERTEKSTOWY - DEFINICJE.....	96
3.2. DZIEDZINOWY SYSTEM HIPERTEKSTOWY JAKO RODZAJ SYSTEMU INFORMACYJNO--WYSZUKIWAWCZEGO .....	103
3.3. STRUKTURA DZIEDZINOWEGO SYSTEMU HIPERTEKSTOWEGO.....	105
3.3.1. CEL I ZAKRES SYSTEMÓW .....	105
3.3.2. ROZWÓJ I ZARZĄDZANIE KOLEKCJĄ.....	107
3.3.3. CHARAKTERYSTYKA TREŚCIOWA I FORMALNA ZASOBÓW. SCHEMATY METADANYCH. ....	112
3.3.3.1. ROADS/IAFA.....	113
3.3.3.2. INTUTE METADATA ELEMENT SET.....	115
3.3.3.3. SCHEMAT SSG-FI .....	116
3.3.3.4. EDNA METADATA STANDARD.....	119
3.3.3.5. SCHEMAT MEATADNYCH RENARDUS.....	122
3.3.4. DOSTĘP PRZEDMIOTOWY .....	123
3.4. TYPOLOGIA DZIEDZINOWYCH SYSTEMÓW HIPERTEKSTOWYCH .....	124
3.5. MODEL ANALITYCZNY DZIEDZINOWEGO SYSTEMU HIPERTEKSTOWEGO.....	125
3.6. GENEZA DZIEDZINOWYCH SYSTEMÓW HIPERTEKSTOWYCH.....	128
3.7. POLSKIE DZIEDZINOWE SYSTEMY HIPERTEKSTOWE .....	132
3.8. PODSUMOWANIE .....	134
<b>CZĘŚĆ DRUGA .....</b>	<b>135</b>
<b>4. WYKORZYSTANIE JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO W ORGANIZACJI</b>	

**INFORMACJI W DZIEDZINOWYCH SYSTEMACH HIPERTEKSTOWYCH ..... 138**

4.1. MIEJSCE I ROLA JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO W DZIEDZINOWYCH SYSTEMACH HIPERTEKSTOWYCH.....	138
4.2. REPREZENTACJA INFORMACJI.....	139
4.2.1. NAZWY OSOBOWE I KORPORATYWNE .....	139
4.2.2. TYP ZASOBÓW .....	141
4.2.3. JĘZYK ZASOBÓW .....	143
4.2.4. DATY .....	143
4.2.5. NAZWY GEOGRAFICZNE.....	144
4.2.6. TREŚĆ. POLIREPREZENTACJA WIEDZY .....	144
4.2.7. DODATKOWE CECHY.....	156
4.3. IDENTYFIKACJA CECH WYSZUKIWAWCZYCH .....	158
4.4. ORGANIZACJA PUNKTÓW DOSTĘPU .....	160
4.4.1. WYSZUKIWARKI .....	161
4.4.2. NARZĘDZIA DOSTĘPU PRZEDMIOTOWEGO.....	170
4.5. PUNKTY DOSTĘPU W REZULTATACH WYSZUKIWANIA.....	174
4.5.1. TRYBY WYŚWIETLANIA REZULTATÓW WYSZUKIWANIA.....	175
4.5.2. PORZĄDKOWANIE REZULTATÓW WYSZUKIWANIA .....	181
4.6. PODSUMOWANIE .....	182

**5. JĘZYK INFORMACYJNO-WYSZUKIWAWCZY W NARZĘDZIACH DOSTĘPU PRZEDMIOTOWEGO WYBRANYCH DZIEDZINOWYCH SYSTEMÓW HIPERTEKSTOWYCH ..... 184**

5.1. BUBL INFORMATION SERVICE.....	185
5.2. CISMef.....	194
5.3. GEM – GATEWAY TO 21ST CENTURY SKILLS .....	200
5.4. INTUTE .....	206
5.4.1. INTUTE: NAUKA, INŻYNIERIA I TECHNOLOGIA (INTUTE: SCIENCE, ENGINEERING & TECHNOLOGY) .....	209
5.4.2. INTUTE: NAUKI SPOŁECZNE (INTUTE: SOCIAL SCIENCES) .....	215
5.4.3. INTUTE: ZDROWIE I NAUKI O ŻYCIU (INTUTE: HEALTH AND LIFE SCIENCES) .....	220
5.4.3.1. MEDYCYNĄ .....	222
5.4.3.2. PIELĘGNIARSTWO, POŁOŻNICTWO, OPIEKA ZDROWOTNA.....	223
5.4.3.3. HISTORIA NATURALNA.....	225
5.4.3.4. HISTORIA MEDYCYNY .....	227
5.5. MATHGUIDE .....	233
5.6. PODSUMOWANIE .....	238

**6. WYKORZYSTANIE JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO W BUDOWIE NARZĘDZI DOSTĘPU PRZEDMIOTOWEGO ..... 239**

6.1. WARIANT AHIERARCHICZNY .....	240
6.1.1. INDEKSY RZECZOWE .....	241
6.1.2. INDEKSY FORMALNE.....	248
6.1.3. INDEKSY MIESZANE .....	253
6.2. WARIANT HIERARCHICZNY.....	255
6.2.1. METODYKA EKSTRAKCJI.....	256
6.2.2. ANALIZA ILOŚCIOWA .....	258
6.2.3. ANALIZA JAKOŚCIOWA .....	264
6.2.3.1. PORZĄDEK KATEGORII W UKŁADZIE .....	265
6.2.3.2. PRZYNALEŻNOŚĆ KATEGORIALNA PUNKTÓW DOSTĘPU .....	267
6.2.3.3. BUDOWA STRUKTUR HIERARCHICZNYCH.....	269
6.2.4. FUNKCJE JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO.....	276
6.3. WARIANT HYBRYDOWY .....	280
6.4. PODSUMOWANIE .....	282

**ZAKOŃCZENIE ..... 284****BIBLIOGRAFIA ZAŁĄCZNIKOWA..... 292****WYKAZ TABEL ..... 301**

<b>SPIS ILUSTRACJI.....</b>	<b>302</b>
<b>INDEKS AUTORSKI.....</b>	<b>304</b>
<b>INDEKS RZECZOWY.....</b>	<b>306</b>
<b>ANEKSY .....</b>	<b>308</b>
ZAŁĄCZNIK 1. LISTA DZIEDZINOWYCH SYSTEMÓW HIPERTEKSTOWYCH W INTERNECIE .....	308
ZAŁĄCZNIK 2. GRUPA REPREZENTATYWNA DZIEDZINOWYCH SYSTEMÓW HIPERTEKSTOWYCH .....	311

## Wykaz użytych skrótów

Skrót	Rozwinięcie
AACR2	Anglo-American Cataloguing Rules - wydanie drugie
CERN	Conseil Européen pour la Recherche Nucléaire
ChWD	Charakterystyka wyszukiwawcza dokumentu
DSH	Dziedzinowe systemy hipertekstowe
DTD	Document Type Definition
FRBR	Functional Requirements for Bibliographic Records
HTML	HyperText Markup Language
IFLA	The International Federation of Library Associations and Institutions
ISBD	International Standard Bibliographic Description
JIW	Język informacyjno-wyszukiwawczy
KBK	Klasyfikacja Biblioteki Kongresu
KDD	Klasyfikacja Dziesiąta Deweya
LCSH	Library of Congress Subject Headings
MARC	Machine Readable Cataloging
MeSH	Medical Subject Headings
OPAC	Online Public Access Catalogue
SGML	Standard Generalized Markup Language
SIW	System informacyjno-wyszukiwawczy
SOW	System organizacji wiedzy
SSOW	Sieciowy system organizacji wiedzy
UKD	Uniwersalna Klasyfikacja Dziesiąta
URL	Uniform Resource Locator
WWW	World Wide Web
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language

## Wstęp

*„Czy to już koniec języków informacyjno-wyszukiwawczych?”\**

„Podstawową niedoskonałością Internetu jako źródła informacji, pomijając jego olbrzymi rozmiar, jest brak jakiegokolwiek kontroli jakości.” [Lancaster 2003, s. 137] Heterogeniczny i rozproszony charakter jego zasobów oraz permanentny i niekontrolowany przyrost informacji powoduje, że mamy do czynienia ze środowiskiem informacyjnym, gdzie z jednej strony obfituje ono w informacje nieaktualne i nieprawdziwe, a z drugiej stanowi nieocenione źródło wartościowych informacji. Drugi sposób postrzegania zasobów informacyjnych Internetu powoduje, że jest on utożsamiany z medium, które zrewolucjonizowało m.in. obieg komunikacji w nauce, przyspieszyło tempo rozwoju wielu dziedzin wiedzy, a także umożliwiło pogłębianie wiedzy i rozwój intelektualny szerokiej grupy jego użytkowników. Pomimo tej poznawczej i użytecznej funkcji zasobów Internetu, dominują tutaj jednak treści niespełniające wymogów wiarygodnych źródeł informacji. Jak zatem „odsiać ziarna od plew”, tak aby efektywnie wykorzystać wiedzę zapisaną w jego zasobach? Skoro każdy może zamieścić w Internecie komunikat dowolnej treści, który nie narusza przepisów prawa, bez ponoszenia odpowiedzialności za jego wiarygodność i zgodność ze stanem faktycznym, który opisuje to mamy do czynienia z sytuacją, gdzie nie ma żadnej kontroli jakości informacji na wejściu do zasobów informacyjnych Internetu. Tym samym „odsiewanie ziarna od plew” musi odbywać się na wyjściu z tego systemu, czyli na poziomie systemów i serwisów informacyjnych, których zbiory informacyjne są podzbiorami zasobu informacyjnego Internetu. W związku z tym, skierowanie uwagi na jakościowe kryteria selekcji i oceny zasobów Internetu włączanych do ich kolekcji ma tutaj priorytetowe znaczenie. Zdaniem Fredricka W. Lancastera filtrowanie jest obecnie jedną z najważniejszych funkcji sieciowych systemów informacyjnych, czyli systemów które przetwarzają zasoby Internetu. Przy czym selekcja informacji zachodzi tutaj na etapie ich streszczania i indeksowania. [Lancaster 20003] To intelektualne podejście do filtrowania zasobów Internetu, w przeciwieństwie do automatycznego, reprezentowanego przez wyszukiwarki internetowe, stanowi tutaj o wartości dodanej

---

\* Tytuł wystąpienia prof. dr hab. Bożenny Bojar podczas IX Forum Informatyki i Technicznej w Zakopanem, 25-28.IX.2007 r.



wyodrębnionej w taki sposób kolekcji. Tym „dodatkiem” jest tutaj gwarancja wysokiej jakości informacji udostępnianej użytkownikowi poprzez sieciowe systemy informacyjne. Tym samym pytanie F. Lancastera, „czy indeksowanie i streszczanie ma przyszłość?” [Lancaster 2003], jest pytaniem retorycznym, którego postawienie ma na celu uświadomienie, zarówno twórcom systemów informacyjnych jak i ich użytkownikom, znaczenia metainformacji dla efektywności procesów wyszukiwania informacji w Internecie.

Z drugiej strony, efektywne funkcjonowanie systemu informacyjno-wyszukiwawczego, obok środków technicznych, zależy przede wszystkim od organizacji informacji, jaką w nim zastosowano. Proces ten ma wymiar intelektualny, a jego efektem jest opracowanie narzędzi dostępu do kolekcji. [Rowley 2000] Zdaniem Elanie Svenonius [Svenonius 2000] organizacja informacji i organizacja innych elementów rzeczywistości nie powinny być porównywane. Jednak zorganizowanie dowolnego zbioru elementów, w tym również zbioru informacji wymaga przede wszystkim zidentyfikowania i scharakteryzowania tego, co chcemy poddać temu procesowi. W odniesieniu do organizacji informacji, identyfikacja polega na ustaleniu elementarnej jednostki poddawanej temu procesowi. Wiąże się to z ustaleniem tego, co ma zostać zorganizowane (np. dane, informacja, obiekt informacyjny czy wiedza?). W dokumentacyjnych systemach informacyjno-wyszukiwawczych wyznacza to format opisu danych oraz narzędzia reprezentacji informacji i wiedzy. Ten ostatni element to podstawowe zadanie dla języków informacyjno-wyszukiwawczych. W tym miejscu warto postawić pytanie: czy w organizacji informacji w środowisku sieciowym jest miejsce dla języków informacyjno-wyszukiwawczych? Te wywodzą się z okresu sprzed „wielkiej mechanizacji”, a w dobie dominacji pełnotekstowego wyszukiwania informacji za pomocą słów kluczowych interesujące może być wskazanie ich nowej roli w sieciowych systemach informacyjnych. Pojawiające się również w polskim środowisku informacji naukowej głosy o końcu języków informacyjno-wyszukiwawczych\*, dodatkowo skłaniają do zajęcia się kondycją tego rodzaju narzędzi.

Przedmiotem rozprawy jest organizacja informacji w dziedzinowych systemach hipertekstowych. Jest to klasa internetowych systemów informacyjno-wyszukiwawczych, których zadaniem jest udostępnianie użytkownikowi informacji o obiektach informacyjnych opublikowanych w Internecie z określonego zakresu

---

\* „Czy to już koniec języków informacyjno-wyszukiwawczych?” Tytuł wystąpienia prof. dr hab. Bożenny Bojar podczas IX Forum Informacji Naukowej i Technicznej w Zakopanem, 25-28.IX.2007 r.

przedmiotowego. Cechą wyróżniającą tego rodzaju systemy informacyjno-wyszukiwawcze, spośród innych systemów, jest udostępnianie użytkownikom informacji o wyselekcjonowanych i wartościowych zasobach informacyjnych World Wide Web. Na wyjściu systemu użytkownik otrzymuje metainformacje charakteryzujące relewantne obiekty informacyjne wraz z informacją o ich lokalizacji w Internecie. Ich dodatkową cechą wyróżniającą są narzędzia dostępu przedmiotowego realizujące dostęp do ich zasobów. Są one zbudowane z kategorii tematycznych, w których na podstawie przyjętych kryteriów grupuje się charakterystyki wyszukiwawcze obiektów informacyjnych.

Zakres rzeczowy rozprawy obejmuje zagadnienia organizacji informacji w hipertekstowych systemach informacyjnych funkcjonujących w Internecie, które spełniają warunki definicyjne dziedzinowego systemu hipertekstowego. Przyjęta w rozprawie interpretacja pojęcia *organizacji informacji* zakłada utożsamianie go z procesami, których celem jest stworzenie narzędzi strukturalnego dostępu do zasobów określonego zbioru informacji. W takim podejściu procesualnym zidentyfikowano etapy: formatowania jednostki opisu, kodowania metainformacji oraz organizowania dostępu.

Celem rozprawy jest pokazanie roli języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych. Tezą rozprawy jest założenie, że **język informacyjno-wyszukiwawczy jest wykorzystywany w procesach składających się na organizację informacji** we wspomnianej klasie systemów informacyjnych.

W celu udowodnienia postawionej w rozprawie tezy o istotnej roli języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych opracowano szczegółowe pytania badawcze:

- w jaki sposób język informacyjno-wyszukiwawczy jest wykorzystywany w procesach składających się na organizację informacji we wskazanym typie systemów informacyjno-wyszukiwawczych?
- jakie typy języka informacyjno-wyszukiwawczego są tam wykorzystywane?
- jakie konkretne języki informacyjno-wyszukiwawcze mają zastosowanie w tego rodzaju systemach?
- jakie zależności mają miejsce pomiędzy organizacją pola semantycznego zastosowanego w systemie języka informacyjno-wyszukiwawczego a narzędziami strukturalnego dostępu do jego zasobów?

Motywacją do podjęcia badań nad wykorzystaniem języka informacyjno-

-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych była odpowiedź na pytanie, czy w środowisku sieciowym zdominowanym przez mechanizmy wyszukiwawcze, automatycznie przetwarzające zapytania informacyjne użytkowników, jest jeszcze miejsce dla manualnego katalogowania zasobów WWW i czy przekłada się to na jakość wykorzystujących to narzędzi wyszukiwawczych. Wybór obiektu badań również nie był przypadkowy. Systemy te od końca lat 90-tych ubiegłego stulecia są dynamicznie rozwijane i wykorzystywane przez internautów jako wartościowe źródło informacji. To właśnie jakościowe kryteria selekcji, manualne katalogowanie zasobów, polityka zarządzania tego rodzaju systemami oraz zaprojektowane narzędzia dostępu stanowią o ich efektywności. Tematyka rozprawy sytuuje się w nurcie rozważań nad sieciowymi systemami organizacji wiedzy z uwzględnieniem kontekstu ich funkcjonowania w ramach procesów zachodzących w organizacji informacji.

W piśmiennictwie naukowym z tego zakresu nie podjęto dotychczas próby opracowania sposobów wykorzystania języka informacyjno-wyszukiwawczego w szerokim kontekście organizacji informacji w tego typu systemach. Wiele publikacji poświęcono natomiast dziedzinowemu systemowi hipertekstowemu, którego model ukonstytuował się jako efekt brytyjskiego programu badawczego *Access to Network Resources*. [Dempsey 2000] Wraz z uruchomieniem nowych systemów ukazują się publikacje wykorzystujące dorobek teoretyczny tego projektu, które uzupełnia się o nowe doświadczenia projektantów. Do najważniejszych prac badawczych nad tym typem systemów należy publikacja Traugotta Kocha [Koch 2000] poświęcona definicji dziedzinowego systemu hipertekstowego. Wyznaczyła ona kierunek dalszych badań oraz zapewniła bazę terminologiczną wykorzystywaną w kolejnych publikacjach. Została opublikowana w specjalnym zeszycie czasopisma *Online Information Review* (2000, vol. 24, nr 1), który w całości poświęcono temu zagadnieniu. Pomimo jej dużego znaczenia, punktem odniesienia w dokumentacjach systemów oraz publikacjach im poświęconych jest dokument elektroniczny w formie raportu z badań nad tą klasą systemów prowadzonych w ramach projektu *DESIRE - Development of a European Service for Information on Research and Education* pt. „Information Gateways Handbook”. [Belcher 1999] Jest to podręcznik, dla projektantów dziedzinowych systemów hipertekstowych zawierający szereg wskazówek dotyczących poprawnego zaprojektowania i zarządzania tego rodzaju systemami.

W polskim piśmiennictwie naukowym rzadko podejmowano próby opracowania tego zagadnienia. Dotyczyły one budowy tego rodzaju systemów [Derfert-Wolf 2004;

Roszkowski 2007a] lub charakterystyki nowopowstających polskich systemów. [Błaszczyk i in. 2001; Ganińska i in. 2006; Derfert-Wolf i in. 2007; Gmiterek i in. 2007; Pomianowicz 2008] Rzadko poruszano szczegółowe zagadnienia związane z funkcjonowaniem tych systemów, jak ocena jakości informacji włączanych do zbiorów informacyjnych systemów [Głowacka 2007], czy też zastosowania jako narzędzia wyszukiwania informacji dla określonych grup użytkowników. [Roszkowski 2007a]

W odniesieniu do wykorzystania języka informacyjnego w organizacji informacji w dziedzinowych systemach hipertekstowych prowadzone prace badawcze zmierzały w większości przypadków do wskazania przydatności tego narzędzia wyłącznie w katalogowaniu zasobów World Wide Web. [Koch, Day 1997; Slavic 2006] Szczegółowe projekty badawcze, których przedmiotem były narzędzia dostępu do zasobów systemu zmierzały w kierunku badań nad zachowaniami użytkowników [Koch, Golub, Ado 2006] lub opracowania mechanizmów jednoczesnego wyszukiwania informacji w kilku kolekcjach (ang. cross-searching, cross-browsing). [Day, Koch, Neuroth 2004] Wśród polskich publikacji należy odnotować artykuły dotyczące wykorzystania języka informacyjno-wyszukiwawczego w tego rodzaju systemach jako narzędzia reprezentacji cech treści i formy katalogowanych zasobów WWW. [Pepol 1999; Sosińska-Kalata 2002; Roszkowski 2005]

Rozprawa ma formę studium analitycznego poświęconego zastosowaniu języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych. Studium analityczne polega na dogłębnej analizie problemu poprzez jego dekonstrukcję na elementy składowe. Gromadzone w ten sposób dane poddawane są obiektywnej ocenie. Celem przeprowadzenia w rozprawie takiego studium było wskazanie sposobów zastosowania języka informacyjno-wyszukiwawczego w procesach składających się na organizację informacji we wskazanej klasie systemów. Forma studium analitycznego wyznaczyła tutaj podstawę metodologiczną rozprawy. Gromadzenie danych poddawanych procesom analitycznym przeprowadzono z wykorzystaniem metody analizy i krytyki piśmiennictwa. Źródłem informacji w tym zakresie były, obok publikacji naukowych, również udostępniane w Internecie dokumentacje techniczne systemów. Perspektywę badawczą w rozprawie wyznaczała struktura dziedzinowego systemu hipertekstowego. Przyjęte podejście „zorientowane na system” miało umożliwić określenie roli języka informacyjno-wyszukiwawczego wyłącznie w ramach struktury systemu, co świadomie wyeliminowało podjęcie rozważań nad otoczeniem systemu. Tym samym w rozprawie celowo zrezygnowano

z badań nad pragmatyką języka informacyjno-wyszukiwawczego w dziedzinowych systemach hipertekstowych. Przyjęta metoda systemowa, jako „pewnego rodzaju filozofia i pewien punkt widzenia badanego przedmiotu” [Ratajewski 1994, s. 113], ograniczyła zakres rozważań nad językiem informacyjno-wyszukiwawczym do elementów strukturalnych systemu informacyjno-wyszukiwawczego. W rozprawie wykorzystano również metodę statystyczną. Posłużono się nią podczas dekompozycji narzędzi dostępu do zasobów systemu, gdzie poszukiwano zależności ilościowych i jakościowych pomiędzy organizacją pola tematycznego systemów a organizacją pola semantycznego zastosowanych w nich językach informacyjno-wyszukiwawczych.

Podstawowym sposobem realizacji dostępu do zasobów dziedzinowych systemów hipertekstowych za pomocą narzędzi dostępu przedmiotowego jest technika przeglądania (ang. browsing). Narzędzia te mają postać zbioru kategorii tematycznych, które na podstawie przyjętych kryteriów grupują charakterystyki wyszukiwawcze obiektów informacyjnych. Tworzą one określone struktury osadzone w hipertekście umożliwiające ich eksplorację. W rozprawie wskazano sposoby wykorzystania elementów strukturalnych języka informacyjno-wyszukiwawczego do opracowania lub generowania narzędzi dostępu do zasobów systemów. Przyjęta w rozprawie koncepcja przedmiotowych punktów dostępu autorstwa B. Hjørlanda i L. Kyllsbech Nielsena [Hjørland, Kyllsbech Nielsen 2001] pozwoliła na przeprowadzenie rozważań przy zachowaniu spójnej podstawy terminologicznej narzędzi pozyskiwania informacji z systemu.

Rozprawa zbudowana jest ze wstępu, sześciu rozdziałów, zakończenia, bibliografii załącznikowej oraz aneksów. Rozdziały 1-3 tworzą część pierwszą rozprawy. Ma ona charakter teoretyczny. Jej celem jest charakterystyka procesów składających się na organizację informacji oraz zdefiniowanie przedmiotu badań, czyli dziedzinowego systemu hipertekstowego. Rozdziały 4-6 tworzą drugą część rozprawy. Jest ona wynikiem badań przeprowadzonych przez autora w celu weryfikacji postawionej w rozprawie hipotezy. Część druga jest poprzedzona wstępem, w którym przedstawiono szczegółowe cele badawcze oraz przyjęte metody i techniki ich realizacji. W obydwu częściach zastosowano jednolitą numerację rozdziałów.

Rozdział pierwszy zawiera wprowadzenie do zagadnienia organizacji informacji w systemach informacyjno-wyszukiwawczych. Przyjęta perspektywa badawcza zakłada koncentrację uwagi na systemach informacyjno-wyszukiwawczych operujących metainformacją. Organizację informacji zdefiniowano w kontekście trzech elementów wyznaczających odmienne punkty widzenia. Są to: przedmiot organizacji informacji,

jednostki oraz procesy zachodzące w ramach tego zagadnienia. Nakreślona w ten sposób perspektywa badawcza pojęcia *organizacji informacji* pozwoliła na jego interpretację w kategoriach celu, którym jest opracowanie narzędzi strukturalnego dostępu do zasobów określonej kolekcji. Realizacja tego celu wiąże się z identyfikacją punktów dostępu do zasobów systemu oraz ich włączeniem w narzędzia realizujące funkcje wyszukiwawcze.

Przedmiotem rozdziału drugiego jest organizacja informacji w zasobach World Wide Web. Scharakteryzowano tutaj narzędzia i techniki organizacji informacji w WWW oraz dokonano charakterystyki hipertekstowego modelu organizacji informacji, który zdaniem autora determinuje możliwości adaptacji tradycyjnych narzędzi organizacji informacji do środowiska sieciowego. Przytoczono również koncepcję architektury informacji, której przedmiotem jest strukturalizacja przestrzeni informacyjnej serwisów i systemów hipertekstowych. Szczególną uwagę poświęcono trybom dostępu do zasobów sieciowych systemów informacyjnych – wykorzystaniu wyszukiwarek internetowych oraz narzędzi dostępu przedmiotowego. W tej części rozprawy scharakteryzowano również rolę języka informacyjno-wyszukiwawczego w organizacji informacji w zasobach WWW w kontekście identyfikacji, reprezentacji i organizacji punktów dostępu.

Rozdział trzeci jest poświęcony charakterystyce dziedzinowego systemu hipertekstowego. Na podstawie przeglądu definicji tego rodzaju systemów, które funkcjonują w piśmiennictwie naukowym, przyjęto w pracy definicję najczęściej rozpowszechnioną, identyfikującą cechy dystynktywne tej klasy systemów informacyjno-wyszukiwawczych. Efektem prac prowadzonych przez autora w tym zakresie jest konstrukcja modelu analitycznego dziedzinowego systemu hipertekstowego.

W rozdziale czwartym scharakteryzowano wykorzystanie języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych. Określono rolę tego rodzaju narzędzi w kontekście reprezentacji informacji, identyfikacji cech wyszukiwawczych oraz organizacji punktów dostępu. Wskazano na sposoby wykorzystania poszczególnych typów i konkretnych języków informacyjno-wyszukiwawczych w omawianej klasie systemów informacyjnych. Celem tego rozdziału była analiza rzeczywistego wykorzystania języka informacyjno-wyszukiwawczego w organizacji informacji we wspomnianym typie systemów informacyjnych.

Rozdział piąty zawiera charakterystykę narzędzi dostępu przedmiotowego

w wybranych dziedzinowych systemach hipertekstowych. Dokonano tutaj dekompozycji ich struktur w celu odnalezienia zależności ilościowych i jakościowych pomiędzy różnymi formami grupowania kategorii tematycznych w narzędziach dostępu przedmiotowego. Badania przeprowadzone na tej grupie zmierzały do ustalenia sposobów wykorzystania języka informacyjno-wyszukiwawczego w budowie tego rodzaju narzędzi.

W ostatnim rozdziale przedstawiono wyniki badań własnych przeprowadzonych przez autora na grupie reprezentatywnej 30 dziedzinowych systemów hipertekstowych. Celem badań było opracowanie wariantów budowy narzędzi dostępu przedmiotowego z wykorzystaniem języka informacyjno-wyszukiwawczego. Wyniki przeprowadzonych badań pozwoliły na opracowanie trzech wariantów budowy tego rodzaju narzędzi: ahierarchicznego (płaskiego), hierarchicznego oraz hybrydowego. Każdy z nich poddano analizie zmierzającej do wykrycia zależności pomiędzy ich elementami strukturalnymi i zastosowanym do ich budowy językiem informacyjno-wyszukiwawczym. W centrum rozważań prowadzonych w tym rozdziale jest język informacyjno-wyszukiwawczy, a konkretnie jego system leksykalny oraz relacje paradygmatyczne, w takim zakresie, w jakim są one wykorzystane w określonym wariantcie narzędzi dostępu przedmiotowego.

Każdy z rozdziałów kończy się podsumowaniem, w którym zawarto wnioski z przeprowadzonych rozważań. Dołączone do pracy aneksy zawierają listę dziedzinowych systemów hipertekstowych zlokalizowanych przez autora w Internecie oraz grupę reprezentatywną systemów, które poddano badaniom. Bibliografia załącznikowa zawiera wykaz 219 pozycji cytowanych w pracy. Aparat pomocniczy w rozprawie składa się z wykazu stosowanych skrótów, indeksu nazwisk, indeksu rzeczowego oraz wykazu tabel i ilustracji.

# CZEŚĆ PIERWSZA

Celem pierwszej części rozprawy jest charakterystyka procesów składających się na organizację informacji oraz zdefiniowanie przedmiotu badań, czyli dziedzinowego systemu hipertekstowego. Informacje zawarte w tej części rozprawy mają za zadanie wprowadzenie do zagadnienia organizacji informacji zarówno w systemach tradycyjnych jak i sieciowych. Przedstawiona analiza narzędzi organizacji informacji w zasobach World Wide Web ma na celu umiejscowienie przedmiotu rozprawy w szerszym kontekście rozważań.

## 1. Organizacja informacji – wprowadzenie do zagadnienia

Przedmiotem pierwszego rozdziału rozprawy jest pojęcie *organizacji informacji*. Intencją autora w tym rozdziale jest zdefiniowanie podstawowych pojęć stosowanych w rozprawie oraz szczegółowa charakterystyka organizacji informacji jako procesu, którego celem jest zapewnienie optymalnej efektywności pozyskiwania informacji z systemu informacyjno-wyszukiwawczego.

### 1.1. Przedmiot

Zakres terminu *organizacja informacji* w najprostszy sposób można określić wyodrębniając tematy szczegółowe oraz analizując zakres publikacji naukowych, których przedmiotem jest to zagadnienie. Do najważniejszych publikacji wydanych w ostatnich kilkunastu latach należy m.in. książka autorstwa Elaine Svenonius pt. *Intelektualne podstawy organizacji informacji (The Intellectual foundation of information organization)*. [Svenonius 2000] Autorka analizując proces organizacji informacji rozpatruje go w kontekście:

- a) charakterystyki samego procesu organizowania,
- b) celów tworzenia opisu bibliograficznego,
- c) cech charakterystycznych jednostek bibliograficznych,
- d) języków opisu bibliograficznego,



- e) zasad sporządzania takiego opisu,
- f) języków opisu rzeczowego,
- g) metod i technik ekstrahowania metainformacji z dokumentu.

W wydanej rok wcześniej publikacji z tego zakresu pt. *Organizacja informacji (The Organization of Information)* [Taylor 1999], jej autorka przyjęła zbliżoną perspektywę badawczą. Arlene Taylor podjęła próbę charakterystyki pojęcia *organizacja informacji* uwzględniając następujące szczegółowe zagadnienia:

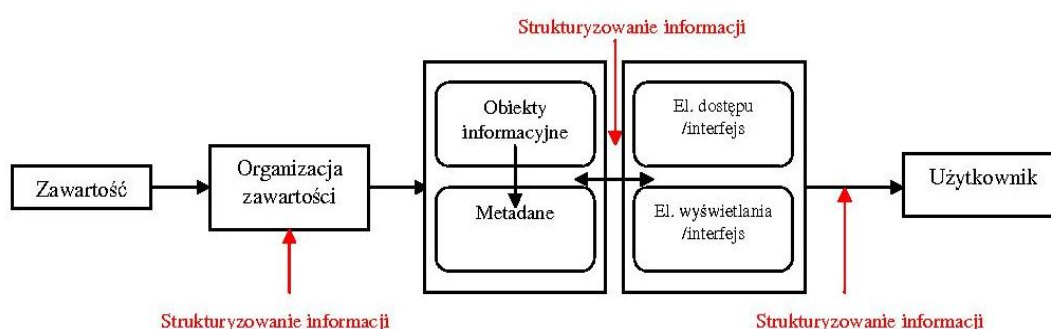
- a) organizacja w ludzkim działaniu,
- b) narzędzia wyszukiwawcze (indeksy, bibliografie, katalogi, bazy danych, itp.),
- c) rozwój organizacji utrwalonej informacji w dziejach,
- d) standardy kodowania (MARC, SGML-DTD, XML)
- e) metadane, formaty opisu danych oraz zasady katalogowania,
- f) werbalna analiza przedmiotowa,
- g) klasyfikacja,
- h) wizualizacja narzędzi organizacji informacji.

Przywoławszy listę szczegółowych zagadnień z dwóch publikacji istotnych z punktu widzenia badań nad organizacją informacji, nadal mamy do czynienia z nieostrym zakresem analizowanego pojęcia. Warto w tym miejscu uwzględnić również jedną z najnowszych publikacji książkowych, w której określa się dodatkowo miejsce organizacji informacji w środowisku cyfrowym. W książce pt. *Organizacja informacji: od półki do sieci* (ang. *Organizing information: from the shelve to the web*) [Chowdhury, Chowdhury 2006], autorzy koncentrują swoją uwagę na następujących zagadnieniach:

- a) ogólna charakterystyka organizacji informacji,
- b) organizacja informacji w środowisku pozabibliotecznym,
- c) katalogowanie,
- d) formaty opisu bibliograficznego,
- e) klasyfikacja biblioteczna,
- f) hasła przedmiotowe i tezaury w organizacji informacji,
- g) organizacja zasobów informacyjnych Internetu,
- h) metadane,
- i) znacznikowe języki opisu,
- j) ontologie internetowe,
- k) architektura informacji,
- l) Semantyczny Web.

Przedstawione szczegółowe tematy badań prowadzonych nad organizacją informacji pozwalają na wyodrębnienie trzech grup zagadnień. Pierwszą z nich stanowi **charakterystyka procesu organizowania informacji** oraz wskazanie na cel i funkcje, jakie pełni on w szerszym kontekście badań nad udostępnianiem informacji. W kolejnej grupie rozważań koncentruje się uwagę na **jednostce poddawanej procesowi organizacji**. Z punktu widzenia kompletności ujęcia zagadnienia oraz jego prawidłowego poznania, istotne jest wyodrębnienie organizowanych jednostek i zidentyfikowanie ich cech konstytutywnych. Trzecią grupę zagadnień tworzą badania prowadzone nad **procesami i narzędziami organizacji informacji**, gdzie analizuje się ich strukturę oraz efektywność.

Organizację informacji można umiejscowić w ramach procesów informacyjnych zachodzących podczas funkcjonowania systemu informacyjno-wyszukiwawczego i jego interakcji z użytkownikiem (Rys. 1.1)



Rys. 1.1. Proces organizacji informacji [Źródło: Rajashekar 2003]

Proces rozpoczyna się od pozyskania do kolekcji obiektów informacyjnych, które są nośnikami informacji i wiedzy (tzw. IKA – information and knowledge access [Rajashekar 2003]). Te, tworząc kolekcję poddawane są strukturyzowaniu. Na poziomie wewnętrznym następuje ukonstytuowanie elementarnej jednostki opisu obiektu informacyjnego, który na podstawie przyjętego formatu opisu danych odwzorowuje ich relewantne cechy. W ten sposób, na podstawie atrybutów obiektów informacyjnych tworzy się wykaz potencjalnych punktów dostępu do kolekcji. Te, z kolei na poziomie zewnętrznym zostają wkomponowane w narzędzia dostępu do zasobów systemu. Podczas interakcji użytkownika z systemem wykorzystanie odpowiednich narzędzi dostępu do zasobów powoduje aktywację określonych punktów dostępu i traktowanie ich jako cech wyszukiwawczych. Strukturyzacja jest kluczowym procesem w ramach

organizacji informacji, gdyż informacja jest wartościowa, ale wyłącznie w zakresie swojego ustrukturyzowania. Z powodu braku struktury, pewnego porządku w procesach tworzenia, dystrybucji i recepcji informacji, ta często nie dociera tam, gdzie jest pożądana i tym samym staje się bezużyteczna. [Rowley 2000]

Podstawowym celem organizacji informacji jest konstrukcja i udostępnienie użytkownikowi narzędzi strukturalnego dostępu do zasobów informacji. [Rowley 2000] Te, angażując procesy ekstrakcji metainformacji o obiektach informacyjnych, które włączone w narzędzia dostępu do zasobów kolekcji pozwalają na efektywne pozyskiwanie przez użytkownika informacji z systemu.

W 1968 roku Gerard Salton zaproponował wprowadzenie terminu *organizacja informacji* jako inną perspektywę badawczą w kontekście rozważań nad organizacją wiedzy. [Salton 1968; Hjørland 2007] U podstaw koncepcji organizacji wiedzy leżą również prace prowadzone przez Henry'ego Blissa w pierwszej połowie ubiegłego stulecia. [Bliss 1933] Celem organizacji wiedzy zdaniem H. Blissa miało być „rzeczowe uporządkowanie kolekcji: umożliwienie wglądu w zawartość wiedzy zapisanej w składających się na nią dokumentach”. [Sosińska-Kalata 2005a, s. 141] Zmiana zaproponowana przez G. Saltona była wynikiem odmiennej interpretacji jednostek, czyli tego, co było organizowane. Definicja terminu *organizacja wiedzy* autorstwa Jacka Andersona zamieszczona w „International Encyclopedia of Information and Library Science” brzmi:

*Charakterystyka (indeksowanie) i organizacja (klasyfikacja) tworzone są dla wyszukiwania komunikatów reprezentujących wiedzę, tekstów w których jest skodyfikowana oraz dokumentów zawierających zagnieżdżone teksty. Wiedza jako taka rezyduje w umysłach i mózgach istot żywych. Jej organizacja w celach wyszukiwawczych poprzez pamięć krótko- i długotrwałą jest kluczowym tematem nauk poznawczych. Informacja naukowa zajmuje się opisem i organizacją artefaktów (komunikatów, tekstów, dokumentów), w których wiedza jest reprezentowana i współdzielona. Te zasoby wiedzy nazywane są również zasobami informacji. Dlatego też termin organizacja wiedzy, w kontekście bibliotekoznawstwa i informacji naukowej, jest skróconą wersją terminu organizacja zasobów wiedzy, co nazywane jest również organizacją informacji”.* [Anderson 2003]

Mamy zatem do czynienia z dwoma rozbieżnymi poglądami. Pierwszym, w którym utożsamia się organizację wiedzy, organizację zasobów wiedzy oraz organizację informacji oraz drugim gdzie dostrzega się istotne różnice szczególnie na poziomie charakterystyki jednostek (informacja – wiedza). Przyjmując pogląd pierwszy można zastosować wspólną definicję dla tych procesów i przyjąć, że „organizacja wiedzy

w wąskim znaczeniu to systemy organizacji wiedzy [...], których przedmiotem zainteresowania są procesy organizowania wiedzy, jak klasyfikowanie, opis dokumentów, katalogowanie formalne, indeksowanie i analiza przedmiotowa”.\* [Broughton i in. 2005 s. 133] Przyjmując zaś drugą koncepcję należy bliżej przyjrzeć się typom jednostek poddawanych procesowi organizacji, aby móc wyraźnie zaznaczyć różnice, jakie występują pomiędzy tymi dwoma podejściami.

## **1.2. Jednostki**

Birger Hjørland [Hjørland 2007b] analizując przytoczoną definicję terminu *organizacja wiedzy* zidentyfikował siedem elementów, które mogą stanowić jednostki tego procesu. Są to:

- komunikat,
- wiedza,
- tekst,
- dokument,
- artefakt,
- informacja,
- zasoby wiedzy.

Do podstawowych procesów analizowanych w ramach informacji naukowej należy według B. Hjørlanda organizacja wiedzy i wyszukiwanie informacji. To sugeruje, że podstawowymi jednostkami poddawanych przetwarzaniu są wiedza i informacja. Jest to błędne przekonanie, gdyż analiza definicji terminu *organizacja wiedzy* pozwala na identyfikację większej liczby obiektów. B. Hjørland na podstawie analizy prac naukowych z zakresu terminologii informacji naukowej [m.in. Anderson 2003 i Garshol 2004] oraz naukoznawstwa uzupełnił listę o dodatkowe elementy, mogące pełnić funkcję jednostek organizacji wiedzy. Są to: artefakty, książki na półkach, pojęcia, dyscypliny naukowe, dokumenty, idee, informacja, zasoby informacyjne, wiedza, twierdzenia wiedzy, reprezentacje wiedzy, zasoby wiedzy, literatura, media, komunikaty reprezentujące wiedzę, trójwymiarowe obiekty, skodyfikowana wiedza, znaki, rozwiązania problemów, tematy, teksty, wypowiedzi, słowa, dzieła. [Hjørland

---

\* Knowledge Organization (KO) in the narrow sense is about knowledge organizing systems (KOS) such as bibliographical records, classification systems (e.g., DDC, LCC and UDC), thesauri, semantic networks and it is about knowledge organizing processes such as classification, document description, "descriptive cataloging", indexing and subject analysis. [Broughton i in. 2005 s. 133]

2007b] Mamy więc do czynienia z zamieszaniem wynikającym ze stosowania terminologii, która nie precyzuje charakteru organizowanych jednostek. Podobna sytuacja ma miejsce w przypadku terminu *wyszukiwanie informacji*. Efektem realizacji tego procesu może być nie tylko informacja, lecz np. metainformacja czy też dokument. Pytanie o różnice pomiędzy organizacją informacji i wiedzy sprowadza się m. in. do analizy procesu informacyjnego. Na potrzeby rozważań został przyjęty uproszczony model, który przedstawia Rys. 1.2.

Wiedza → pojęcia → znak → dane → komunikacja → informacja →  
komunikat (informacja 2) → znaczenie → wiedza

Rys. 1.2. Rodowód procesu informacyjnego [Źródło: Hjørland 2004]

Rys. 1.2. można odczytać następująco: „Osoba A **wie** coś, **konceptualizuje** to i koduje w postaci **znaków** (np. używając słów). Tworzy **dane** (np. ciąg słów) na podstawie znaków, które musi przełożyć na **sygnał**, aby przygotować się do procesu **komunikacji** i przekazać **informacje**. **Komunikat** jest odczytywany przez osobę B poprzez odczytanie jego **znaczenia** i tym samym powiększa zasób jej **wiedzy**”. [Hjørland 2004] W tak przedstawionym schemacie, każdy z jego elementów może stanowić jednostkę całego procesu organizacji.

W piśmiennictwie naukowym z tego zakresu, w zależności od przyjmowanych postaw o charakterze epistemologicznym, próbowano stosować różne sposoby interpretacji tego, co organizowano. Tendencja pozytywistyczna zakłada dążenie do optymalizacji odwzorowania wiedzy o rzeczywistości opisanej w dokumentach za pomocą obiektywnych metod i narzędzi. Tzw. perspektywa zorientowana na użytkownika, zwana również podejściem kognitywnym zmieniła pogląd na naturę organizowanych elementów. Utożsamiano je z terminami *idea*, *komunikat*, *struktura kognitywna użytkownika*. [Smiraglia 2002; Hjørland 2004] Jak zauważa B. Hjørland [Hjørland 2004], większość podmiian terminów stojących za członem *organizacja* nie niosła żadnych zmian teoretycznych leżących u podstaw tych procesów.

Inną perspektywą badawczą jest postrzeganie dokumentu jako rzeczywistej jednostki organizacji informacji. Zakłada ona funkcjonowanie narzędzi dostępu do kolekcji, na którą składają się dokumenty jako nośniki informacji. Tym samym dokument jest w centrum zainteresowania i przedmiotem ekstrakcji z niego metainformacji, które w odpowiedni sposób zorganizowane umożliwią użytkownikowi

dostęp do jego treści. Wraz z takim podejściem pojawił się problem zrewidowania znaczenia terminu *dokument* i określenia jego miejsca również w sieciowym środowisku informacyjnym (np. WWW). [Smiraglia 2001; Capurro, Hjørland 2003; Frohmann 2003] Dotychczasowa interpretacja dokumentu narzuca aspekt formy i fizycznego utrwalenia treści, co w przypadku sieciowego środowiska informacyjnego jest trudno przystawalne. Rewizja znaczenia terminu *dokument* wymaga określenia dodatkowych cech dystynktywnych. W organizacji informacji w zasobach cyfrowych porusza się problemy dotyczące identyfikacji elementarnej jednostki opisu. [Sosińska-Kalata 2005b] Pojawiający się w publikacjach naukowych [Dempsey, Russel 1999; Anderson, Perez-Carballo 2001; Anderson, Perez-Carballo 2005; Spink, Cole 2005] termin *obiekt informacyjny* (ang. *information object* oraz *information-bearing entity* [Miksa 1998; Taylor 1999]) wydaje się nie mieć konotacji związanych z nośnikiem jako fizycznym utrwaleniem. Termin ten może być utożsamiany z tradycyjnym dokumentem (np. w postaci książki, czy artykułu w czasopiśmie), ale przede wszystkim wykazuje zdolność do reprezentowania jednostek przekazu informacji w środowisku sieciowym. Podstawowym kryterium jego zastosowania powinna być spójność treściowa komunikatu. W przypadku środowiska sieciowego mamy do czynienia z informacją o charakterze heterogenicznym, tzn. niejednorodną pod względem przedmiotowym i formalnym (w postaci multimedialnego strumienia obiektów cyfrowych). Tym samym analiza kolekcji, która ma zostać zorganizowana wymaga identyfikacji typów obiektów składających się na nią i wypracowanie pewnego konsensusu, co do formy elementarnej jednostki opisu. Brak zbędnego obciążenia znaczeniowego terminu *obiekt informacyjny* przemawia za możliwością jego zastosowania jako jednostki w procesie organizacji informacji w środowisku sieciowym.

Na zakończenie rozważań nad jednostkami procesów organizacji informacji oraz wiedzy warto przytoczyć interesującą koncepcję B. Hjørlanda. Uważa on, że to co organizujemy, wyszukujemy i pozyskujemy to *twierdzenia wiedzy* (ang. *knowledge claims*). [Hjørland 2004] W jego domenowej koncepcji organizacji wiedzy za jednostkę poddawaną procesom przetwarzania przyjmuje się udokumentowane twierdzenia wiedzy (ang. *documented knowledge claims*). Organizacji poddaje się nie tyle fakty na temat rzeczywistości, ile dzieła (ang. *works*), które zawierają twierdzenia uzasadnione przez teorie wyrastające z różnych perspektyw epistemologicznych. Tym samym taki punkt widzenia zakłada brak obiektywności i neutralności procesu porządkowania wiedzy i jego ukierunkowanie na cel i dostosowanie do potrzeb określonej grupy

odbiorców. Koncepcja organizacji wiedzy B. Hjørlanda ma więc charakter pragmatyczny.

Przytoczone koncepcje jednostek organizacji informacji miały na celu wskazanie na różnorodność przyjmowanych perspektyw badawczych. Terminy *organizacja wiedzy* czy *organizacja informacji* można oczywiście używać w znaczeniu umownym, skróconym, jak zauważył J. Anderson w przytoczonej wypowiedzi. Naukowy charakter badań prowadzonych w tym zakresie obliguje do precyzji i jasnego zdefiniowania przedmiotu dociekań. Na potrzeby dalszych rozważań pojęcia: *organizacja informacji* i *organizacja wiedzy* będą rozróżniane m.in. ze względu na charakter jednostek podlegających procesowi organizacji. W odniesieniu do organizacji informacji, przedmiotem tego procesu jest obiekt informacyjny, który może być utożsamiany w określonych przypadkach z dokumentem. Efektem tego procesu jest identyfikacja punktów dostępu, które reprezentują określone kategorie cech obiektów. Atrybuty te mogą reprezentować formalne aspekty obiektów, np. autor, tytuł, typ obiektu, czas utworzenia i opublikowania, nazwa instytucji sprawczej, nazwa wydawcy, itp. Drugą, możliwą kategorię cech, stanowią atrybuty reprezentujące wymiar treści obiektów informacyjnych. Podczas odwzorowania cech formalnych i treściowych obiektów informacyjnych, każdemu atrybutowi przypisuje się określoną wartość. W procesie organizacji informacji wykorzystuje się narzędzia porządkowania wartości tych atrybutów, tak aby mogły stanowić punkty dostępu do kolekcji, które umożliwią użytkownikowi pozyskanie relewantnych obiektów informacyjnych.

Poprzez przedmiot organizacji wiedzy będzie rozumiana wiedza, czy też twierdzenia wiedzy, które są zawarte w treści obiektów informacyjnych. Tym samym treść obiektów informacyjnych odwzorowana i reprezentowana za pomocą odpowiednich narzędzi będzie przedmiotem zainteresowania organizacji wiedzy, która jest utożsamiana z systemami organizacji wiedzy (SOW). [Broughton i in. 2005] Zastosowanie systemów organizacji wiedzy do organizowania dostępu do określonej kolekcji wiąże się z utworzeniem mapy konceptualnej, reprezentującej fragment rzeczywistości opisany w obiektach informacyjnych. Narzędzia organizacji wiedzy muszą więc dysponować określoną siłą semantyczną do odwzorowywania pojęć składających się na treść obiektów oraz relacji, jakie zachodzą między nimi. Elementy strukturalne takich narzędzi determinują stopień odwzorowania analizowanej rzeczywistości. B. Sosińska-Kalata zauważa: „jest oczywiste, że nie istnieje system organizacji wiedzy, który akceptowaliby wszyscy. Choć opracowanie jednego, powszechnie przyjętego systemu, byłoby może ze względów praktycznych korzystne, to

jest mało prawdopodobne, aby kiedykolwiek taki system powstał. Systemy organizacji wiedzy, jak wszystkie wytwory ludzkiego intelektu, są uwarunkowane kulturą, z której i dla potrzeb której wyrosły. Co jest istotne dla jednej kultury niekoniecznie musi być istotne dla innej. Stąd żyjemy w świecie wielości i różnorodności, i dotyczy to również sposobów organizacji wiedzy. Mimo tej różnorodności dla wszystkich systemów organizacji wiedzy wspólne jest to, że:

- każdy jest emanacją pewnego określonego sposobu patrzenia na rzeczywistość, który implikuje sposób patrzenia na organizowany zasób informacji i jego elementy;
- każdy jest uporządkowanym zbiorem wyrażen reprezentującym pojęcia, które składają się na tę projekcję rzeczywistości;
- za pomocą różnych SOW ten sam obiekt (tekst, obraz, zapis dźwiękowy lub filmowy) może być w różny sposób scharakteryzowany, a zatem różne SOW mogą się wzajemnie uzupełniać;
- swobodne i poprawne posługiwanie się każdym SOW przez człowieka posiadającego pewną wiedzę w dziedzinie poszukiwań wymaga zachowania wystarczającej jedności między pojęciem wyrażanym za pomocą SOW i obiektem, do którego pojęcie to się odnosi; osoba szukająca określonych materiałów używając SOW musi być w stanie powiązać zawartą w nim projekcję pojęć z własnym wyobrażeniem świata”. [Sosińska-Kalata 2005a, s. 147]

Przyjęta w pracy perspektywa zakłada interpretację terminu *organizacja informacji* jako procesu strukturyzacji kolekcji, którego efektem jest identyfikacja relewantnych atrybutów obiektów informacyjnych. Te, włączone w odpowiednie narzędzia pełnią funkcję punktów dostępu do zasobu poprzez traktowanie ich jako cech wyszukiwawczych. Organizacja wiedzy, utożsamiana z systemami organizacji wiedzy, operuje wiedzą publiczną skodyfikowaną w obiektach informacyjnych i reprezentowaną przez system pojęciowy zorganizowany w strukturze i notacji określonych języków informacyjno-wyszukiwawczych. [zob. Sosińska-Kalata 1999] Tym samym zakres terminu *organizacja informacji* obejmuje również w określonym zakresie systemy organizacji wiedzy jako narzędzia odwzorowania i porządkowania pojęć zawartych w treści obiektów informacyjnych. Efektem procesu organizacji informacji jest konstrukcja narzędzi strukturalnego dostępu do kolekcji. Systemy organizacji wiedzy pełnią również funkcję narzędzi dostępu poprzez ich włączenie w proces organizowania informacji w danej kolekcji. Taka koncepcja, charakteryzuje się dużym uproszczeniem,



jednak jej funkcja poznawcza i w pewnym stopniu porządkująca wiedzę na ten temat jest konieczna do przeprowadzenia dalszej części wywodu na temat wykorzystania języka informacyjnego jako narzędzia organizacji informacji.

### **1.3. Procesy**

W organizacji informacji w dokumentacyjnych systemach informacyjno-wyszukiwawczych można zidentyfikować trzy procesy o kluczowym znaczeniu. Są to:

- formatowanie i strukturyzowanie jednostki opisu, równoznaczne z identyfikacją formatu opisu,
- kodowanie metainformacji za pomocą formatu zapisu danych,
- organizowanie dostępu do zasobów z wykorzystaniem przeznaczonych do tego celu narzędzi.

Każdy z nich zachodzi z wykorzystaniem odpowiednich narzędzi przeznaczonych do ich realizacji oraz mniej lub bardziej sformalizowanych procedur zapewniających spójność całego procesu.

#### **1.3.1. Formatowanie jednostki opisu**

Celem katalogowania, czyli procesu tworzenia opisu katalogowego jest opracowanie surogatu dokumentu. Polega to na identyfikacji relewantnych cech dokumentu i przypisaniu im określonych wartości, które mogą stanowić cechy wyszukiwawcze i umożliwić odszukanie go w zbiorze wyszukiwawczym systemu. Reprezentacja tego rodzaju informacji następuje w oparciu o przyjęty format opisu bibliograficznego oraz formaty zapisu poszczególnych kategorii metainformacji, a także procedury postępowania podczas katalogowania. Efektem katalogowania dokumentu jest utworzenie rekordu bazy danych, przechowywanego przez system, który reprezentuje jego istotne cechy stanowiąc jego potencjał wyszukiwawczy. Taki opis pełni również funkcję informacyjną. Podczas analizy rezultatów wyszukiwania pozwala użytkownikowi na podjęcie decyzji o jego wykorzystaniu lub porzuceniu.

Format opisu danych to zestaw atrybutów dokumentów, którym przypisywana jest odpowiednia wartość podczas procesu katalogowania. Formaty opisu danych w dokumentacyjnych systemach informacyjno-wyszukiwawczych zbudowane są zazwyczaj z dwóch części formalnej i rzeczowej, gdzie odwzorowuje się odpowiednio atrybuty obiektu informacyjnego związanego z jego formą oraz treścią.

Zalecenie IFLA o nazwie *Model funkcjonalnych wymagań dla rekordów bibliograficznych – FRBR* (ang. Functional Requirements for Bibliographic Records) zawiera model struktury rekordu bibliograficznego na wysokim poziomie abstrakcji. Jest to model funkcjonalny dla rekordu bibliograficznego, którego charakterystyczną cechą jest ustalenie większej niezależności pomiędzy dziełem jako wytworem intelektualnym a jego fizycznym utrwaleniem. [Pacek 2005] Tworzą go trzy grupy obiektów, które konstytuują model, są to : [Nahotko 2001, s. 14]

1. Produkty działań intelektualnych lub artystycznych:
  - a. Utwór (ang. work) – odrębny twór intelektualny lub artystyczny. Utwór jest abstrakcyjnym tworem; nie istnieje określony materialny obiekt, na który można by wskazać jako na utwór [...];
  - b. Uzewnętrznienie (ang. expression) – intelektualne lub artystyczne wyrażenie utworu „w formie alfanumerycznej, muzycznej, choreograficznej, dźwięk, obraz, obiekt, ruch itp., lub dowolna ich kombinacja. Uzewnętrznienie jest określoną formą intelektualną lub artystyczną którą utwór przyjmuje za każdym razem, gdy jest „realizowany”. Uzewnętrznienie obejmuje np. określone słowa, zdania, akapity, będące rezultatem realizacji utworu w formie tekstu”;
  - c. Materializacja (ang. manifestation) - fizyczna reprezentacja utworu lub jego uzewnętrznienia [...];
  - d. Pozycje (ang. items) - pojedynczy egzemplarz materializacji. Jest to konkretna jednostka fizyczna dzieła. W wielu przypadkach jest to pojedynczy obiekt fizyczny (np. egzemplarz wydawnictwa jednotomowego, pojedyncza kasetka z nagraniem itp.).
2. Jednostki odpowiedzialne za intelektualną lub artystyczną zawartość tych produktów:
  - a. Osoby – twórcy zmarli lub żyjący;
  - b. Ciała zbiorowe (korporatywne) – organizacja lub grupa osób i/lub organizacji działające jako jedność.
3. Jednostki tworzące przedmiot działań intelektualnych lub artystycznych:
  - a. Pojęcia – pojęcie abstrakcyjne lub idea, która może być przedmiotem utworu: dziedzina wiedzy, dyscyplina, szkoły naukowe, teorie, procesy, techniki itp.;

- b. Obiekty – rzeczy materialne, które mogą być przedmiotem utworu: ruchome lub nieruchome obiekty występujące w naturze; stałe i ruchome obiekty będące wytworem ludzkim; obiekty już nieistniejące;
- c. Wydarzenia – działanie lub zdarzenie mogące być przedmiotem utworu: wydarzenia historyczne lub fikcyjne niezależnie od czasu, kiedy się rozgrywają;
- d. Miejsca – lokalizacja: ziemską i pozaziemską, historyczną i współczesną, miejsca geograficzne i jurysdykcja geopolityczną.

Wysoki poziom abstrakcji modelu FRBR stanowi podstawę teoretyczną do opracowywania formatów opisów dla różnego rodzaju dokumentów. Opracowanie przez IFLA logicznych ram dla spójnego modelu opisu danych bibliograficznych było warunkiem koniecznym do prowadzenia dalszych prac nad formatem opisu dokumentów tradycyjnych oraz elektronicznych.

W latach siedemdziesiątych ubiegłego stulecia w ramach prac IFLA opracowano standard opisu danych znany jako *Międzynarodowy Standard Opisu Bibliograficznego* (ang. The International Standard Bibliographic Description - ISBD). Format ten wyznacza osiem stref interpretacji obiektu informacyjnego. Są to: [Taylor 1999, s. 80]

1. Tytuł oraz oznaczenie odpowiedzialności.
2. Wydanie.
3. Typ publikacji/dane dotyczące cech fizycznego utrwalenia.
4. Dane dotyczące wydawcy oraz odpowiedzialności za dystrybucję.
5. Opis fizyczny.
6. Seria.
7. Uwagi.
8. Standaryzowany identyfikator oraz warunki dostępności.

Format ISBD określa również strukturę zapisu danych oraz odpowiednie znaki interpunkcyjne odpowiedzialne za rozpoczynanie, zakańczanie oraz separację poszczególnych stref. Standard ISBD był podstawą do opracowywania wielu narodowych formatów opisu, m.in. drugiego wydania *Anglo-American Cataloguing Rules* [Gorman, Winkler 1978], a także polskiej normy PN-82/N-01152.00 - *Opis bibliograficzny – Postanowienia ogólne*. Dalsze prace nad standardem ISBD doprowadziły do opracowania zaleceń dotyczących formatów opisu dla różnego rodzaju dokumentów, w tym elektronicznych. W 1997 roku opublikowano standard ISBD(ER), czyli zalecenia IFLA dotyczące opisu dokumentów elektronicznych.

Obok formatów opisu danych dla tradycyjnych oraz elektronicznych

dokumentów piśmienniczych funkcjonują również standaryzowane zalecenia, których przedmiotem jest reprezentacja relewantnych cech innych rodzajów (w kategoriach formy, treści lub przeznaczenia) dokumentów. Np. dla dokumentów archiwalnych funkcjonują m.in. AAMP (Archives, Personal Papers and Manuscripts), ISAD(G) (General International Standard Archival Description).

### **1.3.2. Kodowanie metainformacji**

Format opisu danych stanowi wykaz cech obiektu informacyjnego, którym nadawana jest określona wartość podczas procesu katalogowania. Tego rodzaju dane muszą zostać utrwalone w postaci rekordu bazy danych o określonej strukturze. Ta wyznaczona jest przez przyjęty format zapisu danych. Najbardziej rozpowszechnionym sposobem kodowania danych bibliograficznych jest format MARC (Machine Readable Cataloging). Strukturę tego formatu oraz jego pochodnych (np. USMARC, UNIMARC) wyznacza zbiór pól i podpól oraz ich identyfikatorów, które reprezentują określone atrybuty obiektu informacyjnego. Celem formatów opracowanych na podstawie MARC jest umożliwienie automatycznej wymiany i przetwarzania danych bibliograficznych przez zautomatyzowane systemy informacyjne. Format zapisu danych wyznacza jednocześnie logiczną strukturę ich przechowywania. Format kodowania informacji określa formalne zasady tworzenia opisu, a także wskazuje na źródła z których można przejmować wartości dla atrybutów. Są to np. normy dotyczące kodowania nazw języka dokumentu, daty jego publikacji, a także kartoteki wzorcowe oraz słowniki JIW.

### **1.3.3. Organizacja dostępu. Punkty dostępu**

Funkcją narzędzi organizacji informacji jest umożliwienie efektywnego dostępu do zasobów organizowanej kolekcji. Każdy obiekt informacyjny tworzący kolekcję poddawany jest procesowi odwzorowania jego cech formalnych, rzeczowych oraz kodowania tego rodzaju metainformacji. Wybrany atrybutom obiektów zostaje nadany status punktu dostępu. Punkty te zostają włączone w większe struktury realizujące funkcje wyszukiwawcze. Koncepcja punktów dostępu pełni kluczową rolę w organizacji informacji. Wywodzi się z organizowania dostępu do tradycyjnych zasobów w postaci np. zbioru dokumentów w bibliotekach i tym samym związana jest z metodyką tworzenia katalogów bibliotecznych, czy też indeksów alfabetycznych, systematycznych, przedmiotowych oraz innego rodzaju narzędzi umożliwiających dostęp do kolekcji. W terminologii bibliotekoznawstwa i informacji naukowej punkty

dostępu utożsamiane są z hasłami opisu bibliograficznego [Rowley, Farrow 2003, s. 245], które stanowią kryterium porządkowania i wyszukiwania opisów bibliograficznych w zbiorach informacyjnych. Hasłem opisu bibliograficznego może być:

- hasło osobowe, którego jedynym lub pierwszym elementem jest nazwa osobowa,
- hasło korporatywne, czyli takie którego jedynym lub pierwszym elementem jest nazwa ciała zbiorowego (instytucji, organizacji) lub imprezy,
- hasło tytułowe, czyli takie którego jedynym lub pierwszym elementem jest tytuł ujednolicony [Słownik encyklopedyczny 2002, s. 76]

Funkcjonujące normy w tym zakresie określają zasady tworzenia poszczególnych typów haseł opisu bibliograficznego. Normy polskie to:

- PN-N-01228:1994 - Hasło opisu bibliograficznego – Forma nazw geograficznych.
- PN-N-01229:2002 - Hasło opisu bibliograficznego – Hasło osobowe.
- PN-N-01230:2001 - Hasło opisu bibliograficznego – Hasło korporatywne.
- PN-N-01231:2001 - Hasło opisu bibliograficznego – Hasło tytułowe.

Dodatkowym rodzajem punktów dostępu są punkty dostępu przedmiotowego, które za pośrednictwem określonych narzędzi umożliwiają dostęp do obiektów informacyjnych uwzględniając cechy ich treści.

Punkty dostępu pełnią funkcję organizującą jako elementy struktury narzędzia wyszukiwawczego oraz wyszukiwawczą reprezentując określone cechy dokumentów i odsyłając do ich opisów lub lokalizacji w zbiorze informacyjnym. Potencjał informacyjny kolekcji może zostać efektywnie wykorzystany pod warunkiem identyfikacji relewantnych punktów dostępu do tego zbioru.

Reprezentacja informacji o obiektach tworzących zbiór informacyjny warunkuje możliwość identyfikacji punktów dostępu. Proces odwzorowania cech treściowych i wybranych cech formalnych obiektów informacyjnych jest realizowany poprzez języki informacyjno-wyszukiwawcze. Są one „sztucznymi systemami semiotycznymi tworzonymi w celu zapewnienia realizacji określonego typu procesów informacyjnych. Są to wyspecjalizowane, sztuczne systemy językowe przeznaczone do opisu treści i podstawowych cech formalnych dokumentów”. [Babik 2006 s. 191] W dalszej części rozprawy termin *język informacyjno-wyszukiwawczy* będzie rozumiany jako „język sztuczny, którego wyspecjalizowaną funkcją jest funkcja metainformacyjna, polegająca na odwzorowaniu cech treści i/lub cech formalnych dokumentu poprzez

tworzenie ich reprezentacji zwanej charakterystyką wyszukiwawczą dokumentu oraz funkcja wyszukiwawcza, polegająca na sterowaniu procesem selekcjonowania informacji przez użycie wyrażen charakteryzujących warunki, jakie powinien spełniać wyszukany podzbiór informacji”. [Ścibor 1998, s. 25]

Funkcje języka informacyjno-wyszukiwawczego wyznaczone są poprzez jego podstawowe role pełnione w kontekście funkcjonowania w systemie informacyjno-wyszukiwawczym. „Wiążą się one z dwiema podstawowymi kategoriami działań wykonywanych za pomocą JIW w dokumentacyjnych systemach informacyjno-wyszukiwawczych: indeksowaniem dokumentów i określaniem strategii ich wyszukiwania. Stąd wyodrębnia się dwie podstawowe funkcje JIW: funkcję przedstawiania treści i formy dokumentów (funkcja metainformacyjna) oraz funkcję wyszukiwawczą”. [Babik 2006, s. 195]

Dodatkową, uznawaną za wtórną jest funkcja organizacyjna JIW. Jest pochodną wobec funkcji opisowej, metainformacyjnej, ale „ma istotne znaczenie dla sprawnej realizacji funkcji impresywnej tego języka, czyli wyszukiwania dokumentów czy informacji o dokumentach w zbiorze informacyjnym”. [Sosińska 1987, s. 46] Porównuje się ją z funkcją klasyfikacyjną języka naturalnego, która „jest funkcją systemu językowego, a nie jego tekstów”. [Bojar 2005, s. 75] Każdy język pełniący funkcję semantyczną dokonuje klasyfikacji reprezentowanego przez niego fragmentu rzeczywistości. Systemowy charakter funkcji organizacyjnej JIW przejawia się:

- determinacją miejsca informacji opisanej w zbiorze przez zewnętrzną postać wyrażen prostych współtworzących charakterystykę tej informacji (ChWD),
- zależnością pomiędzy organizacją zbioru informacyjnego, a organizacją JIW na osi syntagmatycznej określaną mianem zależności między pionowym i poziomym porządkiem przyjętym w danym języku. [Sosińska 1987, s. 46-47]

Plan wyrażania danego JIW ma istotny wpływ na jego funkcję organizacyjną. Polega to na ustaleniu wartości porządkującej dla jego wyrażen. Zachodzi to zarówno na poziomie elementarnych znaków graficznych, jednostek leksykalnych oraz zdań JIW. „Ustalenie tej wartości dla jednostek leksykalnych uzależnione jest od ich miejsca w paradygmatycznej strukturze języka oraz od roli, jaką pełnią one w organizacji syntagmatycznej”. [Sosińska 1987, s. 47]

Eugeniusz Ścibor podsumowując swoje rozważania nad denotacją i konotacją wyrażen JIW konstatuje:

- języki informacyjne opisują dokumenty, a nie bezpośrednio obiekty opisane w dokumentach;

- desygnatami nazw będących jednostkami leksykalnymi języka informacyjnego są nie same obiekty, o których można orzec te nazwy, lecz dokumenty traktujące o tych obiektach;
- jednostki języka informacyjnego są nazwami klas dokumentów;
- zakresami jednostek leksykalnych języka informacyjnego jako nazw klas dokumentów są zbiory desygnatów w postaci dokumentów scharakteryzowanych za pomocą danych jednostek;
- języki informacyjne nie tylko opisują dokumenty, ale i odsyłają do nich. Funkcja odsyłania jest pełniona pośrednio poprzez słowniki danych języków informacyjnych dzięki umieszczeniu w nich jednostek leksykalnych, które wystąpiły lub mogą w przyszłości wystąpić w charakterystykach wyszukiwawczych dokumentów. Funkcja bezpośredniego odsyłania jest natomiast realizowana przez zastosowanie języka informacyjnego do porządkowania i charakteryzowania dokumentów pochodnych lub pierwotnych. [Ścibor 1998, s. 32]

Szczególne znaczenie ma ostatnie stwierdzenie, w którym identyfikuje się funkcje odsyłania pośredniego i bezpośredniego, która utożsamia jednostki leksykalne JIW z punktami dostępu do zbioru informacyjnego.

Funkcjonująca w informacji naukowej najczęściej cytowana typologia dzieli tego rodzaju języki na:

- języki klasyfikacyjne (klasyfikacje piśmiennictwa),
- języki haseł przedmiotowych,
- języki słów kluczowych,
- języki deskryptorowe,
- języki kodów semantycznych i języki syntagmatyczne,
- języki opisu formalnego (bibliograficznego, katalogowego, cytowań bibliograficznych) [Babik 2006 s. 205-206]

Z punktu widzenia wykorzystania języka informacyjno-wyszukiwawczego jako narzędzia organizacji informacji ważne jest podkreślenie wielowymiarowości tego narzędzia w kontekście efektywności całego procesu. Polega ona, z jednej strony na zapewnianiu reprezentacji wymiaru treści dokumentów w postaci spójnych komunikatów wyrażonych za pomocą środków leksykalnych danego JIW. Z drugiej zaś, na wykorzystaniu struktury paradygmatycznej oferowanej przez JIW do realizacji funkcji heurystycznej i organizacyjnej. Nie wszystkie typy JIW dysponują równą siłą semantyczną i wyszukiwawczą. Tym samym ich wpływ na cały proces organizacji

informacji jest zróżnicowany. Ich efektywność tkwi w możliwościach wykorzystania systemu leksykalnego oraz systemu organizacji pola semantycznego do generowania lub tworzenia narzędzi opartych na koncepcji punktów dostępu. Funkcja wyszukiwawcza i organizacyjna JIW realizowana jest poprzez włączenie ich w narzędzia wyszukiwawcze w postaci np. alfabetycznego wykazu jednostek leksykalnych (struktury równoważnościowe) czy też mono- lub polihierarchicznych struktur porządkujących punkty dostępu, gdzie każde użyte wyrażenie odsyła pośrednio (do ChWD) lub bezpośrednio do relewantnych dokumentów/obiektów informacyjnych.

Koncepcja punktów dostępu leżąca u podstaw narzędzi organizacji informacji przyjmuje formę zależną od technicznych środków realizacji procesów wyszukiwania informacji. W tradycyjnych katalogach bibliotecznych polega to na zastosowaniu odpowiedniego typu katalogu posługującego się konkretnym rodzajem punktu dostępu. Alfabetyczne katalogi autorskie i tytułowe wykorzystują jednocześnie te dwa atrybuty jako punkty dostępu, natomiast katalogi rzeczowe traktują jako punkt dostępu reprezentację treści dokumentu, wyrażoną w systemie leksykalnym wybranego JIW. Przeniesienie tego rodzaju narzędzi do technik komputerowego przetwarzania informacji i do katalogów OPAC niewiele zmienia w samej koncepcji punktów dostępu. Karty katalogowe w tradycyjnych bibliotekach oraz katalogi online wykorzystują ten sam rodzaj danych o dokumentach. [Borgman 1996, s. 495] Zmieniają się jednak techniczne środki realizacji procesów wyszukiwawczych i możliwości zarówno wizualizacji narzędzi jak i technik wyszukiwania. Charles A. Cutter, jeden z pionierów współczesnego bibliotekoznawstwa, na początku dwudziestego wieku określił funkcje tradycyjnych katalogów bibliotecznych. Zaliczył do nich: [Borgman 1996, s. 495]

1. Umożliwienie użytkownikowi odnalezienia książki
  - a) autorstwa danej osoby,
  - b) o określonym tytule,
  - c) na określony temat.
2. Wskazanie co posiada biblioteka
  - a) autorstwa danej osoby,
  - b) na określony temat,
  - c) w określonym rodzaju literatury.
3. Asystowanie w wyborze przez użytkownika dokumentów
  - a) pod względem ich cech formalnych,
  - b) pod względem rodzaju i tematu.

Realizacja tych funkcji jest możliwa poprzez identyfikację relewantnych punktów



dostępu umożliwiającą pozyskanie z kolekcji informacji o jej elementach według wskazanych kryteriów. W odniesieniu do elektronicznych narzędzi wyszukiwawczych, realizacja koncepcji punktów dostępu musi uwzględniać również stosowane techniki wyszukiwania i indeksowania. Wprowadzeniu w latach 60-tych ubiegłego stulecia katalogów OPAC oraz ich dynamiczny rozwój w latach 80 i 90-tych towarzyszyły liczne prace badawcze związane z efektywnością tego narzędzia. Katalogi tzw. pierwszej generacji funkcjonowały na zasadzie prostych prekoordynowanych indeksów generujących wykazy punktów dostępu wybranych ze struktury opisu bibliograficznego. Wraz z rozwojem tego narzędzia modyfikowano zarówno organizację punktów dostępu jak również wprowadzano nowe techniki indeksowania. Katalogi „następnych generacji” stosowały już działania na zbiorach w postaci wykorzystania algebry Boole’a oraz rozbudowane mechanizmy wyszukiwawcze i techniki wizualizacji, sortowania rezultatów wyszukiwania oraz graficzne interfejsy użytkownika. [Antelman, Lynema, Pace 2006] Opracowywane w tym czasie modele katalogów OPAC [Bates 1986; Pejtersen 1992; Hildreth 1995; Drabenstott 1996] doprowadziły do ukonstytuowania się narzędzia, które dzisiaj m.in. dzięki automatycznym technikom przetwarzania i wyszukiwania informacji:

- oferuje wiele punktów dostępu (każde pole rekordu bibliograficznego może posiadać status punktu dostępu),
- umożliwia nieformalne sposoby formułowania zapytania informacyjnego z wykorzystaniem techniki przeglądania np. indeksów,
- stosuje wiele technik wizualnej organizacji katalogów,
- wykorzystuje efektywne techniki indeksowania i wyszukiwania informacji (np. algebra Boole’a),
- umożliwia zdalny dostęp,
- prezentuje aktualne informacje o dostępności elementów zbioru informacyjnego,
- stosuje wyszukiwanie rozproszone (w wielu bazach jednocześnie),
- cechuje się funkcjonalnością i użytecznością. [Hildreth 1995]

Eksploracja potencjału informacyjnego we współczesnych katalogach OPAC poprzez punkty dostępu charakteryzuje się dużą elastycznością i funkcjonalnością, na którą wpływ mają nowe możliwości elektronicznych systemów informacyjno-wyszukiwawczych, ale również wiele lat intensywnych badań teoretycznych i praktycznych.

Specyficznym rodzajem punktów dostępu są elementy opisu bibliograficznego, które odwzorowują cechy treści obiektów informacyjnych. Przypisanie funkcji

wyszukiwawczej elementowi opisu odpowiedzialnemu za reprezentację i przechowywanie tego typu metainformacji jest równoznaczne z nadaniem mu statusu **punktu dostępu przedmiotowego** (ang. subject access point).

Zaproponowana przez B. Hjørlanda oraz L. Kyllsbech Nielsena koncepcja przedmiotowych punktów dostępu do zasobów systemów informacyjnych zakłada interpretację przedmiotu dokumentu jako jego potencjału epistemologicznego i informacyjnego. [Hjørland, Kyllsbech Nielsen 2001] Pojęcie *przedmiotu* wyrażane również poprzez termin *temat* dokumentu czy też funkcję *bycia przedmiotem dokumentu* jest ważnym problemem badawczym, któremu dotychczas poświęcono wiele opracowań. [Ungurian 1976; Hjørland 1992; Woźniak 2000] Przedmiotowe punkty dostępu do zasobów systemu informacyjno-wyszukiwawczego reprezentują wymiar treści dokumentów w postaci odpowiednich atrybutów dokumentów, które otrzymują określone wartości podczas procesu indeksowania. Tym samym źródłem do tworzenia przedmiotowych punktów dostępu są odpowiednie pola formatu opisu odpowiedzialne za reprezentację tego typu metainformacji kodowanych za pomocą przeznaczonych do tego narzędzi (także JIW). Źródła do tworzenia punktów dostępu przedmiotowego to:

- abstrakty,
- bibliografia załącznikowa,
- symbole klasyfikacyjne,
- deskryptory,
- terminy pochodzące z pełnych tekstów,
- słowa kluczowe,
- parateksty (elementy otoczenia tekstu np. tytuł, wstęp, przypisy objaśniające, noty wydawnicze),
- tytuły [Hjørland 2007a].

Każde z nich może posiadać status punktu dostępu do kolekcji, jeśli zostanie włączone w określone mechanizmy wyszukiwawcze.

Od lat 50-tych ubiegłego stulecia prowadzono eksperymenty badawcze nad efektywnością różnych rodzajów punktów dostępu w bibliograficznych bazach danych (np. tytuł, abstrakt, deskryptor, symbol klasyfikacji) oraz ich wpływu na współczynniki trafności i kompletności wyszukiwania informacji. [Hjørland 2007a] Komputeryzacja działalności bibliotecznej i informacyjnej zrewolucjonizowała możliwości wizualizacji narzędzi dostępu do kolekcji. Dotychczasowe ograniczenia wynikające z pojemności karty katalogowej oraz samej budowy tradycyjnych katalogów i kartotek

oraz bibliografii zostały zniesione dzięki wprowadzeniu elektronicznych systemów informacyjno-wyszukiwawczych. Umożliwiło to zastosowanie nowych, efektywniejszych sposobów manipulacji punktami dostępu.

B. Hjørland oraz L. Kyllsbech Nielsen [Hjørland, Kyllsbech Nielsen 2001] zaproponowali taksonomię przedmiotowych punktów dostępu, których źródłem identyfikacji są wymienione elementy opisu bibliograficznego dokumentu (Tab. 1.1.).

Tab. 1.1. Taksonomia przedmiotowych punktów dostępu [Źródło: Hjørland, Kyllsbech Nielsen 2001]

I. Punkty dostępu według agenta lub twórcy
<ul style="list-style-type: none"> <li>- tworzone przez autora (tytuł dokumentu, abstrakt, słowa kluczowe),</li> <li>- tworzone jako wartość dodana: dostarczane przez wydawców, redaktorów (np. tytuł czasopisma, nazwa wydawcy); osoby katalogujące/indeksujące i tworzące abstrakty (np. symbole klasyfikacyjne, deskryptory, słowa kluczowe, abstrakty); recenzentów, czytelników i innych twórców (np. recenzje, statystyki bestsellerów, cytowania, indeksy cytowań (oraz np. tagi w folksonomiach – przyp. Aut.)</li> </ul>
II. Punkty dostępu według rodzaju
<ul style="list-style-type: none"> <li>- werbalne / niewerbalne (w postaci symbolicznej),</li> <li>- długie formy / krótkie formy (abstrakty / słowa kluczowe),</li> <li>- formy kontrolowane / niekontrolowane,</li> <li>- derywowane / nadawane (tytuły / słowa kluczowe),</li> <li>- formy oparte na indeksowaniu fasetowym, kwestionariuszowym / formy nadawane podczas indeksowania współrzędnego,</li> <li>- bezpośrednie / pośrednie (np. deskryptory / tytuł czasopisma, bibliografia załącznikowa); punkty pośrednie są tworzone z myślą o innych celach niż wyszukiwanie informacji; tytuły są bezpośrednimi punktami dostępu gdyż stanowią podstawowe kryterium wyszukiwawcze,</li> <li>- zorientowane na zawartość / zorientowane na zapytanie,</li> <li>- formy prekoordynowane / postkoordynowane,</li> <li>- formy z regułami syntaktycznymi / formy bez reguł syntaktycznych,</li> <li>- tworzone manualnie / generowane automatycznie.</li> </ul>

B. Hjørland [Hjørland 2007a] twierdzi, że wartość wyszukiwawcza określonego rodzaju punktu dostępu przedmiotowego jest względna i różni się dla dokumentów w zależności od dziedziny wiedzy. Proponuje on włączenie oceny wartości wyszukiwawczych punktów dostępu do zaproponowanej przez niego metodologii analizy domen. [Hjørland 2002] Przedmiotowe punkty dostępu rozumiane są tutaj jako semantyczne kondensatory tekstów, których treść reprezentują i do których lokalizacji skierowują.

## **1.4. Podsumowanie**

Przedstawiona koncepcja organizacji informacji zakłada interpretację tego terminu w kategoriach procesu wykorzystującego analizę:

- jego celu,
- jednostki poddawanej organizacji,
- oraz stosowanych narzędzi.

W zakres tego terminu włączono również znaczenie terminu *organizacja wiedzy*, utożsamianego z *systemami organizacji wiedzy* i rozumianego jako narzędzia odwzorowujące, reprezentujące i porządkujące informacje ekstrahowane na podstawie analizy wymiaru treści dokumentu/obiektu informacyjnego. Przedstawiona koncepcja punktów dostępu oraz przedmiotowych punktów dostępu stanowi podstawę teoretyczną do rozważań nad narzędziami organizacji informacji. Ich funkcjonowanie i efektywność uzależnione są od prawidłowej konstrukcji formatu opisu obiektów informacyjnych oraz narzędzi odwzorowujących tego typu metainformacje. Między innymi taką funkcję pełnią języki informacyjno-wyszukiwawcze jako wyspecjalizowane, sztuczne systemy semiotyczne. Cały proces organizacji informacji ma za zadanie udostępnienie użytkownikowi narzędzi strukturalnego dostępu do kolekcji. Zarówno dla zasobów tradycyjnych jak i cyfrowych organizacja informacji określonej kolekcji wiąże się z identyfikacją elementarnej jednostki opisu, która w przypadku zasobów rozproszonych może nosić nazwę *obiektu informacyjnego*. Na cały proces organizacji informacji składają się:

- metody i techniki analizy kolekcji pod kątem wyodrębnienia podstawowej jednostki, poddawanej dalszym manipulacjom,
- zasady i narzędzia odwzorowywania cech formalnych obiektów informacyjnych,
- zasady i narzędzia odwzorowywania cech treści obiektów informacyjnych (systemy organizacji wiedzy),
- zasady budowy narzędzi dostępu do zasobów kolekcji z wykorzystaniem koncepcji punktów dostępu.

## 2. Organizacja zasobów World Wide Web

Rozwój World Wide Web przypada na początek lat 90-tych ubiegłego stulecia i wiąże się z rozwinięciem hipertekstowego modelu organizacji i wizualizacji informacji, do czego przyczyniły się prace m.in. Tima Berners-Lee oraz organizacji CERN (Conseil Européen pour la Recherche Nucléaire). Początkowo WWW powstał jako projekt wymiany dokumentów we wspomnianej sieci CERN, lecz został szybko spopularyzowany i udostępniony do szerokiego stosowania. Do charakterystycznych cech WWW należą:

- organizacja dokumentów do postaci obiektów informacyjnych (lub tzw. cząstek informacyjnych) wykorzystujących wewnętrzną strukturalizację opartą na znacznikowych językach opisu,
- każdy dokument lub obiekt informacyjny posiada swój indywidualny identyfikator lokalizujący do w sieci – URL Uniwersalny Lokalizator Zasobów (ang. Universal Resource Locator),
- URL można wiązać w strukturze hipertekstu z innymi obiektami oraz z wewnętrznymi elementami dokumentu,
- narzędziem prezentacji informacji jest przeglądarka internetowa wyposażona w interaktywny interfejs, który realizuje wysyłanie zapytań z wykorzystaniem mechanizmów wyszukiwawczych lub inne formy dostępu do zasobów WWW,
- komunikacja przeglądarek internetowych z serwerami udostępniającymi kolekcje jest regulowana przez standardowy język komend – HTTP Hypertext Transfer Protocol. [Ellis, Vasconcelos 2000]

WWW utożsamiana jest z multimedialną częścią Internetu, a więc z dokumentami tekstowymi, graficznymi oraz audiowizualnymi, które są udostępniane internautom na określonych warunkach. Z badań prowadzonych w School of Information Management and Systems przy University of California w Berkeley nad przyrostem informacji w różnego rodzaju mediach wynika, że zasób informacyjny Internetu w 2002 roku wynosił 167 terabajtów. Jednocześnie pojemność informacyjna tzw. ukrytego Internetu\* wynosiła około 92 terabajtów. [Lyman, Varian 2003] Ten olbrzymi zasób informacyjny dynamicznie powiększa się, a jednocześnie wiele treści bezpowrotnie ulega utraceniu.

---

\* Ukryty czy też niewidoczny, głęboki Internet to zasoby WWW nieindeksowane przez wyszukiwarki. Należą do niego m.in. treści generowane automatycznie, zawartości niepowiązane z innymi za pomocą hiperłączy, treści o regulowanym dostępie.

Obecną sytuację dotyczącą efektywności realizacji procesów informacyjnych w środowisku sieciowym charakteryzuje tzw. „paradoks Konigera i Janowitza”. [Koniger, Janowitz 1995] Stan ten wyraża się lakonicznym sformułowaniem: „spragniony wiedzy topi się w informacji”. Określa ono sytuację, w jakiej znajduje się współczesny użytkownik wobec heterogenicznych i rozproszonych zasobów sieciowych. Z jednej strony ma do czynienia z olbrzymim zasobem, któremu towarzyszy nadmiarowość informacji, z drugiej brak efektywnych narzędzi wyszukiwania i selekcji powoduje chaos informacyjny i poczucie zagubienia. Tym samym funkcja filtrowania informacji wykorzystująca m.in. kryteria selekcji oraz organizacji informacji wydaje się być jedną z najważniejszych, pożądanых cech współczesnych systemów dostępu do rozproszonych zasobów sieciowych. [Lancaster 2003]

Organizacja informacji w zasobach WWW zostanie przedstawiona w tym rozdziale z dwóch punktów widzenia. Pierwszy będzie uwzględniał organizację zasobów WWW jako środowiska informacyjnego. Drugi, dotyczył będzie organizacji informacji w internetowych systemach i serwisach informacyjnych. Celem tego rozdziału jest przeniesienie aspektu teoretycznego organizacji informacji do sieciowego środowiska informacyjnego WWW i próba określenia roli w tym procesie języka informacyjno-wyszukiwawczego. Rozważania nad tymi zagadnieniami będą umiejscowione w ramach hiperinformacyjnego modelu organizacji, który warunkuje możliwości strukturyzacji przestrzeni informacyjnej w zasobach WWW.

## ***2.1. Hipertekstowy model organizacji informacji***

Koncepcja organizacji informacji w środowisku hipertekstowym zakłada strukturalizację zbioru informacji poddawanego organizowaniu z wykorzystaniem elementów konstytutywnych dla hipertekstu, czyli węzłów i powiązań hipertekstowych. „Podstawowa idea hipertekstu sprowadza się do opracowania metod strukturalizacji i prezentacji informacji zawartej w komunikatach językowych (dokumentach), które zapewniają wzbogacenie jej linearnej organizacji charakterystycznej dla tekstów tradycyjnych dokumentów piśmienniczych”. [Sosińska-Kalata 1999, s. 282] Należą do nich elementy typograficzne tradycyjnych dokumentów (np. tytuły, śródtytuły, indeksy, itp.), ale również nowe elementy strukturalne, których identyfikacja i włączenie w narzędzia dostępu są możliwe dzięki elektronicznemu przetwarzaniu i wizualizacji dokumentów.

### 2.1.1. Geneza hipertekstu

W piśmiennictwie poświęconym genezie hipertekstu [Abramowicz 1996; Carr i in. 1999; Gliński 1994; Januszewska 1992; Wereszczyńska-Cisło 1997] przywołuje się koncepcję Vannevara Busha jako twórcy prototypu maszyny, która wykorzystywała podobny sposób strukturalizacji kolekcji. W 1945 roku w artykule pt. „As we may think” [Bush 1945] przedstawił on koncepcję maszyny o nazwie *Memex*, której głównym zadaniem było usprawnienie pracy badacza poprzez wykorzystanie relacji asocjacyjnych w organizowaniu jego zbioru informacyjnego. „Memex miał postać biurka wyposażonego w ekran, na którym wyświetlano potrzebne informacje. Urządzenie uzupełniała klawiatura oraz zestaw przycisków i dźwigni służących do sterowania. [...] Umożliwiał on tworzenie połączeń pomiędzy dwoma fragmentami informacji z całej biblioteki. [...] Podstawową cechą systemu *Memex* była możliwość kojarzenia ze sobą dwóch dowolnych fragmentów bazy. Tę właśnie cechę przejął hipertekst. Koncepcja Busha oparta została na strukturze ludzkiej pamięci:

*Ludzka pamięć operuje skojarzeniami. Jedno uchwycone zdarzenie jest natychmiast w wyniku myślenia kojarzone z następnym, kojarzenie to odbywa się zgodnie ze skomplikowaną siateczką powiązań zawartą w komórkach mózgu”* [Januszewska 1992, s. 87]

Pomysł Vannevara Busha oparty był na metodzie organizacji dokumentów, która odzwierciedlała nie tyle abstrakcyjne relacje systemowe, co związki o charakterze asocjacyjnym tworzone w procesie percepcji informacji. *Memex* był urządzeniem przystosowanym do indywidualnego użytku, który uwzględniał potrzeby konkretnych użytkowników i pracował na zbiorze informacyjnym złożonym zarówno z dokumentów tekstowych, graficznych jak i dźwiękowych. Niedogodnością tego urządzenia był długi czas oczekiwania na pożądaną dokument, ponieważ system ten pracował na mikrofilmach, które przetwarzane były automatycznie (za pomocą mechanizmów dźwigni i przekładni) nie zaś elektronicznie. Brak wsparcia technicznego uniemożliwił dalszy rozwój tego urządzenia. Po *Memexie* została jednak jak ważna koncepcja organizacji zbioru informacyjnego systemu oparta na wykorzystaniu relacji asocjacyjnych w określaniu połączeń pomiędzy wydzielonymi „porcjami” informacji.

Gwałtowny rozwój badań nad hipertekstem nastąpił dopiero w latach osiemdziesiątych dwudziestego wieku. Postęp technologiczny, a w szczególności coraz szersze zastosowanie komputerów osobistych w życiu codziennym, udostępnił techniczne narzędzia do realizacji koncepcji Vannevara Busha. Wraz z pojawianiem się

publikacji próbujących określić naturę hipertekstu, pojawiły się także próby zdefiniowania tego terminu. B. Wereszczyńska-Cisło proponuje podział definicji hipertekstu, które pojawiły się na przełomie lat osiemdziesiątych i dziewięćdziesiątych ubiegłego stulecia, na cztery grupy odzwierciedlające sposób ujęcia tego zagadnienia. Są to definicje, które:

- kładą nacisk na techniczne aspekty projektowania hipertekstu,
- prezentują techniczne zagadnienia wykorzystania i rozbudowy systemów hipertekstowych przez użytkowników,
- opisują intelektualne procesy towarzyszące projektowaniu i wykorzystaniu hipertekstu,
- skupiają się na podobieństwie do innych systemów pod kątem interakcyjności użytkownika oraz wykorzystania relacji asocjacyjnych. [Wereszczyńska-Cisło 1997, s. 17]

W pierwszej grupie znalazły się opinie, które bardzo ogólnikowo traktują hipertekst jako narzędzie gromadzenia i udostępniania informacji. Spotykamy tutaj zdania typu:

- hipertekst jest sposobem gromadzenia i udostępniania informacji,
- hipertekst to nowa technika organizowania informacji,
- hipertekst to tworzenie i reprezentacja powiązań pomiędzy dyskretnymi fragmentami danych,
- hipertekst jest rodzajową nazwą zlepku technik – stosowanych w celu wprowadzania, gromadzenia, udostępniania, przechowywania i wyszukiwania informacji [...],
- system hipertekstowy to oprogramowanie umożliwiające tworzenie i/lub przeglądanie hipertekstu. [Wereszczyńska-Cisło 1997, s. 18]

Drugą grupę definicji tego terminu stanowią stwierdzenia kładące nacisk na funkcjonalność systemu hipertekstowego. W publikacjach tych, autorzy skupiają swoją uwagę na elementach składających się na hipertekst, czyli węzłach i powiązaniach. „Hipertekst jest charakteryzowany jako narzędzie umożliwiające elastyczność wyszukiwania informacji i stosowanie za każdym razem strategii określanej na bieżąco przez danego użytkownika”. [Wereszczyńska-Cisło 1997 s. 18] W trzeciej grupie wypowiedzi akcentowany jest aspekt nielinearności hipertekstu w porównaniu do percepcji informacji w tradycyjnych dokumentach.

Warunkiem koniecznym dla systemu hipertekstowego jest funkcjonowanie zbioru informacyjnego (zazwyczaj w formie dokumentów tekstowych, ale obecnie równie często są to zbiory o naturze mieszanej zbudowane zarówno z elementów



tekstowych, graficznych i dźwiękowych – wtedy mamy do czynienia, w dużym uproszczeniu z systemem hipermedialnym), elementów strukturalnych tego zbioru w postaci komunikatów zwanych węzłami hipertekstowymi oraz sieci relacji semantycznych zidentyfikowanych pomiędzy węzłami zwanymi powiązaniem hipertekstowymi (także linkami, hiperłączami), a także warunków technicznych umożliwiających funkcjonowanie systemu.

Interpretacja elementów strukturalnych hipertekstu oraz ich organizacji ewoluowała wraz powstawaniem systemów wykorzystujących jego koncepcję. Istota hipertekstu pozostawała niezmienna i zdaniem jego twórcy Teda Nelsona polega na *pisaniu niesekwencyjnym* [Adams 1998], które opiera się na tworzeniu połączeń pomiędzy fragmentami tekstu lub innego nośnika informacji (w rozumieniu logicznym nie zaś fizycznym) na podstawie indywidualnych preferencji autora. Połączenia te, wyznaczają ścieżki eksploracji tekstu przez użytkowników i stanowią wartość dodaną dokumentu. Wraz z rozwojem badań nad hipertekstem, powstające definicje określały jego istotę w sposób coraz bardziej kompleksowy. W 1997 roku B. Wereszczyńska-Cisło przez hipertekst rozumiała „narzędzie informacyjne umożliwiające szybki, zindywidualizowany i dynamiczny dostęp do określonych informacji w odpowiednio ustrukturalizowanej za pomocą relacji semantycznej w tekstowej bazie danych lub sieci takich baz”. [Wereszczyńska-Cisło 1997, s. 17] W wydanym w 2002 roku „Słowniku encyklopedycznym informacji, języków i systemów informacyjno-wyszukiwawczych” pojęcie to interpretowano jako „wielopoziomową strukturę będącą odwzorowaniem cech relewantnych tekstu (także tekstu mieszanego) lub zbioru tekstów, umożliwiającą przekształcanie tekstu linearnego w postać nielinearną (najczęściej za pomocą drzew hierarchicznych oraz sieci) w celu ułatwienia odbiorcy wyszukania zawartych w nim informacji oraz uzupełnienia zbioru informacji wyselekcjonowanych odpowiednio do potrzeb informacyjnych użytkownika. Struktura taka może różnego typu relacjami intertekstualnymi łączyć w jedną całość części jednego dokumentu lub dokumenty zawierające informację dokumentacyjną z odpowiednimi dokumentami źródłowymi lub ich fragmentami odpowiednią informacją faktograficzną, przy czym w jednym hipertekście mogą występować dokumenty różnych typów”. [Słownik encyklopedyczny 2002, s. 78]

### **2.1.2. Struktura hipertekstu**

Węzły oraz powiązania są podstawowymi elementami hipertekstu. „Węzeł

stanowi wydzielony, ze względu na zawartość informacyjną, fragment tekstu (słowo, zdanie, akapit, rysunek, schemat, wideoklip, itp.) lub dokument w bazie danych, określany niekiedy jako jednostka informacyjna albo dyskretna kolekcja informacji”. [Wereszczyńska-Cisło 1997, s. 21] Węzeł hipertekstowy utożsamiany jest z porcją informacji, komunikatem umiejscowionym w określonym miejscu w strukturze dokumentu. W kontekście badań literaturoznawczych nad hipertekstem, węzeł interpretuje się również jako *leksję* i przywołuje tym samym badania Rolanda Barthesa. Autor ten zaproponował koncepcję tekstu idealnego, który „[...] nie ma początku, ma charakter odwracalny, można doń się dostać przez rozliczne wejścia, z których żadne nie powinno pochopnie zostać uznane za główne”. [Barthes, 1999, s 40] Sugeruje on analizę tekstu opartą na jego podziale na bloki znaczeń, szereg krótkich fragmentów „przylegających do siebie” zwanych *leksjami* jako jednostek lektury. Koncepcja „tekstu połamanego” R. Barthesa oraz badania nad tekstem w nurcie postmodernizmu i strukturalizmu można uznać za wartościowy wkład w rozwój badań również nad hipertekstem.

Powiązania hipertekstowe łączą elementy strukturalne dokumentu posiadające status węzła. Relacje pomiędzy węzłami mogą wyrażać:

- związki semantyczne,
- związki syntagmatyczne w postaci relacji:
  - o strukturalnych między węzłami, zwanych też syntaktycznymi lub retorycznymi (np. informacja w A precyzuje, potwierdza, wyjaśnia informację w B albo informacja w A jest niezgodna w B),
  - o związanych z łamaniem hipertekstu (np. wskazujących, że uzupełniające informacje muszą zostać wydrukowane na tej samej stronie),
- związki pragmatyczne (np. po zapoznaniu się z informacją w węźle A należy zapoznać się z informacją w węźle B). [Wereszczyńska-Cisło 1997, s. 21]

Powiązania hipertekstowe są wykładnikiem relacji ustanowionej w zbiorze węzłów. Związki te mogą występować pomiędzy węzłami pochodzącymi z tego samego dokumentu, wtedy mamy do czynienia z relacjami intratekstualnymi lub pomiędzy węzłami z różnych dokumentów, reprezentując tym samym relacje intertekstualne. Powiązania mogą być jedno- lub dwukierunkowe. Opracowanie sieci powiązań pomiędzy węzłami wyznacza możliwości eksploracji zbioru informacji, w którym zastosowano hipertekstową organizację informacji. Aktywacja powiązań hipertekstowych, w które wyposażone są węzły powoduje odpowiedź systemu w postaci wyświetlenia zawartości węzłów powiązanych określoną relacją, ale również

określa sposoby nawigowania, czyli przemieszczania się pomiędzy węzłami i przetwarzania informacji tam zawartych. [zob. Skórka 2006]

W zależności od specyfiki systemów, w których zastosowano hipertekst do organizacji informacji i wizualizacji procesów wyszukiwawczych wykorzystywano różną terminologię dla węzłów i powiązań hipertekstowych. Do najciekawszych i najwcześniejszych należy system NLS (oN Line System) autorstwa D. Engelbarta, który był przeznaczony dla naukowców. Powstał w latach 60-tych ubiegłego stulecia, a w latach 80-tych dzięki modyfikacjom i zastosowaniu nowych technologii komputerowych został przekształcony do postaci systemu o nazwie Xanadu. Lata 80-te charakteryzują się dużą liczbą systemów, w których wykorzystano i ulepszano koncepcje hipertekstu. Wtedy powstały m.in. systemy NoteCards, HyperCard, Virtual Notebook System. Do najbardziej rozpowszechnionego systemu opartego na koncepcji hipertekstu należy niewątpliwie sieć World Wide Web. To właśnie z tą siecią kojarzy się często rzeczywiste wykorzystanie hipertekstu, co jak wskazano jest niepełne, gdyż geneza systemów hipertekstowych sięga lat 60-tych ubiegłego stulecia.

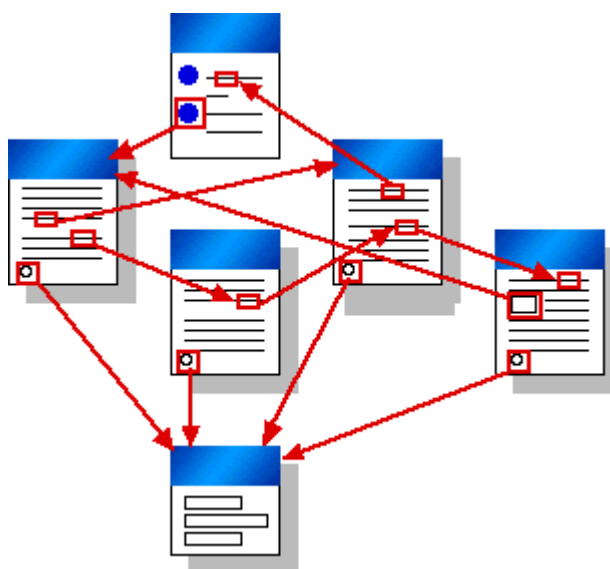
Analiza struktury systemów hipertekstowych zaproponowana przez R. Radę zakłada wyodrębnienie podtypów hipertekstu w systemach informacyjnych, są to [Rada 1991] :

- mikrotekst – dokument tekstowy z wyróżnionymi explicite powiązaniem pomiędzy jego fragmentami (bazującymi na relacjach intratekstualnych),
- makrotekst – system hipertekstowy zbudowany z wielu dokumentów, gdzie wyodrębnione węzły powiązane są różnymi rodzajami powiązań (intra- i intertekstualne),
- tekst grupowy – system hipertekstowy tworzony przez grupę osób,
- ekspertekst – połączenie hipertekstu z systemem ekspertowym, gdzie wykorzystuje się narzędzia sztucznej inteligencji do przetwarzania informacji zgromadzonych w węzłach.

Do podstawowych cech systemów hipertekstowych należą: „nielinearność, elastyczność, przezroczystość, gościnność i kontekstowość”. [Sosińska-Kalata 1999, s. 284] Cechy te dają możliwości kształtowania organizacji informacji w kolekcji poprzez identyfikację i porządkowanie relewantnych punktów dostępu oraz systemów organizacji wiedzy poprzez zastosowanie map konceptualnych, zagnieżdżonych w środowisku węzłów oraz powiązań hipertekstowych. Nielinearność zapewnia dostęp do dokumentów oraz komunikatów w nich zidentyfikowanych niezależnie od sekwencyjnego porządku prezentacji informacji. Elastyczność umożliwia dostosowanie

sieci węzłów do indywidualnych potrzeb użytkowników. Stopień szczegółowości komunikatów posiadających status węzła oraz liczba i rodzaj relacji intra- i intertekstualnych ich łączących jest opracowywany podczas projektowania struktury hipertekstu. Przejrzystość przejawia się w możliwości ekstrahowania z hipertekstu tekstów powiązanych z nim relacjami intertekstualnymi oraz ich elementów strukturalnych, które stanowią człony relacji w powiązaniach hipertekstowych. Gościnność hipertekstu polega na możliwości włączania do struktury zorganizowanej za pomocą takiego modelu nowych punktów dostępu na podstawie potrzeb i preferencji użytkowników. Kontekstowość hipertekstu należy do jego cech charakterystycznych i polega na tworzeniu powiązań pomiędzy węzłami, które odsyłają do informacji o charakterze wyjaśniającym i definiującym, ale również identyfikujących kontekst użycia określonych wyrażen w innych węzłach w strukturze hipertekstu.

Hipertekst jako narzędzie organizacji informacji wyznacza możliwości eksploracji kolekcji poprzez zbiór węzłów reprezentujących określone porcje informacji oraz sieć relacji o charakterze semantycznym, syntagmatycznym, asocjacyjnym i pragmatycznym oznaczonych pomiędzy węzłami i reprezentujących związki intra- oraz intertekstualne (Rys. 2.1).



Rys. 2.1. Struktura hipertekstu

[Źródło:[http://bp2.blogger.com/\\_O1XqChxp7ak/R0wWAUsTxII/AAAAAAAAAAc/bnucHvFCs7s/s1600-h/thema\\_hypertext.gif](http://bp2.blogger.com/_O1XqChxp7ak/R0wWAUsTxII/AAAAAAAAAAc/bnucHvFCs7s/s1600-h/thema_hypertext.gif)]

Zagadnienia istoty i budowy hipertekstu są złożone i poświęcono im

w piśmiennictwie wiele uwagi. Kompleksowa charakterystyka hipertekstowego modelu organizacji informacji wykracza jednak poza ramy rozprawy.

## **2.2. Architektura informacji**

Charakteryzując hiperinformacyjny model organizacji informacji należy zwrócić również uwagę na jego implementację w postaci koncepcji o nazwie *architektura informacji*. Idea architektury informacji polega na wykorzystaniu hipertekstowej strukturyzacji kolekcji w procesach organizowania informacji i realizacji dostępu w sieciowych serwisach i systemach informacyjnych. Termin *architektura informacji* definiowany jest jako:

- proces organizowania nazewnictwa, projektowania nawigacji i systemów wyszukiwawczych pomocnych w odnajdywaniu i zarządzaniu informacją,
- projektowanie strukturalne udostępnianych w sieci środowisk informacyjnych,
- sztuka oraz nauka organizowania informacji w celu ułatwienia ludziom efektywnego zaspokojenia ich potrzeb informacyjnych,
- sztuka, nauka i działalność polegająca na organizowaniu informacji w taki sposób, aby miała ona sens dla jej odbiorców. [Skórka 2002]

Koncepcja architektury informacji wywodzi się z badań nad szeroko rozumianą użytecznością (ang. usability) serwisów internetowych, a jej celem jest organizacja informacji w internetowych serwisach hipertekstowych za pomocą skonstruowanych na te potrzeby odpowiednich narzędzi. Zalicza się do nich:

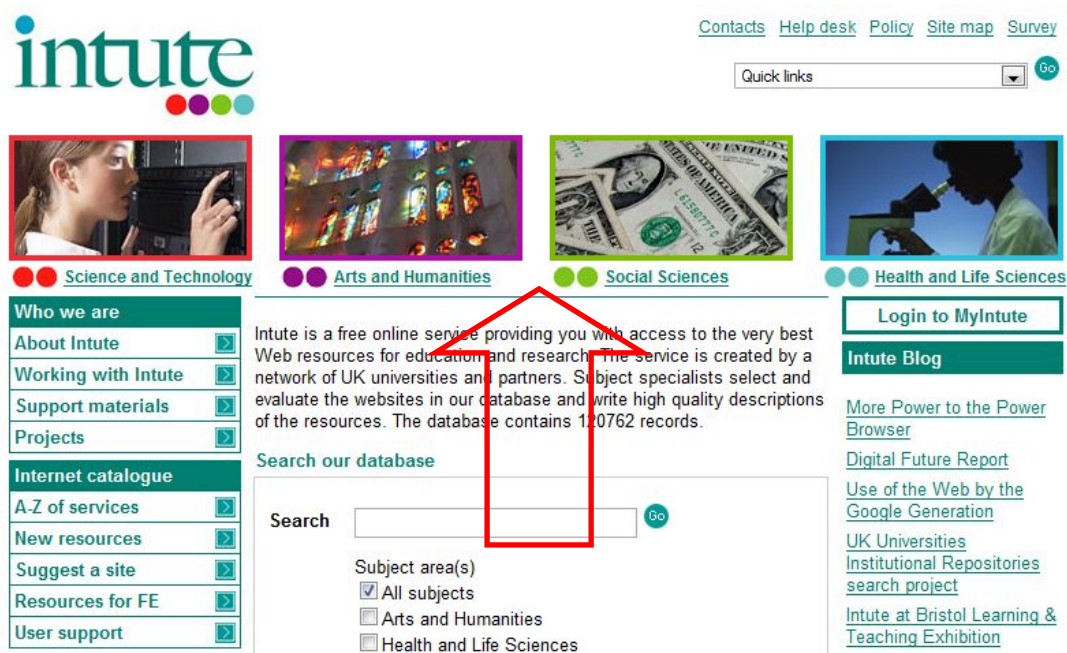
- systemy organizacyjne,
- systemy oznaczania,
- systemy nawigacji,
- systemy wyszukiwawcze. [Morville 2003]

Systemy organizacyjne stanowią podstawowy zrąb architektury informacji. Analiza zbioru informacji, który ma zostać zorganizowany, uwzględnia charakter jednostek ją tworzących zarówno pod kątem formalnym oraz treściowym, a także identyfikację potencjalnych punktów dostępu w postaci węzłów hipertekstowych. Wyodrębnione w ten sposób obiekty informacyjne, które posiadają status węzła lub mają postać całego dokumentu analizuje się pod kątem wchodzenia w relacje z innymi elementami kolekcji. W tym celu tworzy się na potrzeby określonego systemu narzędzia umożliwiające kategoryzowanie treści zwartych w wyodrębnionych jednostkach.

Dopełnieniem systemów organizacyjnych są systemy oznaczania, których celem jest zaopatrywanie schematów organizacyjnych w zbiory leksyki języka naturalnego

używane do nazywania pojęć reprezentujących treść i formę elementów zbioru informacyjnego. W tym celu wykorzystuje się opracowane na potrzeby systemów zbioru słownictwa, ale również podzbiory słowników języków informacyjno-wyszukiwawczych. Systemy oznaczania wykorzystywane są zarówno w indeksowaniu obiektów informacyjnych, jak również do identyfikacji ich elementów typograficznych w postaci np. nagłówków, śródtytułów, które mogą zostać wykorzystane jako węzły hipertekstowe i pełnić funkcję punktów dostępu. [Roszkowski 2004]

Systemy nawigacyjne umożliwiają eksplorację organizacji hipertekstowej i są pochodne wobec systemów organizacyjnych. Wyznaczają porządek punktów dostępu do kolekcji poprzez zestawy węzłów odsyłających do określonych stref zawartości systemu. Składają się z podsystemów: nawigacji globalnej, lokalnej oraz kontekstowej. Narzędzia nawigacji globalnej są stałe dla całego serwisu i mają formę kilku węzłów, których hiperłącza odsyłają do podstawowych grup zawartości serwisu czy też usług przez niego oferowanych (Rys. 2.2.).



Rys. 2.2. Nawigacja globalna w systemie Intute. [Źródło: <http://intute.ac.uk/>]

Na Rys. 2.2. funkcję schematu nawigacji globalnej pełnią cztery węzły o nazwach:

- Science and Technology,
- Arts and Humanities,
- Social Sciences,
- Health and Life Science.

Odsyłają one do czterech głównych grup tematycznych w systemie Intute. „Nawigacja globalna może również stosować podział kategoryalny nieodwzorowujący treści

serwisu, lecz identyfikujący obszary zawartości, do których ma dostęp użytkownik”. [Roszkowski 2004] Nawigacja lokalna jest konsekwencją rozbudowy podziału kategorialnego zastosowanego w nawigacji globalnej. Dla danego terminu, przyjętego za jedną z kategorii najwyższego poziomu w nawigacji globalnej, w trakcie opracowywania systemów organizacyjnych tworzy się zestaw kategorii tematycznych o węższym zakresie znaczeniowym w oparciu o przyjętą zasadę podziału. W ten sposób w nawigacji lokalnej dla każdej z kategorii najwyższego stopnia tworzy się rozgałęzione struktury drzewiaste uwzględniające terminy powiązane ze sobą i jednocześnie pełniące funkcję węzłów, których odesłania hipertekstowe odsyłają do konkretnych stref zawartości, czy obiektów w bazie danych.

W przypadku nawigacji kontekstowej projektuje się dodatkowe narzędzia eksploracji kolekcji, które opierają się na relacjach asocjacyjnych. Dla określonych stref zawartości systemu, czy elementów zbioru wyszukiwawczego, aktualnie wyświetlanych użytkownikowi, projektuje się dodatkowe łącza mające na celu zwiększenie kompletności wyszukiwania. W serwisach edukacyjnych mogą to być np. odwołania do artykułów omawiających pokrewne tematy [Morville, Rosenfeld 2003 s. 142], w serwisach komercyjnych może to być prezentacja produktów z tej samej grupy, itd. Łącza nawigacji kontekstowej w grupie serwisów komercyjnych przyjmują zazwyczaj formę: *„klienci, którzy kupili ten produkt, kupili również:”*. Tego typu schematy nawigacyjne opierające się na relacjach kojarzeniowych prowadzą do wzrostu liczby zamówień (ang. up-sell) oraz poszerzenia zakresu zamawianych produktów (ang. cross-sell). [Morville, Rosenfeld 2003 s. 144]

Celem systemów wyszukiwawczych jest udostępnienie użytkownikowi mechanizmów przeszukujących zawartość bazy danych za pomocą określonego języka zapytań oraz według dostępnych kryteriów wyszukiwawczych. W zależności od modelu wyszukiwania informacji realizowanego w danym systemie oraz sposobów budowy zapytań informacyjnych w systemach wyszukiwawczych stosuje się różnorodne rozwiązania techniczne. Istotą systemów wyszukiwawczych jest ich działanie jako narzędzi komplementarnych w stosunku do niesformalizowanych sposobów pozyskiwania informacji z systemu, jakie towarzyszą eksploracji z wykorzystaniem systemów nawigacyjnych.

Koncepcja architektury informacji jako jednego ze sposobów realizacji hiperinformacyjnego modelu organizacji informacji jest często stosowana w systemach i serwisach komercyjnych. Podstawy teoretyczne tej koncepcji mają swoje korzenie w bibliotekoznawstwie i informacji naukowej, gdyż jeden z jej twórców, Peter Morville

jest z wykształcenia bibliotekoznawcą.

### 2.3. *Metadane*

Terminem metadane, czyli dane o danych, zwykle się określać formaty opisu dla dokumentów elektronicznych opublikowanych w Internecie. [Chu 2003, s. 37] Schemat metadanych można więc interpretować jako język opisu bibliograficznego dla dokumentów sieciowych. Reprezentacja informacji o zasobach cyfrowych w WWW realizowana jest z wykorzystaniem wielu standaryzowanych formatów opisu adekwatnych dla treści lub formy obiektów informacyjnych. Do pożądanych cech schematów metadanych zalicza się:

- modułowość, która polega na komponowaniu schematu z bloków metadanych reprezentujących określone kategorie cech obiektów;
- rozciągłość, czyli zdolność schematu do reprezentacji dodatkowych cech relewantnych lub uszczegóławiania wartości, jakie przyjmują istniejące atrybuty;
- uściślanie, czyli stosowanie adekwatnego stopnia szczegółowości względem potencjalnych potrzeb informacyjnych użytkowników. [Duval i in. 2002]

Zestaw metadanych funkcjonujący pod nazwą Dublin Core (pełna nazwa Dublin Core Metadata Element Set) jest obecnie najbardziej rozpowszechnionym formatem opisu obiektów cyfrowych w Internecie. Został utworzony w 1995 roku podczas tzw. pierwszych warsztatów Dublin Core, na którym opracowano jego podstawy oraz wytyczono strategię rozwoju. W 1999 roku uzgodniona została ostateczna struktura formatu, na którą składa się 15 podstawowych elementów (Tab. 2.1).

Tab. 2.1. Zestaw podstawowych elementów Dublin Core  
[Źródło: Nahotko 2004a]

Element	Charakterystyka
DC.title	Tytuł
DC.creator	Twórca
DC.subject	Opis rzeczowy
DC.description	Opis
DC.publisher	Wydawca
DC.contributor	Współtwórca
DC.date	Data
DC.format	Format
DC.identifier	Identyfikator
DC.source	Źródło
DC.language	Język
DC.relation	Relacja
DC.coverage	Zasięg chronologiczny i geograficzny
DC.rights	Własność



Celem wyodrębnionych elementów jest identyfikacja relewantnych własności obiektów cyfrowych. Do najważniejszych cech charakteryzujących Dublin Core Metadata Element Set należą:

- prostota tworzenia i utrzymania metadanych,
- powszechnie zrozumiała semantyka,
- międzynarodowy zasięg wykorzystania,
- łatwość rozbudowy,
- współdziałanie pomiędzy zasobami i systemami indeksowania. [Nahotko 2004a, s. 124]

Struktura Dublin Core oprócz podstawowych elementów zakłada również uszczegółowianie charakterystyk obiektów cyfrowych poprzez wykorzystanie odpowiednich kwalifikatorów. Pełnią one funkcję uszczegóławiającą i/lub normalizującą schemat kodowania danych. W przypadku pierwszego rodzaju kwalifikatorów uszczegółowienie elementu polega na zawężeniu zakresu, w jakim cecha obiektu informacyjnego jest odwzorowywana. W odniesieniu np. do elementu *data* (ang. *date*) schemat ten dopuszcza następujące kwalifikatory (<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>):

*Data:*  
*utworzenia,*  
*walidacji,*  
*dostępności,*  
*opublikowania,*  
*modyfikacji.*

Celem drugiego rodzaju kwalifikatorów jest „identyfikacja schematów służących interpretacji wartości elementu. Wartość wyrażona przy pomocy schematu kodowania będzie [...] symbolem wybranym ze słownika kontrolowanego (np. symbol z tablic klasyfikacji bądź zestaw haseł przedmiotowych) lub ciągiem sformatowanym odpowiednio według użytej formalnej notacji (np. „2001-01-01” jako standardowe wyrażenie daty)”. [Nahotko 2004a, s. 125] Schemat Dublin Core z uwagi na swoją uniwersalność jest przeznaczony do ogólnego stosowania.

Duża liczba schematów metadanych, funkcjonująca w Internecie jako narzędzie formatowania jednostek opisu, powoduje problemy związane z interoperacyjnością danych, czyli brakiem spójności na płaszczyźnie ich automatycznej wymiany oraz rozproszonego wyszukiwania informacji. Z drugiej strony, dedykowane zestawy metadanych umożliwiają odwzorowanie pożądaných cech obiektów informacyjnych,

specyficznych dla ich treści lub formy, które nie są uwzględniane przez formaty uniwersalne.

#### ***2.4. Reprezentacja metainformacji***

Schemat metadanych wyznacza zestaw atrybutów obiektu informacyjnego, którym nadawana jest wartość podczas automatycznego lub manualnego indeksowania. Tak jak w przypadku zasad kodowania danych bibliograficznych, również metadane zasobów cyfrowych muszą zostać zapisane w postaci komunikatu o określonej strukturze. Standard RDF (ang. Resource Description Framework) opracowany przez World Wide Web Consortium jest specyfikacją modelu metadanych zakładającą ich reprezentację w postaci zdolnej do przetworzenia przez aplikacje internetowe. Schemat Dublin Core identyfikuje zestaw atrybutów obiektu informacyjnego, natomiast RDF określa formalne zasady kodowania wartości tych atrybutów. [Lassila; Swick 1999] Zapis zgodny RDF polega na użyciu trzech kategorii zmiennych: podmiotu, predykatu oraz obiektu. [Nahotko 2004a, s. 154] Wykorzystanie tych elementów oraz zasad gramatyki tego języka opisu danych pozwala na reprezentację metainformacji, która poprzez dołączenie lub zagnieżdżenie w obiekcie informacyjnym może zostać automatycznie przetwarzana.

RDF jest formalną specyfikacją modelu metadanych i standardem reprezentacji metainformacji dla obiektów cyfrowych na wysokim stopniu abstrakcji. W World Wide Web istnieje jednak wiele innych sposobów kodowania metainformacji. Większość z nich wywodzi się z międzynarodowego standardu reprezentacji elementów strukturalnych tekstów w postaci elektronicznej SGML (ang. Standard Generalized Markup Language). Jest on typem znacznikowego języka opisu danych, który został uznany w 1986 roku przez International Organization for Standardization jako standard w tym zakresie. [Sperberg-McQueen, Burnard 1999] Jego podstawową cechą jest opis logicznej struktury elementów dokumentu i jej oddzielenie od prezentacji jego zawartości. Elementy te noszą nazwę znaczników (ang. tags), a ich rodzaj i sposoby użycia są zależne od formalnej definicji typu dokumentu (ang. dokument type definition – DTD). Elastyczność SGML polega w głównej mierze na możliwości opracowania wielu DTD dla określonych typów dokumentów elektronicznych i specyfikowania tym samym ich elementów strukturalnych, które zostają zidentyfikowane i opisane wraz z wykorzystaniem SGML z określonym DTD.

Najbardziej rozpowszechnioną obecnie implementacją SGML jest HTML (ang.

Hypertext Markup Language). Jest to opracowana w 1991 roku przez Tima Berners-Lee definicja typu dokumentu dla dokumentu hipertekstowego. HTML zawiera wykaz znaczników, które można wykorzystać w opisie dokumentu. Prostota HTML jest jednocześnie przyczyną jego ograniczonych możliwości w stosunku do SGML. HTML pozwala na stosowanie wyłącznie gotowego zestawu elementów strukturalnych dokumentu oraz dysponuje środkami o słabej sile reprezentacji relacji strukturalnych pomiędzy jego elementami. Budowanie opisu dokumentu z wykorzystaniem tego języka polega na wykorzystaniu elementów charakterystycznych dla języków znacznikowych: znacznik, element, atrybut oraz dodatkowego łącze (ang. link), który umożliwia reprezentację powiązań hipertekstowych, charakterystycznych dla HTML. Reprezentacja metainformacji w HTML polega na wykorzystaniu podstawowej interpretacji struktury dokumentu, która zakłada opis jego dwóch części: nagłówek (ang. head) oraz części zasadniczej (ang. body). To właśnie w nagłówku dokumentu (części niewyświetlanej użytkownikowi) zidentyfikowano szereg znaczników, które służą do kodowania metainformacji. Umożliwiają one zakodowanie informacji m.in. o autorze dokumentu, dacie jego opublikowania, charakterystykę treści wyrażoną za pomocą słów kluczowych oraz streszczenie jego treści.

Rozwinięciem koncepcji HTML jest język XML, który wyznacza obecnie kierunek rozwoju znacznikowych języków opisu. Rozszerzalny język znaczników (ang. Extensible Markup Language) jest przede wszystkim sposobem reprezentacji ustrukturyzowanych danych. Jego składnia jest bardziej restrykcyjna niż HTML, lecz jednocześnie prostsza i bardziej elastyczna niż SGML, z którego ten się wywodzi. XML pozwala na podstawie przyjętego DTD rozszerzać je o dodatkowe elementy strukturalne dokumentów za pomocą tzw. schematów XML. Te, w przeciwieństwie do DTD, należą do standardu samego XML i są o wiele elastyczniejsze w identyfikacji koniecznych do wyodrębnienia elementów struktury dokumentu. Język XML w bezpośredni sposób formatuje wewnętrzną budowę dokumentu i umożliwia przechowywanie danych wyposażonych w semantykę, a to pozwala na zastosowanie automatycznych mechanizmów przetwarzania informacji. Z poprawnie sformatowanego dokumentu, za pomocą języka XML, istnieje możliwość wygenerowania dla niego zestawu metadanych. Język ten wraz z jego wieloma implementacjami (np. XHTML - Extensible Hypertext Markup Language) jest szeroko stosowanym formatem strukturalizacji i przechowywania danych również w komercyjnych systemach informacyjnych w Internecie.

Reprezentacja metainformacji w zasobach WWW może zachodzić na etapie pre-

i postpublikacji. [Qin 2000] W pierwszym przypadku, mamy do czynienia z dołączaniem ustrukturyzowanych metadanych do tworzonych cyfrowych obiektów informacyjnych. W tym celu wykorzystuje się m.in. wskazane standardy reprezentacji. Jednym z nich jest Encoded Archival Description (EAD). Jest to standard kodowania metainformacji dla dokumentów archiwalnych w SGML. Reprezentacja metainformacji na etapie postpublikacji polega na automatycznym generowaniu wartości atrybutów obiektów cyfrowych do postaci struktury stosowanej w tego rodzaju mechanizmach. Zalicza się do nich przede wszystkim szperacze (ang. crawlers), czyli mechanizmy indeksujące wyszukiwarek internetowych.

## ***2.5. Narzędzia dostępu***

Zagadnienia narzędzi realizujących dostęp do zasobów WWW to przede wszystkim konieczność dostosowania koncepcji punktów dostępu do tego specyficznego środowiska informacyjnego. Hipertekst jako narzędzie strukturyzacji i łączenia dokumentów oraz ich elementów udostępnia szerokie możliwości identyfikacji punktów dostępu. Każdy element strukturalny dokumentu zidentyfikowany jako węzeł hipertekstowy może być jednocześnie punktem dostępu do tego dokumentu. W dalszej części tego rozdziału zostaną scharakteryzowane najczęściej stosowane narzędzia dostępu do zasobów WWW. Będą to: wyszukiwarki internetowe, katalogi internetowe, biblioteki cyfrowe oraz folksonomie.

### **2.5.1. Wyszukiwarki internetowe**

Wyszukiwarki internetowe to mechanizmy realizujące procesy wyszukiwawcze w środowisku WWW. Struktura takiego mechanizmu obejmuje bazę danych z adresami URL i charakterystykami wyszukiwawczymi dokumentów internetowych, mechanizm indeksujący (zwany szperaczem) oraz interfejs użytkownika. Mechanizm indeksujący automatycznie lokalizuje i selekcjonuje zasoby WWW oraz generuje ich charakterystyki wyszukiwawcze, które trafiają do zbioru wyszukiwawczego całego systemu. Interfejs użytkownika umożliwia konstrukcję zapytania informacyjnego.

W większości wyszukiwarek internetowych realizowany jest algorytmiczny model wyszukiwania informacji. Jest on „odzwierciedleniem operacji wykonywanych przez system wyszukiwawczy, prowadzących do wyselekcjonowania z bazy systemu

tych obiektów, które spełniają sformułowane przez użytkownika kryteria wyszukiwawcze. Podstawową czynnością realizowaną przez system, gwarantującą osiągnięcie tego celu, jest porównywanie reprezentacji pytań użytkowników (instrukcji wyszukiwawczych) z reprezentacjami obiektów (rekordami zawierającymi charakterystyki wyszukiwawcze) przechowywanych w zasobach systemu. W wyniku porównywania wyodrębniane są – na podstawie przyjętego kryterium zgodności – rekordy trafne (relewantne)”. [Próchnicka 2004, s. 26] Efektywność tego modelu zależy przede wszystkim od poprawnego sformułowania przez użytkownika zapytania wyszukiwawczego. W tego rodzaju mechanizmach stosuje się tradycyjny algorytm w postaci sekwencji kroków:

- rozpoznanie potrzeby informacyjnej,
- wybór systemu lub kolekcji,
- sformułowanie instrukcji wyszukiwawczej,
- wysłanie zapytania do systemu,
- otrzymanie wyników wyszukiwania w postaci obiektów informacyjnych,
- ocena rezultatów wyszukiwania,
- koniec sesji lub
- przeformułowanie zapytania i kontynuacja od pkt. 4. [Baeza-Yates, Ribeiro-Neto 1999, s. 263]

Przyjęte w systemach kryteria relewancji mają charakter systemowy i sprowadzają się do wyznaczenia w zbiorze wyszukiwawczym takiego podzbioru charakterystyk wyszukiwawczych dokumentów (lub tekstów dokumentów), w których treści wystąpiło wyrażenie równokształtne z zawartością instrukcji wyszukiwawczej. Elementy zbioru wyszukiwawczego spełniające taki warunek zgodności, są wyświetlane użytkownikowi jako odpowiedź systemu na jego zapytanie. Podstawową techniką, na której oparta jest realizacja wyszukiwania informacji jest tzw. dokładne dopasowanie (ang. exact matching) wykorzystujące algebrę Boole’a jako podstawowe działania na zmiennych dwuargumentowych. [Chu 2003, s. 59; Anderson, Perez-Carballo 2005] Wykorzystuje się tutaj następujące operatory:

- AND – operator zawężania wyszukiwania (suma logiczna),
- OR – operator poszerzania wyszukiwania (iloczyn logiczny),
- NOT – wyłączanie niepożądanych rezultatów (zaprzeczenie).

Tego rodzaju technika wyszukiwawcza dzieli zbiór wyszukiwawczy na dwa podzbiory, dokumentów spełniających warunki wyszukiwania, a więc relewantnych oraz dokumentów nirelewantnych. Każdy dokument zbioru relewantnego ma równy status,

zatem nie można w tym przypadku wprowadzać mechanizmów rankingowych dla rezultatów wyszukiwania, gdyż cecha relewantności nie jest stopniowalna.

Przedstawiona charakterystyka przyjmuje najprostszy sposób interpretacji funkcjonowania mechanizmów wyszukiwawczych w Internecie. Obecnie wiele wyszukiwarek internetowych stosuje zaawansowane operacje językowe na tekstach dokumentów internetowych oraz algorytmy zarówno selekcji jak i prezentacji rezultatów wyszukiwania (np. koncepcję zbiorów rozproszonych, współczynniki bliskości semantycznej, częściowe dopasowywanie, algorytmy rankingowe, itd.). Zastosowanie elementów sztucznej inteligencji w wyszukiwarkach internetowych oraz technik indeksowania zasobów WWW zwiększa efektywność wyszukiwania i jednocześnie ogranicza poziom wiedzy użytkownika o systemie wyszukiwawczym do niezbędnego minimum.

Zakres zbioru informacyjnego wyszukiwarki internetowej może mieć charakter globalny (indeksowanie całego zasobu WWW) lub może być ograniczony kryteriami formalnymi (np. indeksowanie kolekcji określonego systemu informacyjnego w Internecie) lub treściowymi (np. indeksowanie zasobów z danej dziedziny wiedzy).

Do najbardziej rozpowszechnionych obecnie globalnych wyszukiwarek internetowych należy niewątpliwie Google ([www.google.com](http://www.google.com)), ale również Excite ([www.excite.com](http://www.excite.com)), HotBot ([www.hotbot.com](http://www.hotbot.com)), Infoseek ([www.infoseek.com](http://www.infoseek.com)), czy też polski Netsprint ([www.netsprint.pl](http://www.netsprint.pl))

Dostęp do zasobów WWW realizowany przez wyszukiwarki internetowe z wykorzystaniem koncepcji punktów dostępu jest uzależniony od zasad konstrukcji zapytania informacyjnego oraz kryteriów wyszukiwawczych udostępnianych przez te mechanizmy. Jednocześnie przyjęty format opisu (najczęściej w formie niejawnej) oraz zasady indeksowania wyznaczają możliwości ich identyfikacji i aktywowania. Dostęp do zasobów realizowany jest najczęściej w trybie uproszczonym, gdzie użytkownik werbalizuje swoją potrzebę informacyjną do postaci ciągu słów kluczowych. Zaawansowani użytkownicy w tym trybie mogą również stosować najczęściej wykorzystywaną algebrę Boole'a jako narzędzie działania na zbiorach oraz identyfikatory dopuszczonych kryteriów wyszukiwawczych. Identyfikatory te są wykładnikami punktów dostępu. Np. wyszukiwarka Google dopuszcza w tym trybie formułowanie zapytania z identyfikatorami poszczególnych atrybutów dokumentów internetowych, których wartości są odszukiwane. Są to np.:

- typ dokumentu – identyfikator: *filetype*:szukana wartość,

- przeszukiwanie kolekcji zamieszczonej pod określonym adresem URL – szukana fraza, identyfikator:*site*:adres witryny,
- poszukiwanie odesłań z WWW do dokumentu internetowego – identyfikator:*link*:adres dokumentu internetowego.

W trybie zaawansowanym realizacja koncepcji punktów dostępu polega na wykorzystaniu do tego celu interfejsu użytkownika, w którym użytkownik może wybrać rodzaj cechy wyszukiwawczej, jaka będzie podstawą do przeprowadzenia wyszukiwania (Rys. 2.3).

Rys. 2.3. Punkty dostępu w interfejsie wyszukiwarki Google. [Źródło: <http://google.pl/>]

Przedstawiony na Rys. 2.3 interfejs zaawansowanego modułu mechanizmu wyszukiwawczego Google udostępnia możliwość dostępu do zbioru wyszukiwawczego z wykorzystaniem zestawu cech wyszukiwawczych. Są to m.in.:

- język dokumentu,
- zasięg geograficzny,
- typ dokumentu,
- data publikacji w WWW.

Cechy te pełnią funkcje punktów dostępu do bazy danych wyszukiwarki, które są aktywowane jako kryteria wyszukiwawcze.

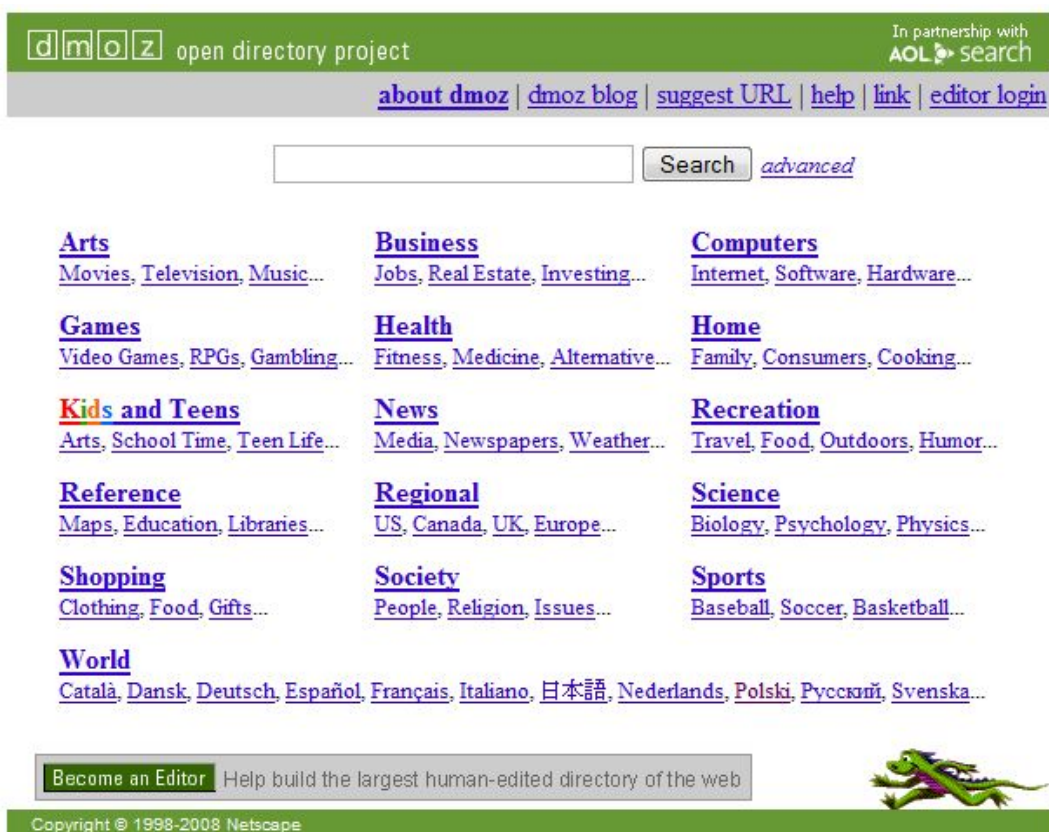
### 2.5.2. Katalogi internetowe

Katalogi internetowe zwane również drzewami przedmiotowymi lub przewodnikami dziedzinowymi [Ellis, Vasconcelos 2000, s. 100] są narzędziami dostępu do zasobów WWW wykorzystującymi ludzki kapitał intelektualny w selekcji i organizacji zasobów włączanych do ich kolekcji. W przeciwieństwie do wyszukiwarek internetowych, to człowiek jest najważniejszym ogniwem ich efektywnego funkcjonowania. Istota katalogów internetowych sprowadza się do selekcji zasobów WWW i włączania ich opisów w narzędzia dostępu przez kompetentne osoby. Zakres katalogu jest wyznaczany przez jego twórców i może mieć charakter ogólny (całe zasoby WWW np. katalog dmoz.org) lub być ograniczony kryteriami formalnymi (np. zasoby danego portalu internetowego – katalog portalu Onet.pl) lub treściowymi. Nie są to jedyne kryteria podziału katalogów internetowych, natomiast w każdym przypadku zbiór warunków koniecznych jest stały. Zalicza się do niego:

- selekcję włączanych zasobów WWW,
- zamieszczanie krótkiej charakterystyki każdego elementu zbioru,
- narzędzia dostępu w postaci:
  - o wyszukiwarki,
  - o oraz najczęściej hierarchicznego układu kategorii tematycznych, w których zgrupowano odesłania do zasobów WWW.

Struktura narzędzi dostępu w postaci układu kategorii tematycznych przypomina budowę katalogu systematycznego lub działowego, jednak w większości przypadków zasady ich budowy podyktowane są względami praktycznymi i często cały układ jest nielogiczny, a przez to niespójny. Źródłem do tworzenia tego rodzaju struktur są często klasyfikacje nauk lub piśmiennictwa, jednak wzorzec ten jest gruntownie modyfikowany do tego stopnia, że pozostaje z niego najczęściej wykaz klas na pierwszym stopniu podziału. Zagnieżdżenie takiego układu w strukturze hipertekstu pozwala użytkownikom na przemieszczanie się pomiędzy węzłami, których nazwy stanowią nazwy kategorii tematycznych (Rys. 2.4).





Rys. 2.4. Struktura katalogu internetowego dmoz.org. [Źródło: <http://www.dmoz.org/>]

Niewątpliwą zaletą katalogów internetowych jest ich zorganizowana struktura, która uznawana jest za ich wartość dodaną. [Ellis, Vasconcelos 2000, s. 100] Do poważnych wad zaliczyć można tutaj opóźnienie we włączaniu nowopojawiających się stron WWW i nastawienie na indeksowanie całych serwisów, a nie pojedynczych dokumentów internetowych. To z kolei przekłada się na stosowanie kategorii tematycznych o stosunkowo szerokim zakresie i nastawieniu na wyszukiwanie na tematy ogólne. [Ellis, Vasconcelos 2000, s. 101]

### 2.5.3. Biblioteki cyfrowe, repozytoria

Termin *biblioteka cyfrowa* (ang. digital library) jest często zastępowany w piśmiennictwie terminami *biblioteka elektroniczna* (ang. electronic library) lub *biblioteka wirtualna* (ang. virtual library). [Radwański 2002; Lesk 2005] Definiowanie biblioteki cyfrowej polega na wprowadzeniu procesów, jakie zachodzą w tradycyjnej bibliotece do środowiska elektronicznego. Mamy zatem do czynienia z elektronicznymi zasobami, elektronicznym opracowaniem, przechowywaniem oraz udostępnianiem. [Radwański 2002] Koncepcja biblioteki cyfrowej pojawiła się na długo przed technicznymi możliwościami jej realizacji. „W 1965 roku J.C.R. Licklider

stworzył pojęcie *biblioteki przyszłości*, odpowiadające jego wizji skomputeryzowanego księgozbioru, natomiast dziesięć lat później F. W. Lancaster pisał o *bibliotece bez papieru*". [Nahotko 2004b] Dynamiczny rozwój badań nad biblioteką cyfrową rozpoczął się na początku lat 90-tych ubiegłego stulecia. Wtedy to powołano w USA Digital Library Initiative, projekt, którego celem było stymulowanie i wyznaczanie kierunkowanie badań nad tym zagadnieniem. W 1994 roku odbyły się w USA pierwsze Warsztaty Bibliotek Cyfrowych (ang. Digital Library Workshop), gdzie m.in. próbowano opracować definicję terminu *biblioteka cyfrowa*. W Europie duży wpływ na rozwój wiedzy w tym zakresie miał projekt badawczy organizowany przez UKOLN (UK Office for Library & Information Networking) o nazwie *eLib Programme*. W rozwój prac nad bibliotekami cyfrowymi zaangażowały się również inne państwa zarówno z Europy jak i spoza niej. Były to m.in. Niemcy z programem GlobalInfo, Australia z m.in. World1, Electronic Reserve Project, Nowa Zelandia z New Zeland Digital Library Project, czy też Japonia z Digital Libraries Project in Japan. [Chowdhury 1999]

Udostępnienie kolekcji bibliotek cyfrowych w Internecie otworzyło nowe możliwości powiększania jego zasobów oraz ustanowiło nowe narzędzie przechowywania, organizacji i dostępu do informacji. Obecnie biblioteki cyfrowe funkcjonują w Internecie zgodnie z wypracowanymi standardami dotyczącymi zarządzania kolekcją, pozyskiwania, katalogowania oraz udostępniania zbiorów. Kierowanie się tego rodzaju wytycznymi zapewnia wysoką jakość świadczonych przez nie usług informacyjnych. W kontekście rozważań nad narzędziami dostępu do zasobów WWW stanowią one z pewnością wartościowe narzędzie wyszukiwawcze oraz źródło informacji. W zależności od rodzaju informacji udostępnianej użytkownikowi na wyjściu można zidentyfikować biblioteki cyfrowe udostępniające:

- cyfrowe odpowiedniki dokumentów tradycyjnych,
- dokumenty, które powstały jako cyfrowe (tzw. born digital),
- informacje o dokumentach cyfrowych wraz z ich lokalizacją w Internecie.

W pierwszym i drugim przypadku może dochodzić do łączenia różnych rodzajów zasobów, natomiast w trzecim, tego typu biblioteki udostępniają wyłącznie metainformacje.

Biblioteki cyfrowe udostępniające dokumenty tradycyjne w cyfrowej postaci włączają się w ich obieg w nowym środowisku informacyjnym, jakim jest Internet. Udostępnianie tego rodzaju zbiorów wiąże się zarówno z rozwiązaniem kwestii technicznego przekształcenia nośnika, jak i z rozstrzygnięciem własności praw

autorskich dla poszczególnych dzieł. Przenoszenie dokumentów tradycyjnych do postaci cyfrowej, niezależnie od sposobu utrwalenia (dokument piśmienniczy, audialny, wizualny, audiowizualny, graficzny) oraz ich udostępnianie w zasobach WWW to olbrzymi wkład tego rodzaju narzędzi zarówno w rozwój nauki (poprzez m.in. upowszechnianie dostępu) jak i w tworzenie zasobu światowego (poprzez archiwizację). Istnieje wiele bibliotek cyfrowych, które restrykcyjnie podchodzą do udostępniania swoich zasobów internautom. Wiąże się to przede wszystkim z koniecznością poniesienia przez użytkownika kosztów dostępu. Z drugiej strony, wiele instytucji, w tym biblioteki, prowadzi biblioteki cyfrowe z wolnym dostępem. Do ważnych inicjatyw związanych z bibliotekami cyfrowymi udostępniającymi cyfrowe wersje dokumentów tradycyjnych należy niewątpliwie „Projekt Gutenberg” ([http://www.gutenberg.org/wiki/Main\\_Page](http://www.gutenberg.org/wiki/Main_Page)). Projekt powstał na początku lat 70-tych ubiegłego stulecia i jego celem jest umieszczanie w Internecie elektronicznych wersji dokumentów tradycyjnych, dla których wygasły majątkowe prawa autorskie.

W Polsce przeważająca liczba bibliotek cyfrowych funkcjonuje na platformie programistycznej dLibra. Istniejąca Federacja Bibliotek Cyfrowych udostępnia mechanizm przeszukiwania polskich bibliotek pracujących w tym systemie.

Specyficznym rodzajem bibliotek cyfrowych są elektroniczne repozytoria. Udostępniają dokumenty elektroniczne zdeponowane tam przez użytkowników. Idea repozytoriów wpisuje się w ruch Open Access, którego celem jest upowszechnienie wolnego dostępu do publikacji elektronicznych w Internecie. Autorzy dokumentów deponują je w repozytorium do otwartego dostępu dla innych użytkowników, wykorzystując swoje prawa do dystrybucji reprintów, postprintów, czy też innych form rozpowszechniania. Zasób repozytorium powiększają również instytucje dysponujące prawami autorskimi do poszczególnych dokumentów. Istniejące w Internecie wykazy repozytoriów elektronicznych ułatwiają odnalezienie konkretnej np. tematycznej kolekcji. Należą do nich m.in. OpenDoar (<http://www.opendoar.org/>), OAIster (<http://oaister.umdl.umich.edu/o/oaister/>), Institutional Archives Registry (<http://archives.eprints.org/>). Do repozytoriów z zakresu bibliotekoznawstwa i informacji naukowej należą m.in. E-LIS (E-prints in Library & Information Science - <http://eprints.rclis.org/>) oraz D-LIST (Digital Library of Information Science & Technology - <http://dlist.sir.arizona.edu/>).

Kolejnym typem bibliotek cyfrowych są kolekcje, których elementami są charakterystyki obiektów informacyjnych w Internecie. Tego rodzaju systemy udostępniają metainformacje w postaci opisu dokumentu internetowego

o zróżnicowanym stopniu szczegółowości. Ich celem jest udostępnianie informacji o wyselekcjonowanych zasobach WWW. Nazywane są często bibliotekami wirtualnymi z uwagi na rodzaj informacji wyjściowej udostępnianej użytkownikowi. Ich celem jest poinformowanie oraz skierowanie użytkownika do konkretnej lokalizacji dokumentu w Internecie. Ich zakres może być szeroki i obejmować jak wiele dziedzin wiedzy (np. The WWW Virtual Library) lub zostać ograniczony do jednej dziedziny lub obszaru tematycznego (np. MathGuide.de – z zakresu matematyki). Tego rodzaju narzędzia dostępu do zasobów WWW prezentują różne podejście zarówno do szczegółowości charakterystyki zasobów Internetu, jak i do ilości metainformacji wyświetlanych użytkownikowi. W najprostszym przypadku sprowadza się to do wykazu adresów dokumentów internetowych spełniających określone kryteria włączenia do kolekcji. W takiej wersji serwisy te nazywane są sieciowymi bibliografiami, tzw. webografiami. [Jachimczyk 2008] Najbardziej zaawansowane w kontekście katalogowania zasobów Internetu oraz organizacji dostępu do kolekcji prezentują dziedzinowe systemy hipertekstowe, zwane również dziedzinowymi bramami tematycznymi, które są przedmiotem rozprawy.

#### **2.5.4. Folksonomie**

Taksonomie tworzone przez użytkowników lub tzw. folksonomie (ang. folksonomies) są współdzielonym narzędziem katalogowania dokumentów graficznych, dźwiękowych, audiowizualnych, hipertekstowych oraz tradycyjnych. [Smith 2007] Są jedną z realizacji koncepcji Web 2.0, która jest obecnie bardzo często poruszonym zagadnieniem związanym ze współczesnymi systemami i serwisami funkcjonującymi w Internecie. Web 2.0 nie jest nową wersją Internetu, lecz innym sposobem interpretacji interakcji pomiędzy użytkownikiem oraz zasobami World Wide Web oraz usługami oferowanymi przez serwisy tam funkcjonujące. Możliwość wpływania przez użytkownika na ilość i rodzaj udostępnianych mu informacji przez aplikacje internetowe zrewolucjonizowała dotychczasowe postrzeganie jego roli, jako biernego konsumenta informacji zgromadzonych w sieci. Web 2.0 przejawia się w aktywności poszczególnego użytkownika, ale również społeczności jaką tworzą grupy osób połączonych wspólnymi potrzebami informacyjnymi, celami, poglądami, zainteresowaniami, itd. [Roszkowski 2007b] Dlatego też Tim O'Reilly charakteryzując to zjawisko używa m.in. terminów interakcja i integracja. [O'Reilly 2005] Do

podstawowych cech serwisów i systemów internetowych opracowanych według koncepcji Web 2.0 należą:

- uwalnianie danych: umożliwianie ich odnajdywania i manipulowania niezależnie od ich pierwotnego przeznaczenia. Wiąże się to pośrednio z ideą Open Access, ale przede wszystkim dotyczy wolnego dostępu do informacji;
- tworzenie wirtualnych aplikacji: udostępnienie wolnego oprogramowania do tworzenia indywidualnych aplikacji przez użytkowników i zamieszczania na ich własnych serwisach (np. Google Maps);
- partycypacja: aktywny udział użytkowników sieci w tworzeniu jej zawartości (np. blogi internetowe, ale również serwisy społecznościowe);
- nastawienie na bezpośrednie funkcjonowanie dla użytkownika: np. umożliwienie tworzenia indywidualnych profili użytkowników i selektywne dystrybuowanie informacji; dostosowanie narzędzi internetowych do zaspokajania zindywidualizowanych potrzeb informacyjnych (np. wyszukiwanie najkrótszych/najszybszych tras przejazdu samochodem z jednego miejsca do drugiego);
- wspólne wykorzystywanie treści, idei i rozwiązań technicznych: oznacza to nowy model ekonomiczny, który przynosi zyski dzięki wspólnemu wysiłkowi stanowiącemu wartość dodaną projektu;
- nastawienie na komunikację, gdzie następuje sprzężenie zwrotne i pełna aktywizacja nadawców (twórców treści) oraz odbiorców w postaci jawnej wymiany informacji;
- łączenie wielu narzędzi pełniących różne funkcje w kompleksowe platformy informacyjne;
- wykorzystanie sztucznej inteligencji do gromadzenia i przetwarzania danych;
- zasada tzw. długiego ogona (ang. long tail): specyfikacja, rozdrobnienie i maksymalizacja oferty (również informacyjnej) skierowanej do małych grup użytkowników o wąskim zakresie zainteresowań i potrzeb informacyjnych;
- zaufanie: przede wszystkim użytkowników do innych użytkowników jako twórców treści oraz w sposobach wykorzystania informacji. [Miller 2005]

Folksonomia jest narzędziem angażującym użytkownika do aktywności intelektualnej, której celem jest charakterystyka tradycyjnych lub internetowych obiektów cyfrowych do postaci ciągu niekontrolowanych słów kluczowych reprezentujących m.in. ich treść, formę lub stosunek katalogującego do nich. Taka charakterystyka udostępniana jest poprzez narzędzie dostępu do zasobów folksonomii,

na które składa się mechanizm wyszukiwawczy oraz, co jest charakterystyczne, mechanizm dostępu w postaci indeksu alfabetycznego użytych słów kluczowych zorganizowanych w strukturze hipertekstowej do postaci tzw. chmury tagów (Rys. 2.5).

The screenshot shows the CiteULike website interface. At the top, there are navigation links for Register, Log in, and FAQ, along with flags for various countries. Below the header, there is a search bar and a 'Search' button. The main content area is titled 'A free online service to organise your academic papers' and lists several recent papers. A red box highlights one of the papers, and a red arrow points from it to the tag cloud on the right. The tag cloud lists various tags such as 'adaptation algorithm', 'bayesian bioinformatics', 'biology book brain cancer', etc.

**OXFORD JOURNALS**  
READ THE LATEST RESEARCH TO HIT THE HEADLINES  
Sponsored link

**CiteULike**

- News
- Discussion

**Journals**

- Browse current issues

**Groups**

- Search groups

Ads by Google

**Real-Time PCR**  
Wide range of instrumentation.

**A free online service to organise your academic papers**  
Some recent papers posted to CiteULike - all mixed together.

- Antioxidant enzymes activity and lipid peroxidation in liver and kidney of rats exposed to cadmium and ethanol.**  
*Food Chem Toxicol*, Vol. 42, No. 3. (March 2004), pp. 429-438.  
by M Jurczuk, MM Brzóska, J Moniuszko-Jakoniuk, M Gałazyn-Sidorczuk, E Kulikowska-Karpińska  
posted to [no-tag](#) by [jvuh](#) on 2008-03-04 09:03:44 as ★★
- Discrepancy between the nephrotoxic potencies of cadmium-metallothionein and cadmium chloride and the renal concentration of cadmium in the proximal convoluted tubules.**  
*Toxicol Appl Pharmacol*, Vol. 130, No. 1. (January 1995), pp. 161-168.  
by C Doran, VH Gattone, CD Klaassen  
posted to [no-tag](#) by [jvuh](#) on 2008-03-04 09:03:37 as ★★
- Experimental model of lead nephropathy. I. Continuous high-dose lead administration.**  
*Kidney Int*, Vol. 41, No. 5. (May 1992), pp. 1192-1203.  
by F Khalil-Manesh, HC Gonick, AH Cohen, R Alinovi, E Bergamaschi, A Multti, VJ Rosen  
posted to [no-tag](#) by [jvuh](#) on 2008-03-04 09:03:31 as ★★
- Cadmium-induced nephropathy in rats is mediated by expression of senescence-associated beta-galactosidase and accumulation of mitochondrial DNA deletion.**  
*Ann N Y Acad Sci*, Vol. 1011 (April 2004), pp. 332-338.  
by A Takaki, S Jimi, M Segawa, H Iwasaki

**Everyone's Tags**  
Most active tags on CiteULike

Filter:

adaptation algorithm  
algorithms analysis and bayesian bioinformatics  
biology book brain cancer  
cell classification  
clustering coding  
cognition collaboration  
communication community complexity control culture  
data database design  
detection development  
disease distributed diversity  
dna drosophila dynamics  
ecology education  
evaluation evolution  
expression fmri gene  
genetics genome graph  
health history human  
imaging information

Rys. 2.5. Chmura tagów w serwisie CiteULike. [Źródło: <http://www.citeulike.org/>]

Każde ze słów kluczowych posiada status węzła hipertekstowego, którego aktywacja powoduje wyodrębnienie podzbioru charakterystyk obiektów cyfrowych, w których je użyto do opisu. Użytkownicy wysyłając charakterystykę obiektu powiększają jednocześnie zbiór informacyjny danej folksonomii. Zarówno dostęp do zasobów folksonomii jak i wprowadzanie nowych charakterystyk zazwyczaj nie jest ograniczone dodatkowymi barierami, czy restrykcjami.

Zakresy folksonomii internetowych są zróżnicowane. Mają one charakter globalny, a więc ich zakres jest bardzo szeroki i potencjalnie angażują każdego internautę, lub wyznaczony kryteriami treściowymi lub formalnymi. Do najpopularniejszych folksonomii internetowych należą:

- del.icio.us (<http://del.icio.us>) - umieszczanie adresów ciekawych i wartościowych stron internetowych wraz z ich opisami, m.in. poprzez słowa kluczowe,

- flickr (<http://www.flickr.com>) – album ze zdjęciami cyfrowymi deponowanymi i opisywanymi przez internautów.

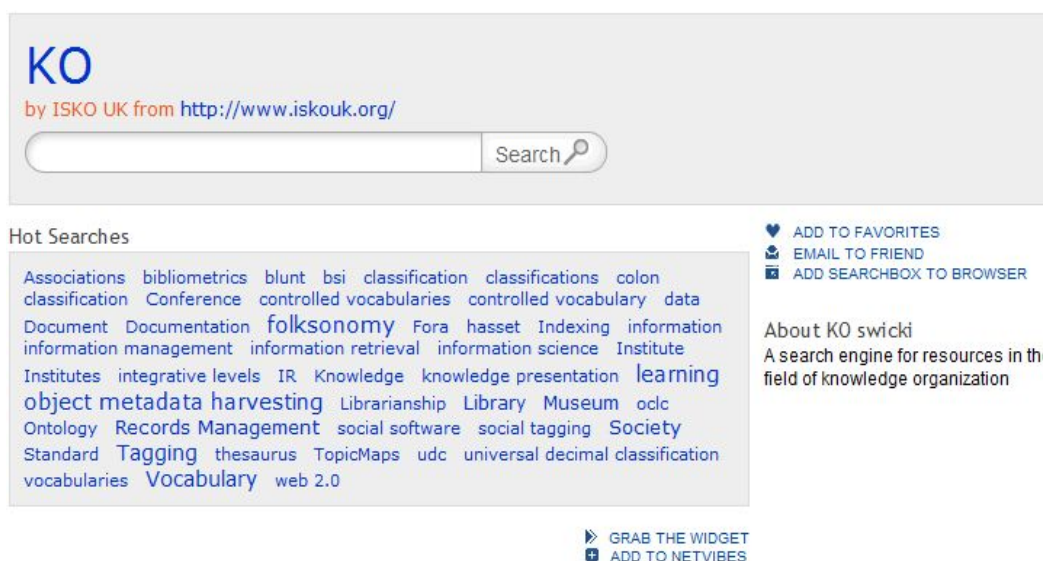
Folksonomie internetowe funkcjonują jako autonomiczne systemy informacyjne oraz jako mechanizmy organizacji i wyszukiwania informacji w większych serwisach i systemach internetowych. Społeczne klasyfikowanie obiektów cyfrowych jest pożądaną cechą internetowych usług informacyjnych realizujących koncepcję Web 2.0. Duża liczba serwisów oraz np. księgarń internetowych (np. Empik.com) daje swoim użytkownikom możliwość wyrażenia opinii o jakości i treści oferowanych produktów w postaci tekstów zbudowanych z ciągu słów kluczowych. Do tego nurtu zaliczyć można również modernizację katalogów OPAC, niestety przede wszystkim bibliotek amerykańskich. Tutaj możliwość opisanie jednostki katalogowej przez użytkownika stanowi wartość dodaną całego opisu. Mamy wtedy do czynienia z pewnego rodzaju polireprezentacją wiedzy. Z jednej strony zobiektywizowany i sformalizowany opis bibliograficzny wraz z charakterystyką treściową wyrażoną w jednym lub kilku językach informacyjnych, z drugiej ciąg słów kluczowych jako efekt postrzegania dokumentu przez użytkownika.

Rola folksonomii jako narzędzia organizacji i dostępu do zasobów WWW polega na aktywizacji internautów do tworzenia jej zbioru informacyjnego i generowania narzędzi dostępu. Analiza leksykograficzna zbioru słów kluczowych stanowiących tzw. zestaw tagów przeprowadzona przez L. Spitieri [Spitieri 2007] w kilku największych tego rodzaju systemów wskazuje m.in. na:

- ich duże nacechowanie emocjonalne,
- subiektywizm w reprezentacji cech treści i formy elementów kolekcji,
- niekonsekwencję stosowania określonej formy wyrażenia.

Subiektywizm i niespójność formalna charakterystyki obiektów informacyjnych w folksonomii nie jest przeszkodą do uznania jej jako wartościowego źródła informacji. Zasoby tego rodzaju usług informacyjnych są tworzone przez użytkowników, więc ich selekcja ma miejsce na poziomie użytkownika, czyli „elementu” systemu najbardziej zainteresowanego wysoką jakością udostępnianych mu informacji. Bezpośredniemu zaangażowaniu użytkownika w tworzenie zasobu informacyjnego internetowego systemu społecznego klasyfikowania, szczególnie o sprecyzowanym zakresie tematycznym, towarzyszy także poczucie odpowiedzialności za własny wkład w taką kolekcję. Tematyczne systemy społecznościowe w Internecie wykorzystujące mechanizmy folksonomii do organizacji i udostępniania informacji są ważnym źródłem informacji o jego zasobach. Do tej grupy należy serwis społecznościowy KO-SWICKI

(<http://ko-swicki.eurekster.com/>), który powstał w ramach brytyjskiego oddziału International Society for Knowledge Organization (<http://www.iskouk.org>) (Rys. 2.6)



Rys. 2.6. Interfejs serwisu KO-SWICKI. [Źródło: <http://ko-swicki.eurekster.com/>]

Serwis ten wykorzystuje mechanizm społecznej klasyfikacji zasobów WWW z zakresu szeroko pojętej organizacji wiedzy. Użytkownicy tworzą jego zasób informacyjny poprzez wysyłanie charakterystyk stron WWW, serwisów informacyjnych, dokumentów cyfrowych, których tematyka należy do wspomnianego zakresu. Każda charakterystyka zawiera wykaz słów kluczowych reprezentujących cechy treści lub formy obiektów informacyjnych. Na ich podstawie generowany jest wykaz tagów zagnieżdżony w strukturze hipertekstowej, które pełnią funkcję punktów dostępu do kolekcji.

Przedstawiona charakterystyka narzędzi organizacji i dostępu do zasobów WWW jest w założeniu niekompletna. Spowodowane jest to immanentną cechą tego środowiska oraz narzędzi w nim funkcjonujących, czyli dynamiką i zmiennością. Nowe koncepcje narzędzi internetowych, jak np. Web 2.0 permanentnie ewoluują pozostawiając wdrożenia nowych mechanizmów wyszukiwawczych. Przytoczony w tej części pracy opis rodzajów narzędzi miał na celu reprezentację jedynie głównych mechanizmów dostępu do zasobów Internetu.

## ***2.6. Tryby dostępu do zasobów w internetowych systemach informacyjnych***

Przedstawiona na początku rozdziału koncepcja architektury informacji, jako płaszczyzny interpretacji organizacji informacji w internetowych systemach informacyjnych, zakłada przynajmniej dwa sposoby pozyskiwania informacji



z zasobów systemu. Pierwszy z nich opiera się na zastosowaniu mechanizmów wyszukiwawczych wykorzystujących proces dopasowywania zapytania użytkownika z odpowiednią charakterystyką (lub fragmentem tekstu dokumentu) w bazie danych. Algorytm procesu wyszukiwania za pomocą takiego mechanizmu jest zróżnicowany i zależy m.in. od przyjętych technik indeksowania oraz kryteriów wyszukiwawczych. Standardowy, uproszczony przebieg tego procesu przedstawiono w podrozdziale dotyczącym wyszukiwarek internetowych (Rozdz. 2.5.1.). W tym przypadku mechanizmy wyszukiwawcze działają na poziomie lokalnym, a więc generują zbiór wyszukiwawczy ze zbioru informacyjnego danego systemu.

Obok sformalizowanych, analitycznych strategii realizowanych podczas pracy z modułem wyszukiwania informacji w systemach informacyjnych WWW, użytkownicy stosują również strategie nieformalne, heurystyczne. Ich cel jest niezmienny, czyli pozyskanie pożądanych informacji, jednak środki realizacji zależą od uwarunkowań środowiska informacyjnego (w tym przypadku od oferowanych przez system narzędzi). Ten sposób pozyskiwania informacji z systemu nazywany jest *dostępem przedmiotowym* (ang. subject access). Charakterystyczną cechą narzędzi dostępu przedmiotowego jest udostępnienie użytkownikowi informacji o zasobach systemu, które poddane odpowiednim procesom transformacji stanowią punkty dostępu do jego zbioru wyszukiwawczego. W odniesieniu do systemów hipertekstowych, narzędzie dostępu przedmiotowego tworzy wykaz kategorii tematycznych reprezentujących wymiar treści, ale również formy elementów zbioru informacyjnego. Są one włączane w określone struktury, co implikuje specyficzną możliwość ich eksploracji przez użytkownika. Zagnieżdżenie ich w środowisku hipertekstowym wyznacza możliwości procesu nawigacji, którego celem jest identyfikacja przez użytkownika wyrażenia lub wyrażenia, których zakres znaczeniowy reprezentuje temat jego poszukiwań. Każde wyrażenie jest jednocześnie nazwą dla kategorii tematycznej, która grupuje odpowiednie odesłania do elementów zbioru informacyjnego. Podczas procesu indeksowania dokumentów internetowych odwzorowuje się ich cechy wymiaru treści i formy. Wartości tych cech stanowią podstawę do budowy określonych struktur porządkujących i organizujących dostęp do kolekcji. Tym samym wyrażenia te pełnią funkcje metainformacyjne i wyszukiwawcze. Przykładem realizacji tego rodzaju dostępu jest mechanizm nawigacji np. w katalogach internetowych czy bibliotekach wirtualnych. Podstawową techniką pozyskiwania informacji z użyciem tego rodzaju narzędzi jest tzw. *przeglądanie* (ang. browsing).

Znaczenie terminu *dostęp przedmiotowy* należy rozpatrywać w kontekście

techniki przeglądania (ang. browsing), która oznacza poszukiwanie informacji przez skanowanie, nawigację i inne podobne aktywności. [Chu 2003, s. 87] Definicja tej techniki autorstwa L. Willa brzmi:

*Przeglądanie oznacza odnajdywanie informacji poprzez analizowanie listy lub sekwencji jednostek, zazwyczaj rozpoczynając od ogólnych i na podstawie informacji tam zawartych podążając w stronę bardziej szczegółowych.* [Will 2006]

Przez wspomniane w definicji jednostki można rozumieć obiekty informacyjne, które mogą mieć formę:

- pełnych tekstów,
- fragmentów dokumentów,
- ich reprezentacji w postaci charakterystyk wyszukiwawczych (o różnym stopniu szczegółowości),
- jednostek schematów kategoryjnych organizujących dokumenty,
- elementów struktury typograficznej dokumentu (np. nagłówki, tytuły itp.).

Technika przeglądania jest przedmiotem badań naukowców z różnych dyscyplin naukowych. Chociaż duże zainteresowanie tym zagadnieniem wykazują specjaliści z zakresu psychologii, w tym ekologii behawioralnej [Wolfe 1994; Logan 2003], to przedmiotem rozważań nad tą techniką będą w dalszej części badania prowadzone w zakresie nauki o informacji.

Wyniki badań prowadzonych przez zespół badawczy pod kierunkiem R. Rice'a [Rice, McCreddie, Chang 2001] wskazują na możliwość identyfikacji czterech wymiarów analizy tej techniki. Są to:

1. Akt skanowania - przeglądanie zakłada skanowanie, które jest charakteryzowane jako oglądanie, analizowanie lub próbkowanie.
2. Obecność lub brak celu - działanie może być zorientowane na cel lub też niezorientowane.
3. Kryteria wyszukiwawcze - mogą być sprecyzowane lub niesprecyzowane.
4. Wiedza użytkownika na temat przeglądanych zasobów.

Pojawiający się tutaj termin skanowanie (ang. scanning), rozumiany jest jako szybkie lecz systematyczne i badawcze przyglądanie się. [Bates 2007] Jest on jednym z podstawowych procesów zachodzących w ramach techniki przeglądania.

Zdaniem Marcii Bates przeglądanie jest złożone z serii przelotnych spojrzeń, mignięć (ang. glipmses), z których jedne prowadzą, czy też wywołują kolejne, inne zaś prowadzą do dokładniejszej eksploracji obiektów lub ich porzucenia. W tym znaczeniu

przeglądanie składa się z określonej liczby startów i zatrzymań oraz towarzyszących im procesów m.in. czytania, próbkowania, selekcji. [Bates 2007] W zależności od stopnia rozpoznania potrzeby informacyjnej przez użytkownika, realizacja techniki przeglądania może przybierać formę:

- zorientowaną (ang. directed) – kiedy użytkownik posiada jasno określoną i zwerbalizowaną potrzebę informacyjną,
- pół-zorientowaną (ang. semi-directed) – eksplorowanie zbioru wyszukiwawczego ze słabo sprecyzowaną potrzebą informacyjną,
- niezorientowaną (ang. undirected) – przeglądanie realizowane bez konkretnego celu i świadomości potrzeby informacyjnej. [Marchionini 1995; Kowalski 1997 za Chu 2003, s. 88]

Inną perspektywę w typologii przeglądania przyjął Gerard Kowalski [Chu 2003, s. 89 za Kowalski 1997], który wyróżnił kolejne pięć rodzajów realizacji tej techniki poszukiwania informacji. Były to:

- Dostęp z wykorzystaniem mechanizmów rankingowych (ang. browse by ranking). Przeglądanie przez użytkowników obiektów uporządkowanych według np. malejącego współczynnika relewancji.
- Dostęp strefowy (ang. browse by zone) – przeglądanie zawartości wybranych pól lub podpól (np. autor, tytuł, opis rzeczowy) w celu określenia potencjalnej przydatności dokumentu.
- Dostęp poprzez wyróżnione elementy (ang. browse by highlighted area). W niektórych systemach informacyjno-wyszukiwawczych wyróżnia się graficznie np. w rezultatach wyszukiwania pewne elementy, w tym przypadku terminy pochodzące z zapytania, aby dać użytkownikowi pewien kontekst w dalszym poszukiwaniu informacji.
- Dostęp kategoryalny (ang. browse by category), którego egzemplifikacją jest narzędzie w postaci układu hierarchicznego kategorii tematycznych. Jednostki leksykalne przyjęte jako nazwy kategorii w układzie hierarchicznym dzielą zbiór wyszukiwawczy na klasy dokumentów.
- Dostęp z wykorzystaniem hiperłączy (ang. browse by hyperlinks) polega na wykorzystaniu węzłów oraz powiązań hipertekstowych w eksplorowaniu przestrzeni informacyjnej. H. Chu uważa, że „hipermedialne systemy informacyjno-wyszukiwawcze są jednym z najlepszych, jeżeli nie najlepszym, środowiskiem informacyjnym dla realizacji tej techniki”. [Chu 2003, s. 89]

Zdaniem wielu badaczy [Kwasnik 1992; Chang, Rice 1993; Marchionini 1995; Kowalski 1997; Bates 2002; Chu 2003; Sciliabi in. 2003] obydwie sposoby pozyskiwania informacji, czyli wyszukiwanie za pomocą wyszukiwarek i narzędzi dostępu przedmiotowego, należy traktować komplementarnie. M. Bates szczególnie podkreśla niezrozumiałą dychotomię i traktowanie przeglądania w opozycji do wyszukiwania informacji, która funkcjonuje w środowisku pracowników informacji naukowej. [Bates 1989]

Pomimo różnic wynikających np. ze sposobu realizacji tych procesów ich cel pozostaje wspólny, zaspokojenie potrzeby informacyjnej poprzez odnalezienie relewantnych informacji. H. Chu [Chu 2003, s 91-93] na podstawie pięciu zmiennych dokonał charakterystyki wyszukiwania za pomocą wyszukiwarek oraz narzędzi dostępu przedmiotowego. Zaliczył do nich:

- **Potrzebę informacyjną.** W przypadku jasno sprecyzowanej potrzeby informacyjnej, zastosowanie wyszukiwarki wydaje się być właściwym rozwiązaniem. Przeglądanie jest efektywne, kiedy użytkownik posiada szerokie i niesprecyzowane potrzeby informacyjne, lub kiedy chce sprawdzić czy w zbiorze informacyjnym znajdują się jakiegokolwiek relewantne dokumenty.
- **Wydajność.** Wyszukiwanie za pomocą współczesnych wyszukiwarek przebiega szybko. System natychmiast wysyła odpowiedź w postaci rezultatów wyszukiwania. Przeglądanie jest czasochłonne i w przypadku nieskutecznych narzędzi dostępu może spowodować u użytkownika uczucie zagubienia w przestrzeni informacyjnej lub przeładowania informacyjnego. Poza tym użytkownik w danej chwili ma dostęp nie do całego zbioru wyszukiwawczego, lecz do jego części aktualnie eksplorowanej.
- **Obciążenie kognitywne.** Proces wyszukiwania informacji można podzielić na trzy główne etapy: reprezentacja zapytania, przeprowadzenie wyszukiwania, ocena rezultatów. W przypadku użycia wyszukiwarki obciążenie kognitywne użytkownika jest wysokie podczas etapu pierwszego i trzeciego. Przeglądanie jest wysoce interaktywne i jego efektywność zależy przede wszystkim od zaangażowania użytkownika podczas całego procesu pozyskiwania informacji i reakcji na odpowiedzi systemu.
- **Przypadkowe odkrywanie informacji** (ang. serendipity) . Sytuacja, w której użytkownik przypadkowo odnajdzie użyteczne, lecz aktualnie nieposzukiwane

informacje podczas interakcji z mechanizmem wyszukiwawczym jest teoretycznie niemożliwa. Ten proces oparty jest bowiem (w większości przypadków) na kryteriach relewancji systemowej, czyli na porównywaniu zawartości instrukcji wyszukiwawczej z odpowiednimi polami w charakterystyce wyszukiwawczej i zwracaniu dokumentów, gdzie takie podobieństwo (na różnych poziomach szczegółowości) istnieje. Przeglądanie dopuszcza prawdopodobieństwo odnalezienia takich informacji, na co ma wpływ szerokie wykorzystanie informacji kontekstowych.

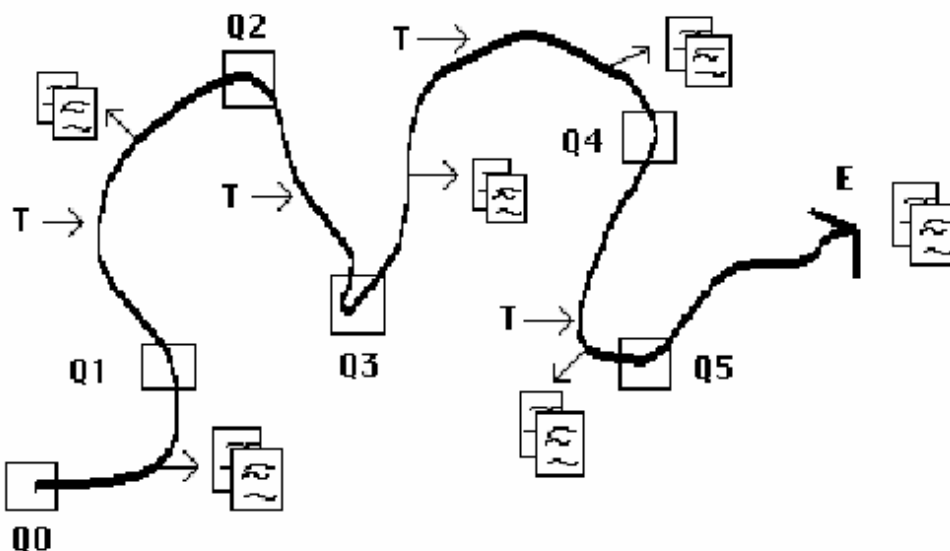
- **Wymagania.** Wyszukiwanie informacji za pomocą wyszukiwarek wymaga od użytkownika znajomości technik stosowanych w systemie (np. formułowania instrukcji wyszukiwawczej), czyli pewnego zasobu wiedzy o systemie. Przeglądanie nie jest uzależnione w takim stopniu od tego typu wymagań.

W Tab. 2.2. przedstawiono podsumowanie dokonanego porównania wyszukiwania informacji oraz przeglądania.

Tab. 2.2. Porównanie wyszukiwania i przeglądania. [Źródło: Chu 2003, s. 93]

Aspekt / Podejście	Potrzeba informacyjna	Wydajność	Obciążenie kognitywne	Szczęśliwy traf	Wymagania
<b>Wyszukiwanie</b>	Szczegółowa i znana	Wysoka	Małe	Rzadko	Duże
<b>Przeglądanie</b>	Szeroka i niepewna	Niska	Duże	Często	Małe

Technika przeglądania może nasuwać skojarzenia z metodą pozyskiwania informacji opracowaną przez M. Bates pod nazwą *berrypicking* (zbieranie jagód) [Bates 1989] (Rys. 2.7.).



Rys. 2.7. Technika ewoluującego zapytania informacyjnego - berry picking M. Bates [Źródło: Bates 1989]

Jak zauważa sama autorka [Bates 2007] jej metoda charakteryzuje poszczególne „epizody” podczas wyszukiwania informacji, w tym również technikę przeglądania. Zatem metoda ta, charakteryzująca się specyficznym procesem ewoluowania zapytania (ang. *evolving query*), obejmuje również opisywaną technikę. [Hildreth 1995]

Przedstawiona krótka charakterystyka tej niewątpliwie złożonej techniki pozyskiwania informacji, jaką jest przeglądanie miała na celu jej umiejscowienie w szerszym kontekście rozważań nad narzędziami dostępu przedmiotowego i wskazanie na podstawowe procesy, jakie zachodzą podczas jej realizacji przez użytkownika.

Implementacja narzędzi dostępu przedmiotowego w internetowych systemach informacyjnych polega na identyfikacji i organizacji węzłów hipertekstowych pełniących funkcje wyszukiwawcze. Status węzła pełni najczęściej jednoelementowe wyrażenia języka naturalnego przyjęte do reprezentacji wymiaru treści i/lub formy obiektów informacyjnych w kolekcji. Przypisanie im funkcji punktów dostępu wyznacza zakres reprezentowanych przez nie cech wyszukiwawczych. Organizacja tych punktów dostępu przyjmuje formę od prostych struktur monorelacyjnych (np. alfabetyczne wykazy słów kluczowych), przez struktury hierarchiczne (np. struktura hierarchiczna kategorii tematycznych w katalogach internetowych), aż po zaawansowane struktury wykorzystujące fasetową organizację pola tematycznego\* lub

\* Pole tematyczne systemu informacyjno-wyszukiwawczego, zakres systemu informacyjno-wyszukiwawczego - cecha systemu informacyjno-wyszukiwawczego wyznaczana przez sumę logiczną tematów dokumentów zawierających informację relewantną dla użytkowników systemu informacyjno-wyszukiwawczego, tworzących zbiór informacyjny systemu [...]. [Słownik encyklopedyczny informacji..., s. 197]

mechanizmy oparte na ontologiach internetowych. Do budowy tego rodzaju narzędzi wykorzystuje się często języki informacyjno-wyszukiwawcze, traktując ich słowniki jako źródło leksyki dla nazw kategorii tematycznych, a także wykorzystując organizację ich pól semantycznych.

## ***2.7. Rola języka informacyjno-wyszukiwawczego w organizacji informacji w World Wide Web***

Rola języka informacyjno-wyszukiwawczego w organizacji informacji wynika z realizacji jego podstawowych funkcji, jakie pełni w systemie informacyjnym. Te, jak wspomniano, sytuują się w kontekście:

- ciągu operacji wprowadzania informacji do zbioru, analizując treść i tworząc jej skondensowaną reprezentację z wykorzystaniem reguł i środków JIW,
- ciągu operacji wyszukiwania dokumentów, dyskursie polegającym na wprowadzaniu opracowanego ciągu znaków JIW i komend wyszukiwawczych oraz sterowaniu wprowadzonymi znakami JIW poprzez porównanie ich z ze znakami składającymi się na ChWD i wydawaniu tych dokumentów, które na podstawie przyjętego kryterium relewancji będą odpowiedzią na zapytanie użytkownika [Sosińska 1987, s. 41],
- odsyłania bezpośredniego i pośredniego jednostek leksykalnych JIW do dokumentów pierwotnych lub pochodnych w zbiorze informacyjnym.

Przyjmując koncepcję punktów dostępu jako perspektywę dla interpretacji organizacji informacji, rola języka informacyjno-wyszukiwawczego może być rozpatrywana w kontekście:

- identyfikacji punktów dostępu,
- przypisywania określonych wartości punktom dostępu,
- organizacji punktów dostępu.

### **2.7.1. Identyfikacja punktów dostępu**

Zastosowany w systemie informacyjnym określony język opisu bibliograficznego determinuje zestaw cech dla każdego elementu zbioru informacyjnego, których wartości są określane na etapie indeksowania. Zdolność tego typu języków informacyjnych do identyfikacji relewantnych atrybutów dokumentów

ma bezpośredni wpływ na możliwości wyszukiwania informacji w systemie. Stopień szczegółowości formatu opisu musi być dostosowany zarówno do potrzeb informacyjnych użytkowników systemu, jak i do specyfiki dziedziny wiedzy, która wchodzi lub obejmuje zakres pola informacyjnego systemu. Język opisu bibliograficznego wyznacza zestaw atrybutów obiektów informacyjnych, które będą odwzorowane podczas procesu katalogowania. Sporządzony w ten sposób zestaw metadanych musi uwzględniać cechy wyszukiwawcze obiektów informacyjnych. Terminem tym określa się „te spośród elementów opisu [...], które występują lub mogą występować w pytaniach [...] i według których może być prowadzone wyszukiwanie”. [Babik 1992, s. 81] Cechy wyszukiwawcze obiektów informacyjnych ustala się na podstawie analizy zbioru informacyjnego oraz analizy zbioru pytań użytkowników. Badania nad potrzebami informacyjnymi grupy użytkowników docelowych systemu informacyjnego są nieocenionym źródłem informacji również w kontekście identyfikacji cech wyszukiwawczych. Język opisu bibliograficznego formalizuje płaszczyzny interpretacji obiektu informacyjnego do postaci zbioru jego cech, których wartości są pożądane do pozyskania podczas wyszukiwania informacji. Uznanie określonej z nich za cechę wyszukiwawczą jest równoznaczne z nadaniem jej statusu punktu dostępu. Atrybuty obiektu informacyjnego wyodrębniane przez język opisu bibliograficznego uzależnione są od formy jego utrwalenia lub specyfiki zakresu tematycznego, który reprezentuje. Tradycyjne języki opisu bibliograficznego identyfikują cechy obiektów w zależności od rodzaju dokumentów, o których informacje udostępnia system. Tym samym funkcjonują formaty opisu dla dokumentów zwartych, ciągłych, norm i patentów, materiałów kartograficznych, itd. W przypadku WWW pojawia się problem heterogeniczności jego zasobów. Format opisu dla dokumentów elektronicznych opublikowanych w WWW powinien uwzględniać wszystkie potencjalne cechy formalne obiektów. Dzięki tendencji do standaryzacji opisu zasobów sieciowych powstał m.in. zestaw metadanych Dublin Core, który identyfikuje 15 podstawowych elementów odpowiadających cechom formy i treści obiektów informacyjnych. [Tab. 2.2. Rozdz. 2.] Ten język opisu bibliograficznego wyznacza zestaw elementów obowiązkowych i opcjonalnych, których wartości należy odwzorować podczas indeksowania. Umożliwia to tworzenie m.in. profiliów aplikacyjnych opartych na Dublin Core, które w dużym uproszczeniu, wykorzystują jego elastyczną strukturę umożliwiającą dostosowanie schematu do potrzeb konkretnego systemu. Obok schematu Dublin Core, który ma szerokie zastosowanie, istnieje wiele języków opisu bibliograficznego przeznaczonych do opisu zasobów sieciowych.



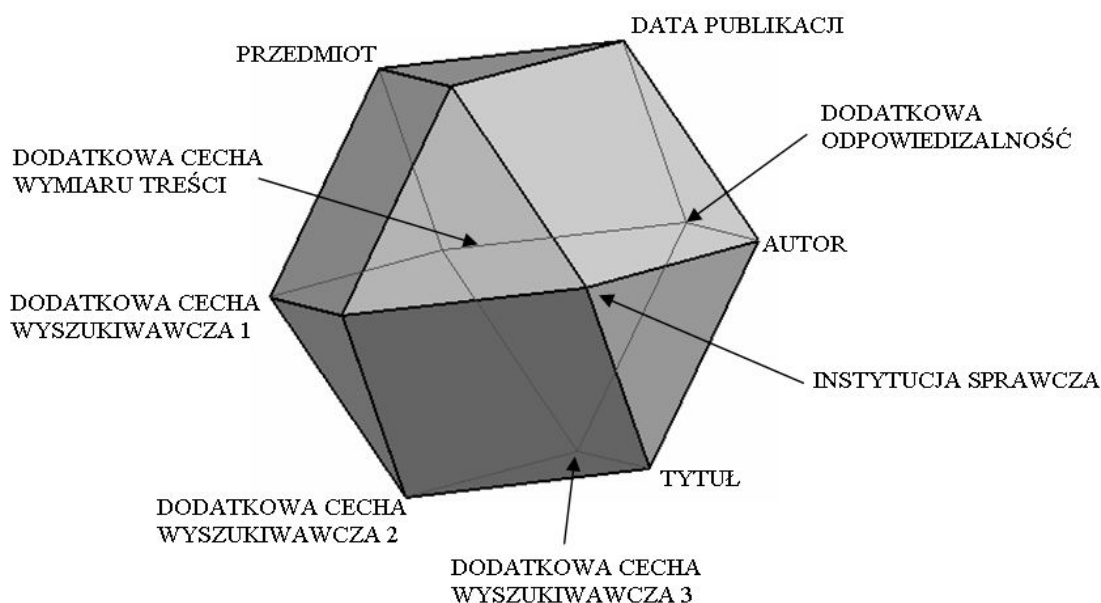
Jednym z tego typu języków opisu danych jest format Metadane Obiektu Edukacyjnego (ang. Learning Object Metadata), opracowany przez zespół specjalistów związany z Institute of Electrical and Electronics Engineers (IEEE). Celem tego formatu jest identyfikacja relewantnych atrybutów obiektów informacyjnych, które pełnią funkcję edukacyjną. Za pomocą tego schematu można efektywnie odwzorować požądane przez użytkowników informacje o obiekcie edukacyjnym. Wspomniany język opisu bibliograficznego przeznaczony jest dla systemów informacyjnych zarządzających zarówno zbiorami tradycyjnymi, jak i elektronicznymi. Podstawowa struktura formatu zakłada wyodrębnienie dziewięciu kategorii danych o obiekcie. Są to:

- dane ogólne – charakterystyka obiektu edukacyjnego jako całości,
- cykl trwania (ang. lifecycle) - informacje o genezie i aktualnym stanie obiektu informacyjnego oraz zmianach, jakie zachodziły w jego budowie i treści w czasie,
- meta-metadane – informacje o samym formacie opisu danych,
- techniczna charakterystyka obiektu edukacyjnego oraz wymogów technicznych, których spełnienie jest konieczne do jego wykorzystania,
- poziom charakterystyki edukacyjnej obiektu informacyjnego,
- prawa autorskie do obiektu edukacyjnego,
- relacje, w jakie wchodzi obiekt edukacyjny z innymi obiektami i źródłami informacji,
- adnotacja, charakterystyka słowna,
- klasyfikacja – reprezentacja treści i/lub formy za pomocą wybranego języka informacyjnego typu klasyfikacji. [Hodgins 2002]

Przedstawiony język opisu bibliograficznego wyznacza płaszczyzny interpretacji obiektów informacyjnych, w ramach których można wyodrębniać cechy wyszukiwawcze równoznaczne z punktami dostępu. Tego typu języki uwzględniają, w zestawie atrybutów obok wskazywania na istotne cechy formy, również te które dotyczą wymiaru treści. Wiąże się to z wyodrębnieniem w formacie opisu jednego lub kilku atrybutów pełniących tę funkcję, gdzie do odwzorowania tego rodzaju informacji wykorzystuje się określone języki informacyjno-wyszukiwawcze.

Rola języka informacyjno-wyszukiwawczego w kontekście identyfikacji punktów dostępu, w ramach procesu organizacji informacji, polega na sprowadzeniu niezidentyfikowanego obiektu informacyjnego do postaci „bryły o określonych kształtach”, której płaszczyzny wyznaczają kategorie cech, a wierzchołki konkretne atrybuty (Rys. 2.8). Tego rodzaju modelowanie jest funkcją języków opisu

bibliograficznego, które w przypadku zasobów cyfrowych utożsamiane są ze schematem metadanych.



Rys. 2.8. Modelowanie przykładowej jednostki opisu. [Źródło: opracowanie własne]

Zdolność tego rodzaju języków do identyfikacji relewantnych cech obiektów informacyjnych przekłada się na możliwości wyszukiwawcze całego systemu informacyjnego.

### 2.7.2. Kodowanie

Język informacyjny w postaci formatu opisu wyznacza zbiór cech obiektu informacyjnego, które będą odwzorowywane podczas indeksowania, i które będą mogły potencjalnie posiadać status punktu dostępu, czyli są cechami wyszukiwawczymi. W procesie katalogowania każdej ze zidentyfikowanych cech nadaje się określone wartości zgodne z ich postacią w obiekcie informacyjnym. W tym celu do odwzorowania i skodyfikowania wartości poszczególnych cech wykorzystuje się również języki informacyjne zaprojektowane do pełnienia tej funkcji. Na tym etapie, zwanym kodowaniem metainformacji, funkcja języka informacyjno-wyszukiwawczego polega na udostępnianiu jego zasobu leksykalnego oraz reguł łączenia poszczególnych jednostek leksykalnych w większe całości znaczeniowe. Przyjęcie stałej formy dla tej samej wartości umożliwia uniknięcie problemu synonimii wyszukiwawczej oraz zminimalizowanie ryzyka błędnego zapisu powodującego obniżenie efektywności wyszukiwania informacji.

W zależności od rodzajów atrybutów, narzędzia służące do kodowania ich wartości mogą odwzorowywać cechy formy i treści. Cechy formalne obiektów mogą być reprezentowane za pomocą zbiorów słownictwa kontrolowanego, list wartości, kartotek wzorcowych, a także dokumentów normatywnych (Tab. 2.3).

Tab. 2.3. Potencjalne źródła wartości dla cech formalnych. [Źródło: opracowanie własne]

Cecha	Potencjalne źródło wartości
Autor, twórca	kartoteka wzorcowa nazw osobowych
Tytuł	kartoteka wzorcowa tytułów ujednoliconych
Typ obiektu (forma utrwalenia)	lista wartości, słownik języka informacyjnego (odpowiednie jednostki leksykalne)
Data utworzenia	schemat kodowania (np. dd-mm-rrrr)
Miejsce wydania	lista wartości, tezaurus nazw geograficznych
Wydawca	kartoteka wydawców
Zasięg geograficzny	tezaurus nazw geograficznych, dokumenty o charakterze normatywnym (dot. skrótów)
Język obiektu	lista wartości, dokumenty o charakterze normatywnym (dot. skrótów)

Przedstawione w Tab. 2.3 potencjalne źródła wartości dla wybranych cech wymiaru formy obiektów informacyjnych ilustrują jedynie możliwości zastosowania zbiorów słownictwa kontrolowanego w tym zakresie.

Kodowanie informacji odwzorowujących wymiar treści obiektów informacyjnych to realizacja podstawowej (metainformacyjnej) funkcji języka informacyjno-wyszukiwawczego. Rola języka informacyjnego polega tutaj na umożliwieniu przeprowadzenia przez indeksatora procesu określania przedmiotu i treści obiektu i zakodowania go z wykorzystaniem środków leksykalnych języka. Forma i struktura komunikatów zawierających tego rodzaju informacje uzależniona jest od rodzaju notacji (paranaturalna/sztuczna) oraz reguł gramatycznych języka informacyjnego użytego do tego celu. Stopień kontroli słownictwa wyznacza również zasady pragmatyki zastosowanego języka informacyjnego. Ten w różnym stopniu sformalizowany sposób reprezentacji treści jest często uzupełniany poprzez możliwość odwzorowania tej samej kategorii cech w postaci charakterystyki słownej, streszczenia, abstraktu, a także dołączenia np. spisu treści dokumentu.

### 2.7.3. Organizacja punktów dostępu

Rola języka informacyjnego w organizacji punktów dostępu polega na umożliwieniu realizacji jego funkcji wyszukiwawczej, czyli drugiej podstawowej

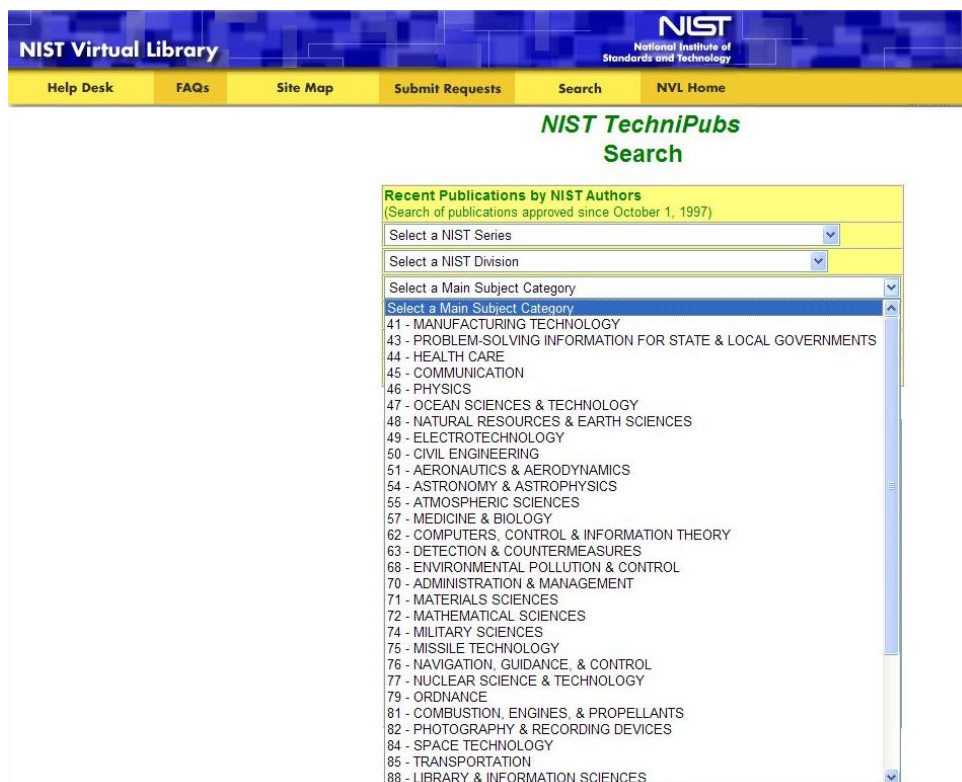
funkcji. W procesie wyszukiwania informacji w systemie informacyjno-wyszukiwawczym zachodzi interakcja użytkownika z mechanizmem wyszukiwawczym oraz dodatkowymi narzędziami umożliwiającymi dostęp do zasobów systemu. Organizacja punktów dostępu w narzędziach realizujących dostęp do informacji z wykorzystaniem języka informacyjno-wyszukiwawczego zostanie scharakteryzowana z perspektywy ich włączenia w mechanizmy wyszukiwawcze oraz narzędzia dostępu przedmiotowego, a więc dwóch głównych sposobów pozyskiwania informacji z systemów informacyjnych w World Wide Web.

### **2.7.3.1. Mechanizmy wyszukiwawcze**

Istota wykorzystania przez użytkownika mechanizmów wyszukiwawczych (rozumianych tutaj jako mechanizmy automatycznie przeszukujące zasoby kolekcji) sprowadza się do poprawnej werbalizacji jego potrzeby informacyjnej i jej manifestacji w postaci zapytania informacyjnego, które zostaje wysłane do mechanizmu wyszukiwawczego\*. Rola JIW w tworzeniu punktów dostępu w internetowych mechanizmach wyszukiwawczych polega na wykorzystaniu jego słownika do budowy zapytania informacyjnego. Mechanizmy wyszukiwawcze udostępniają użytkownikowi możliwości wyboru kryteriów wyszukiwawczych, które mają postać cech wyszukiwawczych zidentyfikowanych w strukturze formatu opisu. Zawężenie wyników wyszukiwania informacji do przeszukiwania wartości określonego zbioru atrybutów jest jednym ze sposobów wyboru punktu dostępu. Udział JIW w funkcjonowaniu mechanizmów wyszukiwawczych polega tym samym na udostępnieniu przez system możliwości pobierania przez użytkownika jednostek leksykalnych z jego słownika do konstruowania zapytania informacyjnego (Rys. 2.9, 2.10) Angażuje to użytkownika do manualnego pobierania poszczególnych jednostek i łączenia ich za pomocą przyjętych w systemie technik wyszukiwawczych (np. algebry Boole'a).

---

\* Cały przebieg procesu wyszukiwania jest o wiele bardziej skomplikowany i angażuje użytkownika wymuszając określony algorytm postępowania. Na potrzeby pracy zastosowano jednak tego rodzaju uproszczenie oraz przyjęto perspektywę, w której centrum jest system oraz operacje jakie zachodzą podczas procesu wyszukiwania. Zabieg ten podyktowany jest skupieniem uwagi wyłącznie na udziale języka informacyjnego w tym procesie.



Rys. 2.9. Dostęp do listy wartości dla cechy kategoria tematyczna w NIST Virtual Library.

[Źródło: <http://nvl.nist.gov/>]



Rys. 2.10. Dostęp do słownika JIW poprzez mechanizm wyszukiwawczy w systemie Virtual Health Library [Źródło: <http://www.virtualhealthlibrary.org/php/index.php?lang=en>]

Postęp w automatycznym przetwarzaniu danych umożliwia także przeprowadzanie takiego zabiegu automatycznie. Zintegrowanie mechanizmu wyszukiwawczego ze

słownikiem danego JIW umożliwia automatyczne tłumaczenie zapytania użytkownika do postaci ciągu jednostek leksykalnych równokształtnych z ich odpowiednikami w słowniku zastosowanego języka informacyjnego. Jest to możliwe poprzez zastosowanie zaawansowanych technik dokonujących transformacji językowych na zbiorze właściwych elementów opisu z wykorzystaniem ustrukturyzowanego zasobu leksykalnego derywowanego ze słownika danego JIW. Najprostszym rozwiązaniem w ramach takiego podejścia jest dążenie do redukcji synonimii wyszukiwawczej poprzez zastosowanie narzędzia zwanego pierścieniem synonimów (ang. synonym ring). [Leise, Fast, Steckel 2003; Tomaszczyk 2007] Polega ono na wiązaniu relacją ekwiwalencji znaczeniowej określonej grupy terminów, ale wyłącznie na potrzeby wyszukiwawcze. [Tudhope, Koch, Heery 2006] Powstała w ten sposób struktura oparta na relacji semantycznej jest wykorzystywana do tzw. ekspansji zapytania (ang. query expansion) o dodatkowe warianty językowe dla wprowadzonego przez użytkownika słowa lub słów kluczowych. Wykorzystanie tego typu narzędzia pozwala na efektywne przeprowadzenie przez użytkownika sesji wyszukiwawczej bez konieczności posiadania przez niego wiedzy o formie terminów reprezentujących jego potrzebę informacyjną przyjętych w procesie indeksowania. Zakres zastosowania pierścieni synonimów jest duży. Obok różnych wariantów językowych dla terminów przyjętych do oznaczania tych samych pojęć, można stosować je również do:

- akronimów nazw własnych i ich pełnych form (np. PTIN, Polskie Towarzystwo Informacji Naukowej),
- różnej pisowni nazw własnych (np. William Shakespeare, William Szekspir),
- terminów naukowych i popularnie stosowanych (np. kwas acetylosalicylowy, aspiryna) [Leise, Fast, Steckel 2003],
- błędnego i poprawnego zapisu tej samej nazwy (np. klsyfikacja, klasyfikacja).

Organizacja punktów dostępu do kolekcji w mechanizmach wyszukiwawczych z wykorzystaniem języka informacyjno-wyszukiwawczego realizowana jest poprzez sterowanie procesem budowy zapytania za pomocą wyrażeń przejmowanych z jego słownika. Obok atrybutów obiektów reprezentujących wymiar treści, na poziomie mechanizmu wyszukiwawczego wykorzystuje się również kontrolę formalną dla wielu pozostałych rodzajów cech wyszukiwawczych. W takich przypadkach wartości dla wybranych cech wyszukiwawczych mogą pochodzić z kartotek wzorcowych lub kontrolowanych list wartości. Podczas tworzenia zapytania informacyjnego użytkownik ma do nich bezpośredni dostęp. Włączanie zbiorów słownictwa kontrolowanego w interfejs mechanizmu wyszukiwawczego ułatwia budowanie zapytania

informacyjnego z wykorzystaniem zasobu leksykalnego reprezentującego określone kategorie cech obiektów informacyjnych. Jednocześnie minimalizuje się ryzyko pojawienia się synonimii wyszukiwawczej poprzez możliwość wyboru form preferowanych dla wyrażen wyszukiwawczych.

### **2.7.3.2. Narzędzia dostępu przedmiotowego**

Rola języka informacyjno-wyszukiwawczego w organizacji punktów dostępu w narzędziach dostępu przedmiotowego funkcjonujących w systemach i serwisach informacyjnych w WWW polega na wykorzystaniu jego organizacji pola semantycznego do budowy tego typu narzędzi. Wyrażenia mające status punktu dostępu zostają zorganizowane do postaci struktur równoważnościowych lub schematów hierarchicznych. Każda z użytych jednostek leksykalnych JIW jest nazwą dla odpowiedniej kategorii tematycznej. Zorganizowane do postaci struktury hierarchicznej lub ahierarchicznej tworzą narzędzie dostępu do kolekcji. Ich włączenie w organizację hipertekstową pozwala użytkownikowi na eksplorację takiego narzędzia i wybór wyrażenia, które w sposób pośredni lub bezpośredni odsyła go do relewantnych obiektów informacyjnych.

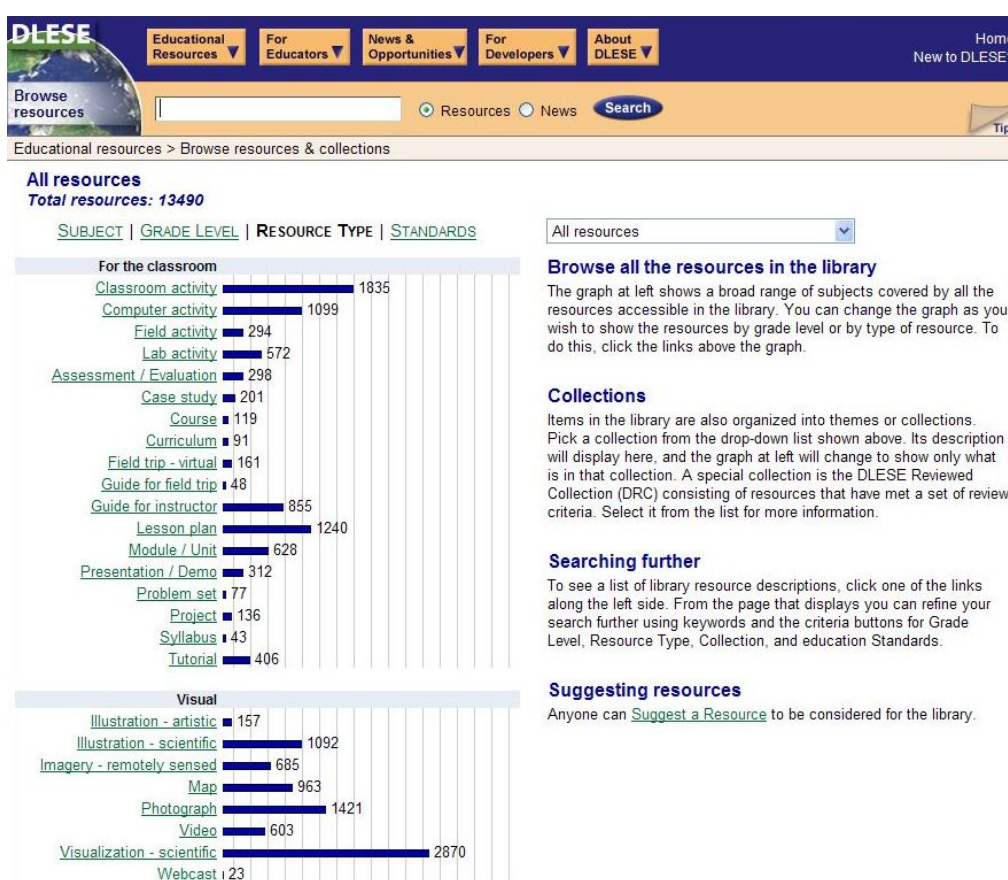
#### **2.7.3.2.1. Struktury równoważnościowe**

Dystrybucja punktów dostępu w narzędziach dostępu przedmiotowego w strukturach równoważnościowych polega na generowaniu list wartości dla określonych kategorii punktów. Narzędzia te pełnią funkcję indeksów, których elementy (wejścia) skierowują do określonych stref zawartości systemu, obiektów informacyjnych, czy też ich charakterystyk wyszukiwawczych. Zasadę porządkującą wyznacza alfabetyczny układ wyrażen. Ich zagnieżdżenie w strukturze hipertekstu powoduje, że każde z nich pełni funkcję węzła, którego aktywacja powoduje uruchomienie odesłania hipertekstowego realizującego operację wyodrębniania ze zbioru wyszukiwawczego elementów spełniających tak ustanowione kryteria relewancji systemowej i wyświetlania rezultatów tej operacji w odpowiedni sposób.

Narzędzia dostępu przedmiotowego wykorzystują JIW w takim zakresie, w jakim jest on użyty podczas indeksowania. Tym samym generowanie indeksów, których hasła pełnią funkcje dostępu, polega na automatycznym wyborze wartości dla danej cechy ze wszystkich rekordów w bazie, wyposażenia każdej z nich w adres

rekordu i uwzględnienia tych danych w odesłaniu hipertekstowym opracowanym dla każdego hasła.

Do najczęściej występujących cech wyszukiwawczych stanowiących podstawę do generowania tego typu narzędzi należą podstawowe atrybuty identyfikujące obiekt informacyjny (np. autor, tytuł) oraz atrybuty reprezentujące jego treść, a także specyficzne dla zakresu tematycznego systemu, dodatkowe cechy wyszukiwawcze (np. poziom kształcenia dla systemów z zakresu edukacji). Ze względu na niejednorodność formalną zasobów WWW, w systemach informacyjnych działających w tym środowisku jako dodatkową kategorię punktów dostępu wykorzystuje się również alfabetyczne wykazy wartości dla tej cechy (rodzaj źródła) (Rys. 2.11).



Rys. 2.11. Alfabetyczny wykaz wartości punktów dostępu dla cechy *rodzaj źródła* w Digital Library of Earth System Education [Źródło: <http://www.dlese.org>]

Najczęściej praktykowanym rozwiązaniem równoważnościowej organizacji punktów dostępu jest generowanie wartości dla cechy *temat* lub *przedmiot obiektu informacyjnego*. Narzędzie dostępu zbudowane według tego kryterium przyjmuje wówczas postać zbioru wyrażen w porządku alfabetycznym, których suma zakresów wyznacza aktualny zakres pola tematycznego systemu. Kryterium włączenia danego wyrażenia do narzędzia jest jego wystąpienie w odpowiednim polu charakterystyki



wyszukiwawczej (Rys. 2.12).



Rys. 2.12. Wykaz słów kluczowych w narzędziu dostępu przedmiotowego w Śląskiej Bibliotece Cyfrowej. [Źródło: <http://www.sbc.org.pl/dlibra/>]

Organizacja punktów dostępu w narzędziach wykorzystujących struktury równoważnościowe może przyjmować formę mieszaną lub selektywną. W pierwszym przypadku polega to na konstrukcji narzędzia dostępu, w którym występują różne rodzaje punktów dostępu. Oznacza to np. włączanie w porządku alfabetycznym słów kluczowych reprezentujących wymiar treści oraz wartości dla cechy rodzaj źródła, czyli wymiaru formy. Selektywna organizacja punktów dostępu polega na generowaniu osobnych wykazów wartości dla konkretnych cech wyszukiwawczych. Np. w Digital Library of Earth System Education (Rys. 2.11) narzędzie dostępu przedmiotowego wykorzystujące równoważnościową strukturalizację punktów dostępu jest zbudowane z czterech indeksów, których podstawą konstrukcji były wartości dla cech:

- przedmiot,
- poziom kształcenia,
- typ źródła,
- zakres standardu edukacyjnego.

Narzędzia dostępu przedmiotowego oparte na strukturach równoważnościowych wykorzystują punkty dostępu w postaci atrybutów obiektów reprezentujących poza treściowymi także inne cechy. Tym samym realizują wielowymiarowy tryb dostępu do kolekcji. Język informacyjno-wyszukiwawczy traktowany jest tutaj jako źródło leksyki

do oznaczania wartości poszczególnych rodzajów atrybutów, które stanowią podstawę do generowania narzędzi dostępu do kolekcji.

### **2.7.3.2.2. Struktury hierarchiczne**

Narzędzia dostępu przedmiotowego wykorzystujące hierarchiczną strukturyzację zbioru słownictwa, które pełnią funkcję punktów dostępu charakteryzują się dużym zróżnicowaniem zarówno w budowie jak i w praktycznym zastosowaniu. Rola języka informacyjno-wyszukiwawczego w budowie tego typu narzędzi polega na zastosowaniu organizacji jego pola semantycznego do tworzenia struktury narzędzia dostępu przedmiotowego. Duży wpływ na konstrukcję hierarchicznych narzędzi dostępu przedmiotowego mają JIW o rozbudowanej paradygmatyce (szczególnie klasyfikacje).

Opracowanie własnych schematów hierarchicznych wiąże się z dużym nakładem prac, jednak wpływa to korzystnie na efektywność narzędzia. Uwzględniają one bowiem specyfikę zakresu przedmiotowego zbioru informacyjnego oraz potrzeby informacyjne potencjalnych użytkowników. [Bechler, Knight, Place 1999] Zaletą tego rodzaju schematów jest ich elastyczność pod względem absorpcji nowych tematów i obszarów wiedzy wchodzących w zakres danego systemu. Do wad takiego rozwiązania należą:

- czasochłonność i duży nakład środków,
- konieczność zaangażowania specjalisty dziedzinowego,
- duże ryzyko subiektywizmu, co może prowadzić do braku spójności,
- słabe rozpowszechnienie schematu, które jest przyczyną małej jego znajomości wśród użytkowników i tym samym trudności w jego obsłudze. [Bechler, Knight, Place 1999]

Tworzenie własnego schematu jest konieczne w przypadku, kiedy nie ma zgodności zakresowej z istniejącymi schematami np. klasyfikacyjnymi zapewniającej wystarczający poziom efektywności narzędzia. Dostosowywanie istniejących schematów klasyfikacyjnych do postaci struktur hierarchicznych kategorii tematycznych uznaje się za bardziej efektywne rozwiązanie. [np. Koch, Day 1997] Do niewątpliwych zalet wykorzystania tego typu języków informacyjnych należą:

- **dostęp** do zasobów - jest uproszczony do wyboru odpowiedniej klasy z układu hierarchicznego, której zakres odpowiada zakresowi tematycznemu poszukiwanych dokumentów;

- **poszerzanie i zawężanie zapytania informacyjnego** - układ hierarchiczny klas w prosty i intuicyjny sposób umożliwia zawężenie i poszerzenie zapytania poprzez wykorzystanie relacji hierarchicznych zastosowanych podczas budowy powiązań między klasami;
- **kontekst znaczeniowy** – wyszukiwanie z wykorzystaniem terminów w układzie hierarchicznym jest ułatwione przez kontekst semantyczny dla danego terminu, który tworzą terminy z jego otoczenia;
- **dostęp dla wielu języków etnicznych** - klasyfikacja jest językiem o notacji sztucznej, która jest niezależna od języków etnicznych. Pozwala to na ich zastosowanie w kolekcjach wielojęzycznych.
- **podział i obsługa bazy danych** – ze sklasyfikowanego dużego zbioru może zostać logicznie wydzielony mniejszy, jeżeli jest to wymagane w celu poprawy efektywności wyszukiwania;
- **interoperacyjność** – sztuczna notacja oraz powszechność stosowania, szczególnie uniwersalnych schematów klasyfikacyjnych, umożliwia wykorzystanie spójności na płaszczyźnie wymiany danych pomiędzy kilkoma systemami i zastosowanie jednoczesnego przeszukiwania wielu baz danych;
- **stabilność** – ugruntowany schemat klasyfikacyjny nie jest zagrożony zaniknięciem, więcej, duże i powszechnie stosowane schematy są aktualizowane i udoskonalane;
- **powszechna znajomość** – szerokie wykorzystanie klasyfikacji w bibliotekach do opisu i porządkowania zbiorów wpływa na znajomość tych narzędzi przez użytkowników bibliotek, a obecnie internautów. Stopień znajomości jest różny dla poszczególnych schematów klasyfikacyjnych jest jednak jest większy niż w przypadku schematów hierarchicznych zaprojektowanych na potrzeby konkretnych systemów;
- **dostępność** – wiele tablic klasyfikacji dostępnych jest w formie elektronicznej gotowej do dostosowania. [Koch, Day 1997]

Stosowanie języków typu klasyfikacji jako narzędzia porządkowania i dostępu do informacji, gdzie organizacja pola semantycznego charakteryzuje się podziałem logicznym (rozłącznym i adekwatnym) powoduje również pewne bariery. Podział logiczny wpływa czasami na rozdzielanie dokumentów powiązanych ze sobą. Można ten problem rozwiązać wykorzystując np. relację asocjacyjną. Największe i najbardziej rozpowszechnione klasyfikacje stanowią spuściznę XIX wieku (Uniwersalna Klasyfikacja Dziesiąta, Klasyfikacja Dziesiąta Deweya, Klasyfikacja Biblioteki

Kongresu) i są odzwierciedleniem postrzegania rzeczywistości sprzed dwóch wieków. Poza tym, charakteryzują się dużym opóźnieniem w asymilacji nowych dziedzin i zagadnień, co w przypadku środowiska internetowego, które żywiłowo reaguje na nowe kierunki w badaniach naukowych, może być problemem w efektywnym odwzorowaniu nowych zakresów badawczych.

Dostosowanie istniejącego schematu klasyfikacyjnego w narzędziach dostępu przedmiotowego uwarunkowane jest zakresem zbioru informacyjnego. Najczęściej wykorzystuje się w tym celu podział klas na pierwszym stopniu i jego kontynuację do ustalonego stopnia głębokości. Powstała w ten sposób struktura jest modyfikowana do postaci hierarchicznego systemu kategorii tematycznych wyrażonych poprzez nazwy naturalne. [Sosińska-Kalata 2002b, s. 212] Rezygnacja ze sztucznej notacji podyktowana jest względami praktycznymi i ma na celu ułatwienie korzystania ze schematu. Najlepiej widoczne jest to w katalogach internetowych o zakresie uniwersalnym. Dotyczy to szczególnie rozbudowy kategorii *nauka*, który częściowo przypomina wykaz klas pierwszego stopnia w powszechnie stosowanych klasyfikacjach bibliotecznych (np. UKD, KDD czy KBK).

Struktury hierarchiczne w katalogach internetowych charakteryzują się również obecnością klas o charakterze formalnym. Mają one przeważnie formę etykiet: *aktualności, ludzie, informatory, media*, itp.

Istota funkcjonowania tego rodzaju narzędzi polega na wyposażeniu każdej klasy posiadającej jednocześnie status punktu dostępu w odesłanie hipertekstowe, którego aktywacja powoduje:

- wyświetlenie obiektów informacyjnych (lub ich charakterystyk), które zostały zgrupowane w danej klasie,
- wyświetlenie rozbudowy danej klasy na kolejnym stopniu podziału z wykorzystaniem relacji hierarchicznych.

Struktura hierarchicznych narzędzi dostępu przedmiotowego w katalogach internetowych zazwyczaj odwzorowuje strukturę współczesnej nauki. Realizuje się to często w postaci struktur polirelacyjno-polisemantycznych. [Sosińska-Kalata 2002, s. 217] Tutaj również względy pragmatyczne wymuszają wielokrotną lokalizację tej samej klasy w wielu miejscach w strukturze hierarchicznej. Jak zauważa B. Sosińska-Kalata [Sosińska-Kalata 2002b, s. 217] pełnią one funkcję odsyłaczy uzupełniających. Przyjmuje się, że rozbudowa kategorii tematycznej następuje przy minimalnej pojemności około 20-25 odesłań. [Sosińska-Kalata 2002b, s.215]

Serwis internetowy *Cyberstack's* prowadzony w latach 1996-1998 przez

G. McKiernana [McKiernan 1996] zawiera wykaz serwisów i systemów internetowych, w których do organizacji punktów dostępu w stosuje się m.in. funkcjonujące klasyfikacje biblioteczno-bibliograficzne. Wśród wymienionych, najczęściej stosowaną klasyfikacją jest Klasyfikacja Dziesiąta Deweya (20 wskazań), Klasyfikacja Biblioteki Kongresu (7 wskazań) oraz Uniwersalna Klasyfikacja Dziesiąta (5 wskazań).<sup>\*</sup> Do pozostałych należą: Göttinger Online Klassifikation, Sveriges Allmänna Biblioteksforening, Engineering Information (Ei) Classification Codes, Mathematics Subject Classification, Nederlandse Basisclassificatie, ACM Computing Classification System, Klasyfikacja National Library of Medicine.

Szczegółowe badania nad wykorzystaniem klasyfikacji w organizacji zasobów Internetu prowadzono w ramach projektu Unii Europejskiej *DESIRE Development of a European Service for Information on Research and Education* pt. „Rola schematów klasyfikacyjnych w opisie i dostępie do zasobów Internetu”. [Koch, Day 1997] W dokumencie tym wskazano m.in. sposoby wykorzystania tego typu języków informacyjnych do porządkowania informacji i organizacji dostępu do dokumentów w WWW poprzez ich adaptację do środowiska hipertekstowego. Zauważono nieadekwatność postrzegania rzeczywistości przez klasyfikacje nauk wywodzące się jeszcze z XIX wieku i reprezentowane przez największe i jednocześnie najczęściej stosowane trzy klasyfikacje - KBK, UKD, KDD. Wskazano również na potencjalną efektywność schematów fasetowych w organizacji zasobów sieciowych, którą reprezentują m.in. Klasyfikacja Dwukropkowa Ranganathana czy Klasyfikacja Bibliograficzna Blissa (BC2). Zidentyfikowano podstawowe wady stosowanych klasyfikacji w postaci trudności w asymilacji nowych obszarów wiedzy oraz sztywności zasad podziału logicznego, która może wpływać na dyspersję w pokrewnych zagadnięć. Duża „popularność” KDD w organizacji zasobów Internetu wynika, jak stwierdzono, z jej większej elastyczności niż KBK oraz mniejszego stopnia skomplikowania niż w przypadku UKD. Drugi etap badań w ramach projektu DESIRE zmierzał do wykorzystania metod lingwistycznych w automatycznym klasyfikowaniu zasobów WWW. Doświadczenia z prac w ramach DESIRE wykorzystywano podczas podobnych projektów badawczych zmierzających do automatycznej kategoryzacji obiektów informacyjnych z wykorzystaniem istniejących schematów klasyfikacyjnych. Były to m.in.:

---

<sup>\*</sup> Dane pochodzą z 2001 roku i aktualność jest dyskusyjna. Wskazują one jednak na wykorzystanie przede wszystkim ugruntowanych, znanych i uniwersalnych klasyfikacji piśmiennictwa do organizacji zasobów internetowych.

- The Nordic WAIS/WWW Project z użyciem UKD,
- Projekt Gerhard (German Harvest Automatem Retrieval Directory) z użyciem UKD,
- Projekt Scorpion (OCLC) z użyciem KDD.

W ramach projektu DESIRE dokonano analizy wykorzystania schematów klasyfikacji w organizacji i dostępie do zasobów serwisów i systemów internetowych w odniesieniu do:

- Klasyfikacji Dziesiątej Deweya,
- Uniwersalnej Klasyfikacji Dziesiątej,
- Klasyfikacji Biblioteki Kongresu,
- Nederlandse Basisclassificatie,
- Sveriges Allmänna Biblioteksförning,
- Iconclass,
- National Library of Medicine Classification,
- Engineering Information (Ei) Classification Codes,
- Mathematics Subject Classification,
- ACM Computing Classification System.

Różne sposoby wykorzystania schematów klasyfikacyjnych w narzędziach dostępu przedmiotowego wpływają na wizualną postać samych narzędzi. „W niektórych serwisach przeprowadzenie wyszukiwania za pomocą określonej klasyfikacji wymaga znajomości odpowiednich symboli, które czasami – aczkolwiek nie zawsze – można ustalić przeglądając internetową kopię tablic klasyfikacyjnych. W innych symbole klasyfikacyjne stanowią nazwy hiperłączy, obok których prezentowane są odpowiedniki słowne; użytkownik nawiguje po strukturze internetowej wersji klasyfikacji podobnie, jak czyniłby to przeglądając tradycyjne tablice klasyfikacji wyliczającej, jednak wybór któregoś z symboli w tym przypadku umożliwia mu natychmiastowe wyszukanie dokumentów, którym jest on przypisany. W jeszcze innych serwisach i wirtualnych bibliotekach operuje się po prostu wyrażeniami języka naturalnego, którym przyporządkowany jest symbol niewidoczny dla użytkowników; w tym przypadku frazy wyświetlane w porządku systematycznym są nazwami tematów, które często klasyfikuje się za pomocą symboli rozwiniętych lub złożonych o wysoce skomplikowanej strukturze”. [Sosińska-Kalata 2002b, s. 409]

Specyficznym narzędziem hierarchicznej organizacji punktów dostępu są taksonomie internetowe. Znaczenie terminu *taksonomia* w kontekście rozważań nad językami informacyjno-wyszukiwawczymi, a w szczególności klasyfikacją odnosi się

do zasad podziału i systematyzacji obiektów z uwzględnieniem w szczególności ich cech fizycznych. [Anderson, Carballo 2005] Ma on powszechne zastosowanie w naukach przyrodniczych, gdzie klasyfikacja i kategoryzacja organizmów przypomina raczej systematykę jednorodnych przedmiotów empirycznych.

Nowy, zaktualizowany punkt widzenia na to narzędzie organizacji wiedzy prezentują twórcy schematów nawigacyjnych oraz narzędzi dostępu przedmiotowego w systemach hipertekstowych. Tutaj taksonomia rozumiana jest jako struktura identyfikująca grupy kategorii tematycznych wraz z relacjami semantycznymi zachodzącymi między nimi, które reprezentują pole tematyczne określonego zbioru informacji. [Roberts 1999; Saeed, Chaudhry 2002] System strukturyzacji wiedzy w taksonomiach przypomina połączenie klasyfikacji z tezauresem. [Zhonghong, Khoo 2006] Wizualizacja drzewiastej dystrybucji kategorii tematycznych opracowywana jest na podstawie hierarchicznego układu klas np. w istniejących klasyfikacjach. Forma kategorii tematycznych jest determinowana postacią jednostek leksykalnych o notacji paranaturalnej, pochodzących ze zbiorów słownictwa o różnym stopniu kontroli. W ten sposób mamy do czynienia z układem hierarchicznym punktów dostępu, których forma przejmowana jest na podstawie kryterium zgodności zakresowej ze słownikami języków informacyjno-wyszukiwawczych o notacji paranaturalnej. Do każdej kategorii tematycznej z taksonomii często opracowuje się uproszczony artykuł hasłowy. Tworzą go dodatkowo wyrażenia pozostające z aktualnie wybranym w relacji synonimii wyszukiwawczej oraz np. skojarzeniowej. [Saeed, Chaudhry 2002] Znaczna część taksonomii funkcjonujących jako narzędzia dostępu przedmiotowego lub nawigacji nie reprezentuje takiego zaawansowanego sposobu organizacji punktów dostępu. W większości przypadków budowa taksonomii ogranicza się jedynie do hierarchicznego wykazu kategorii tematycznych.

Kryterium funkcjonalności taksonomii opiera się na ustaleniu konsensusu pomiędzy jej szerokością i głębokością. [Sosińska-Kalata 2005a] Względy funkcjonalne decydują tutaj o znacznym uproszczeniu samej struktury hierarchicznej poprzez zubożenie relacji semantycznych zachodzących pomiędzy kategoriami. Mamy więc do czynienia z wykorzystaniem struktur polirelacyjnych lub monorelacyjnych o różnym stopniu spójności. Taksonomie wykorzystujące relację generyczną właściwą (zakres szerszy/zakres węższy) nazywane są generalizującymi lub specjalizującymi, a oparte na relacji mereologicznej noszą nazwę agregujących, skupiających [Daconta, Orbst, Smith 2003, s. 146] Duża liczba taksonomii internetowych charakteryzuje się niekonsekwencją w stosowaniu wspomnianych relacji paradygmatycznych. Jest to

tłumaczone dążeniem do zapewnienia funkcjonalności całej struktury, jednak odbywa się kosztem spójności podziału. Pomimo elastyczności i pozorów funkcjonalności, taksonomie internetowe zajmują dalekie miejsce wśród narzędzi organizacji wiedzy pod względem ich potencjału semantycznego. Nie jest to spowodowane ułomnością samego narzędzia, lecz nieefektywnym jego wykorzystaniem w wielu systemach i serwisach internetowych.

S. Conway i C. Sligar twierdzą, że struktura taksonomii jest zdeterminowana również przez jej funkcję w systemie informacyjnym. Identyfikują oni cztery zmienne, które powinny być uwzględniane podczas budowy taksonomii: [Conway, Sligar 2002]

1. rodzaj narzędzia – schemat nawigacyjny, struktura organizacyjna serwisu,
2. wielkość (liczba etykiet),
3. forma (struktury ahierarchiczne, mono- lub polihierarchiczne),
4. zakres stosowania (lokalna, globalna, współdzielona).

Czynniki te poprzez identyfikację miejsca taksonomii w systemie wpływają zarówno na jej strukturę, jak i na sposoby jej implementacji.

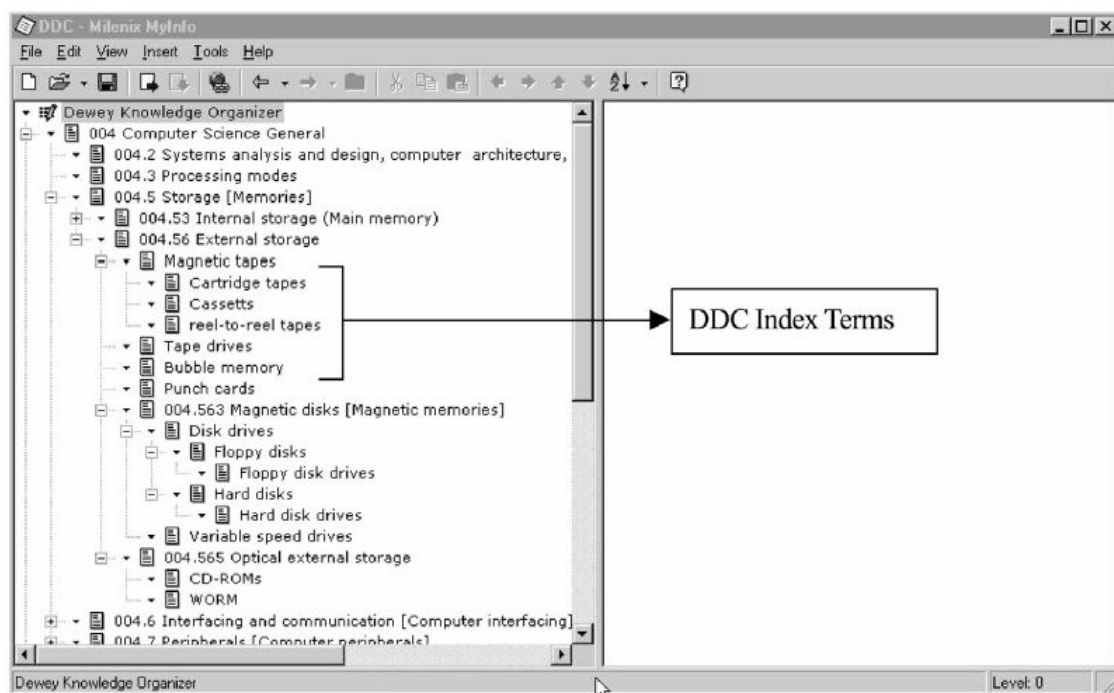
Taksonomia jest wykorzystywana do organizacji zawartości kolekcji oraz jako narzędzie dostępu do zasobów systemów komercyjnych i intranetów. Postrzegana jest jako mapa wiedzy ułatwiająca nawigację i dostęp do kapitału intelektualnego intranetu przedsiębiorstwa. [Zhonghong, Khoo 2006] Reprezentując zawartość kolekcji może realizować również cele biznesowe poprzez odpowiednie dostosowanie jej struktury, co wiąże się ze zmianą perspektywy interpretacji rzeczywistości dokumentacyjnej.

Szerokie wykorzystanie taksonomii jako jednego ze sposobów na odnalezienie roli klasyfikacji w organizacji zasobów cyfrowych ma swoje odbicie również w specyfice badań naukowych nad tym narzędziem. W centrum rozważań sytuuje się strukturę taksonomii jako mapę konceptualną organizowanej kolekcji oraz analizuje się możliwości wykorzystania klasyfikacji jako szkieletu strukturalnego narzędzia. Metodyka adaptacji klasyfikacji do budowy taksonomii opracowana przez H. Saeed oraz A.S. Chaudhry na bazie Klasyfikacji Dziesiątej Deweya (Rys. 2.13) zakłada:

1. Wybór odpowiednich klas i ich rozbudowy jako podstawy strukturalnej taksonomii.
2. Oznaczenie każdej z klas terminami indeksowymi z tablic.
3. Włączenie zbioru słownictwa kontrolowanego w hierarchię taksonomii poprzez:
  - a. odwzorowanie bezpośrednie (ang. direct mapping) – łączenie odpowiedników słownych w tablicach z wyrażeniami tezauryusa pod względem zgodności zakresowej,



- b. odwzorowanie pośrednie (ang. indirect mapping) – kryterium zakresu pojęcia reprezentowanego przez symbol klasyfikacyjny decyduje o włączeniu w jego otoczenie dodatkowych terminów z tezaury. [Saeed, Chaudhry 2002, s. 578-579]



Rys. 2.13. Tworzenie taksonomii na podstawie KDD. [Źródło: Saeed, Chaudhry 2002, s. 579]

Każdy z etapów tworzenia taksonomii na bazie schematu klasyfikacyjnego jest obarczony dodatkowymi czynnościami, a przedstawiony algorytm postępowania zakłada jedynie procesy związane z dostosowaniem schematu klasyfikacji, nie zaś cały przebieg tworzenia taksonomii. [por. Tomaszczyk 2007]

Prowadzone badania nad zastosowaniem schematów klasyfikacyjnych w narzędziach dostępu do kolekcji mają charakter wieloaspektowy. Podczas analizy tego rodzaju struktur dokonuje się często dekonstrukcji takich układów oraz odszukuje korelacje z istniejącymi schematami stanowiącymi podstawę do ich budowy. W nurt takich badań można włączyć eksperymenty m.in. Alana Weathleya oraz Diane Vizine-Goetz. [Wheatley 2000; Vizine-Goetz 2002] Celem badań A. Weathleya była analiza struktury katalogów wybranych portali internetowych. Hierarchiczne narzędzia dostępu przedmiotowego zostały określone przez niego mianem *drzew przedmiotowych* (ang. subject trees). Podsumowując swoje badania wyciągnął on następujące wnioski:

- drzewa przedmiotowe zbudowane są z 12-20 kategorii tematycznych na najwyższym poziomie,

- w kategoriach tematycznych pierwszego poziomu zgrupowano bardzo mały odsetek odesłań do obiektów informacyjnych (stanowią jedynie podstawę do dalszej rozbudowy),
- 2/3 odesłań zgrupowane jest na poziomach 4-5,
- drzewa przedmiotowe realizują efektywnie dostęp przedmiotowy do zasobów bez konieczności adaptowania tradycyjnych schematów bibliotecznych,
- drzewa przedmiotowe umożliwiają dostęp do zasobów za pomocą struktur płaskich i hierarchicznych bez konieczności formalnej specyfikacji relacji łączących wyrażenia stanowiące nazwy dla poszczególnych klas.

Publikacja A. Wheatleya była podstawą dalszych badań w tym zakresie realizowanych przez D. Vizine-Goetz. [Vizine-Goetz 2002] Porównała ona strukturę katalogów internetowych portali Yahoo! oraz LookSmart z dystrybucją klas w wydaniach skróconym i rozszerzonym Klasyfikacji Dziesiątej Deweya. Opracowana przez D. Vizine-Goetz metodologia ekstrakcji struktur hierarchicznych z katalogów internetowych uwzględnia zasadę identyfikacji rzeczywistego miejsca kategorii tematycznej w strukturze hierarchicznej. Przyjęła ona perspektywę badawczą, w której:

- symbole z poddziałami wspólnymi są traktowane jako należące do niższego poziomu niż klasa, na podstawie której zbudowano symbol rozwinięty,
- symbole złożone identyfikuje się z poziomem symbolu stanowiącego podstawę jego utworzenia i umiejscawia o jeden poziom niżej.

Ilościowy rozkład klas oraz kategorii tematycznych w analizowanych narzędziach przedstawia Tab. 2.4.

Tab. 2.4. Wyniki badań D. Vizine-Goetz [Źródło: Vizine-Goetz 2002]

Poziom	Liczba klas KDD wersja skrócona	Liczba klas KDD wersja rozszerzona	Liczba klas Yahoo!	Liczba klas LookSmart
1	10	10	14	13
2	99	99	420	160
3	897	897	3509	1681
4	4415	8121	7562	8604
5	8574	11372	11058	9191
6	9986	12360	6823	1021
7	5590	7354	15352	0
8	1820	2917	16632	0
9	535	923	2376	0
10	93	248	0	0
11	20	36	0	0
12	9	11		
13	0	2		
Razem	32048	44350	63746	20670

Wyniki badań D. Vizine-Goetz wskazują, że dystrybucja 80% łącznej liczby klas w analizowanych katalogach internetowych ma miejsce na poziomach od trzeciego do

piątego. Średnia pojemność kategorii tematycznej wynosi od 9 odesłań (dla Yahoo!) do 12 (dla LookSmart). KDD ma najgłębszą strukturę z analizowanych schematów (13 poziomów), LookSmart ([www.looksmart.com/](http://www.looksmart.com/)) najpłytszą (sześć poziomów). Yahoo! ([www.yahoo.com/](http://www.yahoo.com/)) i KDD udostępnia zbliżoną liczbę kategorii tematycznych łącznie na poziomach od czwartego do siódmego (40795 Yahoo! i 38207 rozszerzona wersja KDD). Yahoo! udostępnia dwa razy więcej kategorii tematycznych na poziomie siódmym niż rozszerzona KDD, chociaż może to być spowodowane przyjętą metodologią analizy planu notacyjnego symboliki KDD i przypisywania poziomu w hierarchii. Zarówno D. Vizine-Goetz jak i A. Wheatley zgodnie twierdzą, że:

- klasyfikacje biblioteczne są bardziej efektywne w porównaniu z projektowanymi schematami stosowanymi w Internecie pod względem ogólnego zakresu tematycznego,
- klasyfikacje biblioteczne dysponują strukturami hierarchicznymi dostatecznie głębokimi i szerokimi do zastosowania jako narzędzie dostępu przedmiotowego,
- notacja hierarchiczna w klasyfikacjach pozwala na derywację struktur hierarchicznych i wykorzystania ich do budowy taksonomii opartej na etykietach o notacji paranaturalnej.

Badania D. Vizine-Goetz i A. Wheatleya są istotne z punktu widzenia analizy gotowego narzędzia dostępu pod kątem dystrybucji kategorii tematycznych oraz rozkładu odesłań do obiektów informacyjnych na poszczególnych poziomach. Pokazują fragmentację pola tematycznego kolekcji z wykorzystaniem taksonomii internetowych.

W nurt badań jakościowych włączają się eksperymenty Chaima Zinsa [Zins 2002] oraz zespołu badawczego pod kierunkiem Kim Kyung-Sun. [Kyung-Sun i in. 2006] Badania Ch. Zinsa nad organizacją klas w narzędziach dostępu dużych portali internetowych doprowadziły do wyodrębnienia podstawowych kategorii pojęciowych. Przeprowadzona przez niego analiza materiału badawczego (hierarchicznych narzędzi dostępu przedmiotowego w dziewięciu dużych portalach internetowych) umożliwiła identyfikację ośmiu kategorii pojęciowych, które reprezentują klasy jako punkty dostępu. Są to:

1. Przedmiot (ang. subject).
2. Obiekt (ang. object).
3. Aplikacja (ang. application).
4. Użytkownicy (ang. users).
5. Lokalizacja (ang. location)

6. Informator (ang. reference).
7. Medium (ang. media).
8. Język (ang. language).

Kategoria *przedmiot* jest reprezentowana przez jednostki leksykalne odwzorowujące wymiar treści dokumentów internetowych, np. *zdrowie, nauka, filozofia, historia*, itd. Wymiar ten stanowi podstawę strukturalizacji kategorii tematycznych w analizowanych portalach. Kategoria *obiekt* reprezentuje jednostki leksykalne odnoszące się do nazw obiektów rzeczywistości pozajęzykowej, o których informacje zawarte są w dokumentach, np. *muzea, organizacje, stowarzyszenia*. *Aplikacja* jest nazwą kategorii grupującej jednostki leksykalne odnoszące się do specyficznych dla środowiska internetowego procesów i zastosowań, np. *czat, zakupy internetowe, e-mail, wyszukiwarka*. Kategoria *użytkownicy* została wyodrębniona na podstawie jednostek leksykalnych reprezentujących zidentyfikowane grupy użytkowników, dla których przeznaczone są źródła internetowe. Kategoria *lokalizacja* jest reprezentowana przez jednostki leksykalne wskazujące na odniesienie treści obiektu informacyjnego do przestrzeni geograficznej. W kategorii *informator* grupuje się obiekty informacyjne, które są informatorami bezpośrednimi lub pośrednimi. W analizowanych przez Ch. Zinsa przypadkach do tej kategorii zaliczono wyrażenia, które pełnią funkcję nazw klas grupujących odesłania np. do encyklopedii, słowników i katalogów internetowych, ale również elektronicznych czasopism i dokumentów kartograficznych. Kolejna kategoria – *medium*, została wyodrębniona na podstawie kategoryzacji zbioru wyszukiwawczego według rodzaju medium będącego podstawowym nośnikiem obiektu. Ch. Zins do zakresu tej kategorii włącza medium jako nośnik informacji oraz technologię zaangażowaną do jej kodowania i rozkodowania. Kategoria *język* reprezentuje jednostki leksykalne będące nazwami języków etnicznych, w których wyrażono treść dokumentów, np. *język polski, język francuski, język niemiecki*.

Każdą z wyodrębnionych przez Ch. Zinsa kategorii, zwanych przez niego również fasetami przyporządkował on do jednego z dwóch typów charakteryzujących płaszczyzny reprezentacji dokumentu internetowego. Jest to typ zorientowany na treść (ang. content-related), do którego zaliczył kategorie *przedmiot, obiekt, aplikacja, użytkownik* i *lokalizacja* oraz typ zorientowany na format (ang. format-related) czyli *informator, medium* oraz *język*.

Model kategorii Zinsa został zweryfikowany oraz częściowo zmodyfikowany przez zespół badawczy pod kierunkiem K. Kyung-Sun. [K. Kyung-Sun 2006] Przedmiotem badań tego zespołu była struktura katalogów internetowych pięciu wersji

językowych katalogu portalu Yahoo! Na podstawie zgromadzonego materiału badawczego w postaci zbioru jednostek leksykalnych użytych jako nazwy klas grupujących odesłania do dokumentów internetowych dokonano jego analizy pod kątem przynależności do określonej kategorii modelu Zinsa. Tab. 2.5 przedstawia porównanie modelu Zinsa oraz jego zmodyfikowanej wersji opracowanej przez K. Kyung-Sun.

Tab. 2.5. Porównanie modelu Zinsa i jego zmodyfikowanej wersji [Źródło: Kyung-Sun 2006].

Model Zinsa	Wersja zmodyfikowana
Przedmiot	Przedmiot
Aplikacja	Aplikacja
Lokalizacja	Lokalizacja
Medium	Medium/Typ/Format/Narzędzie
Obiekt	Organizacja
Użytkownicy	Ludzie
Informator	Informacje ogólne
Język	Język

Zmodyfikowana wersja modelu Ch. Zinsa zakłada zmianę zakresu oraz pominięcie niektórych kategorii. Sytuacja taka ma miejsce np. w przypadku kategorii *aplikacja*, która tutaj grupuje jednostki leksykalne wykorzystywane do odwzorowania cech produkt/usługa, dostęp płatny/bezpłatny, tryb dostępu online/offline. Tym samym zakres kategorii *medium/typ/format/narzędzie* został zwiększony o dodatkowe kryteria, które u Zinsa włączono do kategorii *aplikacja*.

Budowa i zastosowanie struktur hierarchicznych jako narzędzia organizacji punktów dostępu do kolekcji z wykorzystaniem języków informacyjno-wyszukiwawczych jest złożonym zagadnieniem wymagającym odrębnych badań. W tym miejscu, celem autora było wskazanie głównych sposobów zastosowania języków informacyjnych, w szczególności klasyfikacji, w hierarchicznych narzędziach dostępu przedmiotowego. Interesującym zagadnieniem wymagającym poświęcenia szczególnej uwagi jest wykorzystanie klasyfikacji fasetowych do klasyfikowania zasobów WWW i organizacji dostępu. Wartościowe źródło informacji o projektach, których przedmiotem jest tego typu organizacja wiedzy stanowi zamieszczona w Internecie bibliografia pt. *Putting facet on the Web* autorstwa W. Dentona. [Denton 2003]

Możliwa i uzasadniona wydaje się być interpretacja hierarchicznej organizacji punktów dostępu w narzędziach dostępu przedmiotowego jako projekcji struktury pola tematycznego dla danej kolekcji.

#### 2.7.4. Sieciowe systemy organizacji wiedzy

Narzędziem organizacji informacji wykorzystującym koncepcję przedmiotowych punktów dostępu, stosowanym w środowisku internetowym są sieciowe systemy organizacji wiedzy. Charakterystyka tego narzędzia pozwoli na dostrzeżenie zmian, jakie zachodzą w procesach porządkowania informacji w sieciowym środowisku cyfrowym oraz na udział w tych przeobrażeniach tradycyjnych narzędzi organizacji wiedzy.

Narzędzia porządkowania informacji o treści obiektów cyfrowych noszą nazwę *sieciowych systemów organizacji wiedzy* (ang. networked knowledge organization systems) (SSOW). Koncepcja SSOW jest nierozzerwalnie związana z teorią oraz praktyką systemów organizacji wiedzy stosowanych do porządkowania tradycyjnych dokumentów. Wykorzystanie systemów organizacji wiedzy w środowisku sieciowym jest przedmiotem ożywionej dyskusji badaczy z różnych środowisk m.in. od czasu konferencji poświęconej bibliotekom cyfrowym ACM Digital Libraries w 1998 roku w Pittsburgu w USA. [Hodge 2000] Do 2005 roku tego rodzaju spotkania miały charakter warsztatów naukowych organizowanych w ramach corocznych konferencji The Joint Conference on Digital Libraries w USA (JCDDL) [Tudhope 2005], a w latach 2006-2007 odbywają się podczas European Conference on Digital Libraries zorganizowanych odpowiednio w Hiszpanii i na Węgrzech. Platformą wymiany informacji z tego zakresu stał się m.in. serwis internetowy oraz grupa dyskusyjna Networked Knowledge Organization Systems/Services (<http://nkos.slis.kent.edu/>) umieszczone na serwerze Kent State University's School of Library and Information Science w USA. [Koch 2006; Tudhope 2007]

Sieciowe systemy organizacji wiedzy pełnią funkcje map semantycznych pomocnych w wyszukiwaniu i pozyskiwaniu informacji, które są zagnieżdżone w strukturach serwisów internetowych. [Tudhope, Koch 2003] Metafora mapy użyta w tym stwierdzeniu konotuje zarówno narzędzie pomocne w odnajdywaniu określonych obiektów informacyjnych jak i pewnego rodzaju obraz rzeczywistości odwzorowanej przez nie.

Sieciowe systemy organizacji wiedzy funkcjonują jako narzędzia porządkowania informacji poprzez strukturalizację przestrzeni informacyjnej danej kolekcji. Jest ona projekcją odwzorowywanego fragmentu rzeczywistości dokumentacyjnej za pomocą zbiorów słownictwa o różnym stopniu kontroli. Celem takiego procesu jest osiągnięcie maksymalnego poziomu optymalizacji procesów wyszukiwawczych poprzez

identyfikację i charakterystykę obiektów informacyjnych i włączenie ich w strukturę relewantnych z punktu widzenia użytkownika sieci relacji semantycznych. „Poszukiwanie znaczenia w przestrzeni bez formy” [Brooks 2001] jest zatem istotą funkcjonowania tych narzędzi. Nowe lokalizacje znaczenia w rozszerzalnych technologiach kryją się m.in. w:

- strukturze źródła informacyjnego, pośrednio w znacznikach, bezpośrednio w atrybutach kwalifikatorów jako nośników znaczenia;
- relacjach pomiędzy źródłami informacji, pośrednio w powiązaniach hipertekstowych, bezpośrednio w atrybutach kwalifikatorów jako nośników znaczenia. [Brooks 2001]

W piśmiennictwie z zakresu SSOW funkcjonuje typologia tego rodzaju narzędzi uwzględniająca zakres odwzorowania relacji zachodzących pomiędzy kodyfikowanymi w nich pojęciami. [Sosińska-Kalata 2005a, s. 149] Według takiego kryterium podziału w grupie SSOW można wyodrębnić:

- listy terminów,
- klasyfikacje, kategoryzacje i taksonomie,
- listy relacyjne. [Hodge 2000; Sosińska-Kalata 2005; Woźniak-Kasperek 2006]

Listy terminów reprezentują najprostszy model organizacji wiedzy stosowany w środowisku sieciowym. Narzędzia te przyjmują formę zbiorów słownictwa kontrolowanego służącego do odwzorowywania cech opisywanego fragmentu rzeczywistości za pomocą równoważnościowych struktur monorelacyjnych. Zalicza się tutaj najczęściej wykorzystywane kartoteki wzorcowe, glosariusze (jako listy terminów należących do wąskiej specjalności), słowniki (wykazy terminów wraz z definicjami) oraz indeksy nazw geograficznych.

Cechą drugiej grupy SSOW jest „bezpośrednie oznaczanie związków hierarchicznych pomiędzy pojęciami reprezentowanymi przez terminy lub symbole klasyfikacyjne”. [Sosińska-Kalata 2005a, s. 152] Chociaż zakresy pojęć *klasyfikacja*, *kategoryzacja* i *taksonomia* są ściśle rozróżniane, to w praktyce, co zauważono wcześniej, stosowane są zamiennie. [Woźniak-Kasperek 2006] SSOW należące do tego typu charakteryzuje wykorzystanie struktur hierarchicznych do organizowania pojęć oraz do realizacji dostępu do obiektów informacyjnych czy stref zawartości przez nie reprezentowanych.

Listy relacyjne jako kolejny typ SSOW charakteryzują się niejednorodnością pod względem struktur leżących u podstawy narzędzi tej klasy. Należą do niej tezaury, sieci semantyczne oraz ontologie. Tezaury są bardzo rozpowszechnionym

SSOW realizującym zarówno funkcje organizacyjną jak i wyszukiwawczą. Narzędzie to przeznaczone jest przede wszystkim do realizacji funkcji wyszukiwawczej, co w środowisku sieciowym polega np. na zagnieżdzeniu go w strukturze mechanizmu wyszukiwawczego systemu. Wykorzystanie zasobu leksykalnego tezaursu w postaci struktury węzłów i powiązań hipertekstowych umożliwia eksplorację takiego schematu i pobieranie terminów, które tworzą instrukcję wyszukiwawczą.\* Sieci semantyczne, mapy myśli, a także ontologie wykorzystywane są do porządkowania pojęć i terminów za pomocą wyodrębnionego zbioru relacji i sieci powiązań. Charakteryzuje je różny stopień formalizacji oraz zakres odwzorowywanych relacji. Na temat tego rodzaju narzędzi istnieje duża liczba publikacji [np. Vickery 1997; Qin, Paling 2001; McGuinness 2003; Woźniak-Kasperek 2004; Gliński 2005; Shirky 2005; Gliński 2006] dlatego też na potrzeby pracy ograniczono się wyłącznie do tej krótkiej charakterystyki.

## ***2.8. Podsumowanie***

Przedstawione w rozdziale zagadnienia dotyczące organizacji zasobów World Wide Web oraz strukturalnych narzędzi dostępu do zasobów serwisów i systemów internetowych nie są wyczerpujące. Ograniczona objętość pracy oraz obawa o naruszenie spójności wywodu nie pozwalają na szczegółową analizę tego zagadnienia. Celem tego rozdziału było zidentyfikowanie głównych obszarów badań nad tymi zagadnieniami. Na podstawie koncepcji punktów dostępu przedstawionej w poprzednim rozdziale wskazano rolę języka informacyjno-wyszukiwawczego w organizacji i dostępie do informacji w narzędziach dostępu przedmiotowego oraz mechanizmach wyszukiwawczych. Polega ona na udostępnieniu jego systemu leksykalnego i organizacji pola semantycznego oraz dostosowania tych elementów strukturalnych do identyfikacji punktów dostępu. Hipertekstowy model organizacji informacji, będący immanentną cechą zasobów WWW, determinuje możliwości wizualizacji schematów organizacyjnych w narzędziach dostępu. Kluczem do efektywności procesów organizacji informacji i organizacji wiedzy jest przede wszystkim elastyczność struktur porządkujących, którą w warstwie technicznej udostępnia hipertekst, a która na poziomie konceptualnym jest możliwa do odwzorowania poprzez wprowadzenie jednej z metod organizacji wiedzy.

---

\* Nowe sposoby wykorzystania tezaursów w środowisku sieciowym przedstawia Barbara Sosińska-Kalata. [Sosińska-Kalata 2006]



### **3. Charakterystyka i budowa dziedzinowego systemu hipertekstowego**

Tematem rozważań trzeciego rozdziału rozprawy jest charakterystyka dziedzinowego systemu hipertekstowego (DSH). Na podstawie przyjętej na potrzeby pracy definicji tego rodzaju systemów zostaną scharakteryzowane jego elementy strukturalne. Celem rozdziału jest również weryfikacja hipotezy o przynależności dziedzinowych systemów hipertekstowych do klasy dokumentacyjnych systemów informacyjno-wyszukiwawczych. Przedstawiona wieloaspektowa analiza tego typu systemu informacyjno-wyszukiwawczego pozwoli na konstrukcję modelu analitycznego dziedzinowego systemu hipertekstowego.

#### ***3.1. Dziedzinowy system hipertekstowy - definicje***

W literaturze przedmiotu występuje wiele definicji terminu *dziedzinowy system hipertekstowy* oraz wiele nazw mających w jego postaci wspólny desygnat. Są to m.in.:

- *brama dziedzinowa* (ang. subject gateway, subject-based gateway) [Brümmer 1998; Worsfold 1998],
- *brama dziedzinowa o kontrolowanej jakości* (ang. quality-controlled subject gateway) [Koch 2000],
- *drzewo tematyczne* (ang. subject trees) [Chowdhury 1999; Wheatley 2000],
- *mechanizm naprowadzający* (pathfinder) [Chowdhury 1999],
- *tematyczna brama informacyjna* (ang. subject-based information gateway). [Monopoli, Bawden 2000]

Dziedzinowy system hipertekstowy wykazuje kilka cech wyróżniających go spośród innych sieciowych systemów informacyjnych. Są to m.in.:

- zakres systemu ograniczony jest do jednej i/lub kilku dziedzin wiedzy, czy też sfer działalności ludzkiej,
- baza danych stanowiąca zbiór wyszukiwawczy systemu zbudowana jest z rekordów w postaci charakterystyk wyszukiwawczych obiektów informacyjnych w zasobach WWW,
- użytkownik na wyjściu z systemu otrzymuje rezultaty wyszukiwania w postaci charakterystyk wyszukiwawczych dokumentów wraz z odesłaniami hipertekstowymi do ich lokalizacji w Internecie,

- odwzorowanie cech formy i treści dokumentów internetowych jest wykonywane przez specjalistów dziedzinowych, nie zaś generowane automatycznie.

Wymienione cechy dają obraz usługi, której zadaniem jest zaspokojenie potrzeb informacyjnych użytkownika poszukującego informacji z określonej dziedziny wiedzy poprzez wskazanie mu w Internecie lokalizacji relewantnych dokumentów.

Celem przytoczenia różnych definicji DSH jest wskazanie zarówno rozbieżności w jego interpretowaniu oraz zidentyfikowanie ich cech wspólnych. Za punkt wyjścia do rozważań przyjęto definicje najbardziej rozpowszechnionej nazwy dla tego modelu udostępniania informacji w Internecie, czyli wyrażenia *brama dziedzinowa*.

W projekcie „Reynard: Academic Subject Gateway Service Europe” znanym pod nazwą *Renardus*, którego celem było stworzenie europejskiej sieci akademickiej dziedzinowych bram informacyjnych zamieszczono następującą definicję tej usługi:

*Tematyczne przewodniki po zasobach udostępniające odesłania hipertekstowe do źródeł internetowych (dokumentów, stron internetowych i elektronicznych usług informacyjnych). Cechą charakterystyczną tych przewodników jest opis zasobów oraz ich klasyfikacja przedmiotowa.* [Gateways Defined]

Autorzy projektu badawczego „IMesh Toolkit” funkcjonującego w latach 1999-2001, którego jednym z celów było opracowanie oprogramowania dla dziedzinowych systemów hipertekstowych, podają w dokumentacji projektu następującą charakterystykę tego terminu:

*Brama dziedzinowa to serwis internetowy, który realizuje dostęp (poprzez wyszukiwanie i przeglądanie) do zasobów online z określonego zakresu przedmiotowego. Opisy udostępniane przez bramy dziedzinowe są sporządzane manualnie, rzadziej generowane.* [IMesh Toolkit]

Termin *brama dziedzinowa* został zdefiniowany również przez Australian Subject Gateways Forum, instytucję o zasięgu narodowym w Australii, której celem jest m.in. opracowywanie strategii rozwoju i określania kierunków badań dotyczących katalogowania zasobów Internetu:

*Mechanizm sieciowy umożliwiający dostęp do ocenianych przez specjalistów, wysokiej jakości źródeł informacji (elektronicznych) zidentyfikowanych jako istotne, z punktu widzenia badań naukowych w konkretnej dziedzinie wiedzy.*

Uproszczona definicja bramy dziedzinowej występująca w dokumentacji tej instytucji brzmi:

*Internetowa kolekcja opisów źródeł i informacji o ich lokalizacji z danego zakresu przedmiotowego, zorganizowana według układu przedmiotowego lub*

*dyscyplinarnego, gdzie o włączeniu do niej decyduje opublikowany zbiór kryteriów selekcji. [Australian Subject Gateways: Definitions]*

lub:

*Mechanizm webowy oparty na udostępnianiu kolekcji wysokiej jakości zasobów sieciowych rozpoznanych jako znaczące dla badań w danej dyscyplinie. [Campbell 2000]*

W projekcie Unii Europejskiej pod nazwą *DESIRE - Development of a European Service for Information on Research and Education* zagadnieniom dziedzinowego systemu hipertekstowego poświęcono wiele uwagi. Do najważniejszych publikacji z okresu funkcjonowania tego projektu należy niewątpliwie praca zbiorowa pt. „*Information Gateways Handbook*”. [Belcher i in. 1999] Ma ona formę podręcznika, który jest poświęcony metodyce projektowania i wdrażania bram informacyjnych. Chociaż w publikacji tej, autorzy posługują się wyrażeniem *brama informacyjna*, to prezentowane tam informacje dotyczą *de facto* dziedzinowych systemów hipertekstowych. Dokumentacja projektu DESIRE zawiera następującą charakterystykę *dziedzinowych bram informacyjnych*, która brzmi:

*W przypadku bram dziedzinowych, świadomie kładzie się nacisk na zaangażowanie wykwalifikowanego czynnika ludzkiego w ocenie jakości informacji wybranych źródeł informacyjnych w Internecie. Podstawowe działanie, czyli selekcja oraz odwzorowanie metainformacji (tutaj ang. attributing meaning – przypisywanie znaczenia) jest wykonywane manualnie. [Hofman, Worsfold 1997]*

Anna Brümmer, która przeprowadziła szereg badań nad wybranymi bramami dziedzinowymi scharakteryzowała ten rodzaj usługi informacyjnej dostępnej w Internecie następująco:

*Bramy informacyjne oparte na dziedzinowym dostępie do informacji mają charakter wejść tematycznych do zasobów internetowych ocenianych pod kątem jakości informacji. [...] Zasoby bram dziedzinowych poddawane są selekcji według przyjętych i opublikowanych kryteriów oceny jakości informacji oraz poddawane procesowi katalogowania. Dodatkowo, ich zasoby są katalogowane z wykorzystaniem schematów klasyfikacyjnych. Odpowiednie symbole klasyfikacyjne, przechowywane w rekordach bazy danych, mogą być wykorzystane do tworzenia narzędzi dostępu przedmiotowego. [...] W porównaniu do zwykłych list z odesłaniami hipertekstowymi do zasobów Internetu, wdrożenie i zarządzanie bramami dziedzinowymi jest pracochłonne.*

*Jednocześnie wysoka jakość informacyjna kolekcji oraz standaryzowane charakterystyki jej elementów pozwalają użytkownikowi na obiektywną ocenę udostępnianych tam informacji.* [Brümmer 2000]

Najważniejszą pracą, z punktu widzenia uporządkowania terminologii, dotyczącą dziedzinowych systemów hipertekstowych stanowi artykuł autorstwa Traugutta Kocha opublikowany w zeszycie czasopisma „Online Information Review”, który w całości poświęcono temu zagadnieniu. [Koch 2000] Informacje zawarte w tym artykule stanowią od daty jego wydania, czyli od 2000 r., podstawowe źródło, które cytuje większość badaczy. T. Koch rozróżnia znaczenie wyrażen *brama dziedzinowa* (ang. subject gateway) i *brama dziedzinowa o kontrolowanej jakości* (ang. quality-controlled subject gateway). Dla pierwszego wyrażenia proponuje definicję:

*Brama dziedzinowa jest usługą internetową opartą na systematycznym odkrywaniu (ang. systematic discovery) jego zasobów. Udostępnia odesłania hipertekstowe do zasobów (dokumenty, obiekty informacyjne, strony internetowe i inne) dostępnych poprzez Internet. Funkcjonowanie takich systemów oparte jest na charakterystyce jego zasobów. Ważną funkcją realizowaną przez nie jest dostęp przedmiotowy do kolekcji.*

Według T. Kocha brama dziedzinowa jest usługą informacyjną, gdzie kataloguje się rozproszone zasoby Internetu. Na wyjściu użytkownik otrzymuje jednak ich surogaty - charakterystyki wyszukiwawcze obiektów informacyjnych. Bramy dziedzinowe funkcjonują w oparciu o sporządzane manualnie opisy wyselekcjonowanych zasobów Internetu. T. Koch zaznacza, że istnieją dodatkowe kryteria nieuwzględnione w definicji, które umożliwiają zidentyfikowanie tego typu serwisów hipertekstowych. Są to:

- a) cel działania systemu, który zazwyczaj służy potrzebom nauki, nauczania i kultury,
- b) zakres systemu – ograniczony zazwyczaj do jednej dziedziny wiedzy (choć funkcjonują również DSH o zakresie uniwersalnym),
- c) przyjęcie określonego poziomu analizy obiektu informacyjnego (np. serwis internetowy, strona internetowa, rekord bazy danych) i ustalenie formatu opisu dla poszczególnych rodzajów elementarnej jednostki opisu,
- d) finansowanie systemu – zazwyczaj przez instytucje z sektora publicznego.

Autor ten uważa, że definicja bramy dziedzinowej jest nieostra i oprócz kryterium zakresu oraz warunku charakterystyki treściowej i formalnej zasobów nie podaje więcej cech wyróżniających. Wprowadził on nowy termin *brama dziedzinowa o kontrolowanej*

*jakości* w celu rozróżnienia tych dwóch usług. Zaproponował następującą interpretację:

- a) *brama dziedzinowa* - lista odesłań hipertekstowych wzbogacona o elementy opisu wraz z uproszczonym narzędziem dostępu przedmiotowego,
- b) *brama dziedzinowa o kontrolowanej jakości* - system wykorzystujący efektywne standardy kontroli jakości udostępnianych informacji i procedury systemowe wraz z wyczerpującą charakterystyką formalną i treściową zasobów oraz rozbudowanym narzędziem dostępu przedmiotowego.

Taki punkt widzenia zakłada, że brama dziedzinowa o kontrolowanej jakości stanowi podklasę internetowych systemów informacyjnych zwanych bramami dziedzinowymi. Jej definiowanie przez naukowców zajmujących się tym zagadnieniem [np. Kirremuir 1999; Heery 2000; Koch 2000; Neuroth, Koch 2001] polegało przede wszystkim na identyfikacji warunków, jakie musi spełniać hipertekstowy system informacyjny, aby można było określić go mianem bramy dziedzinowej o kontrolowanej jakości. Są to [Roszkowski 2005 za Koch 2000]:

1. **Tworzenie** tego rodzaju systemów zakłada funkcjonowanie zespołu osób z przypisanymi do nich określonym zakresem obowiązków (specjaliści dziedzinowi, redaktorzy, osoby odpowiedzialne za indeksowanie, itd.).
2. **Rozwój kolekcji** jest oparty na polityce gromadzenia i selekcji indeksowanych zasobów z uwzględnieniem ich jakości oraz dążeniem do kompletności.
3. Odpowiednie **zarządzanie kolekcją** ma na celu zapewnienie aktualności odesłań do dokumentów internetowych poprzez m.in. sprawdzanie aktualności ich adresów URL, aktualizację charakterystyk wyszukiwawczych (w razie zmian w zawartości dokumentów), usuwanie odesłań do dokumentów nieistniejących.
4. **Charakterystyka zasobów** jest jednym z najważniejszych procesów podczas tworzenia i utrzymania dziedzinowych systemów hipertekstowych. Zakłada ona:
  - tworzenie lub przejmowanie zestawu metadanych, przystosowanego do rodzaju opisywanych zasobów,
  - zgodność z funkcjonującymi standardami metadanych,
  - odpowiednie wyświetlanie pól wraz z ich zawartością użytkownikowi,
  - formalizację oraz dążenie do spójności stworzonego lub przejętego modelu opisu zasobów,
  - odwzorowanie treści i formy obiektów informacyjnych z wykorzystaniem języków informacyjno-wyszukiwawczych.

5. **Dostęp przedmiotowy** polega na wykorzystaniu zbioru kategorii tematycznych, których nazwy mają zazwyczaj formę wyrażen języka naturalnego. Grupują one charakterystyki wyszukiwawcze dokumentów spełniające kryteria zgodności zakresowej. Narzędzie to jest zagnieżdżone w strukturze hipertekstu, co umożliwia użytkownikowi jego eksplorację w celu odnalezienia kategorii, w której zgromadzono relewantne ChWD. W zależności od przyjętych dla danego systemu założeń, zbiór kategorii ma zazwyczaj formę struktury hierarchicznej lub ahierarchicznej (w porządku alfabetycznym).
6. **Standardy.** Zastosowanie funkcjonujących rozwiązań przyjętych za standardowe ma na celu umożliwienie współdziałania (ang. interoperability) wielu systemów np. na płaszczyźnie wymiany danych.

Badania przeprowadzone w 1999 roku przez Johna Kirremuira [Kirremuir 1999] nad bramami dziedzinowymi należą do pierwszych podejmujących próby systematyzacji wiedzy na ten temat. Charakterystyka bram dziedzinowych, które nazwał on *systemami wyszukiwania zasobów o wysokiej jakości* (ang. quality resource discovery system) była oparta na opracowanych kryteriach decydujących o włączeniu analizowanego serwisu tematycznego do grupy reprezentującej ten model. Były to: [Kirremuir 1999]

- umieszczanie obok odesłania hipertekstowego jego charakterystyki zawierającej zestaw metadanych,
- kontrola jakości udostępnianych przez system informacji, która miała polegać na posługiwaniu się określonymi zasadami selekcji źródeł informacji oraz instrukcjami katalogowania zasobów Internetu,
- współpraca z instytucjami naukowymi lub specjalistami zarządzającymi systemem i odpowiedzialnymi za jego kolekcję,
- regularne sprawdzanie poprawności odesłań hipertekstowych oraz monitorowanie zmian w treści i formie skatalogowanych zasobów sieciowych,
- znacząca wielkość zbioru informacyjnego (w przypadku tego kryterium J. Kirremuir włączył do badania systemy, które opisywały minimum kilkaset źródeł internetowych).

Przytoczone definicje oraz ich analizy charakteryzują dziedzinowy system hipertekstowy w kategorii usługi informacyjnej, której celem jest udostępnienie użytkownikowi informacji o relewantnych zasobach Internetu, poddanych ocenie jakościowej, których charakterystyki wyszukiwawcze są tworzone przez wyspecjalizowany zespół osób zarządzających zasobem systemu zgodnie

z wyznaczoną polityką selekcji, gromadzenia, opracowania i udostępniania informacji. Zaprojektowany zestaw metadanych oraz wykorzystanie języka informacyjno-wyszukiwawczego do reprezentacji cech formalnych i treściowych zasobów internetowych gwarantuje wysoką jakość informacji udostępnianych użytkownikom.

To, co odróżnia bramy dziedzinowe o kontrolowanej jakości informacji od np. katalogów internetowych (np. Yahoo!, czy katalogu portalu Onet.pl) to przede wszystkim [Koch 2000]:

- zaangażowanie specjalistów dziedzinowych, tj. osób odpowiedzialnych za zarządzanie projektem (ciężar i odpowiedzialność za efektywność bram dziedzinowych spoczywa na konkretnych osobach, które realizują powierzone im zadania),
- respektowanie polityki zarządzania kolekcją, która zakłada m.in. szczegółowe zasady selekcji i katalogowania zasobów,
- odwzorowywanie cech formalnych i treściowych w postaci charakterystyk wyszukiwawczych oparte jest na funkcjonujących standardach dotyczących katalogowania dokumentów elektronicznych oraz formatach ich sporządzania.

Według T. Kocha wymieniona charakterystyka dziedzinowych bram informacyjnych o kontrolowanej jakości pozwala na zbudowanie definicji, która brzmi:

*Dziedzinowe bramy informacyjne o kontrolowanej jakości są usługami internetowymi, które polegają na stosowaniu zbioru wskaźników jakościowych w celu metodycznego pozyskiwania informacji o zasobach internetowych. Czynniki ludzki wykorzystywane jest w procesie selekcji zasobów (opartej na kryteriach decydujących o włączeniu źródła do zbioru informacyjnego) oraz sporządzania wyczerpujących opisów tych zasobów z wykorzystaniem standaryzowanych zbiorów metadanych. Efektywne zarządzanie kolekcją polega m.in. na regularnym sprawdzaniu poprawności adresów URL zasobów internetowych oraz na ich aktualizowaniu. Zasadniczym celem tej usługi jest umożliwienie użytkownikowi dostępu do zasobów, zaindeksowanych za pomocą zbiorów słownictwa kontrolowanego, poprzez wykorzystanie efektywnych narzędzi dostępu przedmiotowego w postaci m.in. głębokich struktur klasyfikacyjnych. [Koch 2000]*

Taka definicja dziedzinowej bramy informacyjnej o kontrolowanej jakości będzie stanowiła podstawę do rozumienia tego terminu w dalszej części pracy. Pozostaje jeszcze wskazanie przesłanek do stosowania nazwy *dziedzinowy system hipertekstowy*. Jak już wspomniano, w literaturze przedmiotu funkcjonuje wiele nazw w języku

angielskim, które mają wspólny desygnat w postaci omawianego typu systemów. Istnieje zatem potrzeba stworzenia lub przejęcia z języka obcego jednej ogólnie akceptowanej nazwy. W polskim piśmiennictwie naukowym rzadko podejmowano zagadnienia związane z tymi systemami [Sosińska-Kalata 2002; Derfert-Wolf 2004; Roszkowski 2007a], co nie sprzyjało ugruntowaniu pozycji dla jednej preferowanej nazwy. Tłumaczenie wyrażen *brama tematyczna* (ang. subject gateway), *brama tematyczna o kontrolowanej jakości* (ang. quality-controlled subject gateway) czy też *centra informacyjne* (ang. hub) nie przyjęły się w polskiej terminologii z tego zakresu i nie utożsamiają tego modelu z jednym z rodzajów internetowych systemów informacyjno-wyszukiwawczych. Dlatego też, zdaniem autora wyrażenie *dziedzinowy system hipertekstowy* nie ma charakteru kalki językowej oraz co najważniejsze charakteryzuje jego podstawowe cechy, utożsamiając jego desygnaty z rodzajem systemów informacyjno-wyszukiwawczych. Zatem w tekście pracy do oznaczania typu systemu informacyjno-wyszukiwawczego, którego definicje przedstawiono w tej części, stosowane będzie wyrażenie *dziedzinowy system hipertekstowy*.

### **3.2. Dziedzinowy system hipertekstowy jako rodzaj systemu informacyjno-wyszukiwawczego**

Charakterystyka dziedzinowego systemu hipertekstowego, przedstawiona w przytoczonych definicjach, umożliwi identyfikację tego rodzaju usługi internetowej jako systemu informacyjno-wyszukiwawczego (SIW) w rozumieniu definicji tego terminu stosowanej w informacji naukowej. Tym samym stało się możliwe określenie jego miejsca w typologii systemów informacyjno-wyszukiwawczych. Zdaniem autora istnieją przesłanki do przyjęcia takiej interpretacji. Wyrażenie *usługa*, jakiego używa się wobec DSH, deprecjonuje znaczenie oraz nakład pracy, jaki konieczny jest do zaprojektowania, wdrożenia i zarządzania takim systemem, jak również wagę wykorzystanych metodologii i narzędzi. Kolejny argument przemawiający za taką interpretacją to spełnianie przez DSH warunków definicyjnych wyrażenia *system informacyjno-wyszukiwawczy*. Za punkt wyjścia dla potwierdzenia tego założenia przyjęto definicję terminu *system informacyjno-wyszukiwawczy* zamieszczoną w „Słowniku encyklopedycznym informacji, języków i systemów informacyjno-wyszukiwawczych”, która brzmi:

*System przetwarzający informacje [...], który z informacji wejściowych, obecnie często w postaci dokumentów, tworzących zbiór informacyjny, tworzy odpowiednio*



*ustrukturalizowany zbiór wyszukiwawczy poprzez odpowiednie transformacje, a następnie dokonuje na nim dalszych transformacji, a więc wyszukiwania informacji na podstawie określonej dla danego systemu relacji relewancji technicznej. Informacje spełniające warunek relewancji technicznej tworzą zbiór informacji wyjściowej i udostępniane są użytkownikowi systemu. Zależnie od rodzaju informacji wyjściowej, a co za tym idzie i rodzaju transformacji tworzących z informacji wejściowej zbiór informacyjny, wyróżnia się systemy informacji dokumentacyjnej i systemy informacji faktograficznej.* [Słownik encyklopedyczny 2002, s. 261-262]

Analiza znaczenia tego wyrażenia przeprowadzona przez E. Artowicz [Artowicz 1997, s. 9] wskazała, iż można w odniesieniu do SIW wyróżnić jego cechy dystynktywne, konieczne i wystarczające do odróżnienia ich od innych systemów. Są to:

- zbiór informacyjny o określonej strukturze,
- język informacyjno-wyszukiwawczy (JIW),
- transformacja (przetwarzanie informacji).

W odniesieniu do omawianego modelu systemu hipertekstowego można przyjąć, że:

- jego zbiór informacyjny stanowią wyselekcjonowane internetowe obiekty informacyjne spełniające odpowiednie kryteria polityki ich gromadzenia,
- język informacyjno-wyszukiwawczy jest narzędziem pełniącym funkcję metainformacyjną, wyszukiwawczą oraz organizującą. W DSH jego wykorzystanie przybiera charakterystyczne formy, co zostanie przedstawione w dalszej części rozprawy,
- transformacja informacji dokonywana jest według przyjętych dla danego systemu zasad.

Powyższe stwierdzenia pozwalają na pełnoprawne zastosowanie terminu *system informacyjno-wyszukiwawczy* wobec dziedzinowych systemów hipertekstowych oraz zakwalifikowanie DSH do typu dokumentacyjnych systemów informacyjno-wyszukiwawczych. Autorzy *Słownika encyklopedycznego informacji, języków i systemów informacyjno-wyszukiwawczych* pod tym wyrażeniem zamieścili następującą definicję:

*System informacji dokumentacyjnej - system informacyjno-wyszukiwawczy, którego zbiór wyszukiwawczy składa się z dokumentów pochodnych, najczęściej w postaci charakterystyk wyszukiwawczych dokumentów. W systemie takim informacje wyjściowe są więc metainformacjami, informującymi użytkownika systemu o tym, jakie dokumenty odpowiadają potrzebom informacyjnym tego użytkownika. System*

*informacji dokumentacyjnej informacje włączane do zbioru wyszukiwawczego może tworzyć samodzielnie, wówczas informacją wejściową jest zbiór dokumentów podlegających opracowaniu dokumentacyjnemu w samym systemie, a więc gdy na podstawie informacji wejściowych tworzy się zbiór metainformacji w postaci charakterystyk wyszukiwawczych dokumentów poprzez transformacje streszczania, wyboru słów kluczowych i indeksowanie przy użyciu odpowiedniego JIW, przy czym streszczanie jest transformacją fakultatywną. [Słownik encyklopedyczny 2002, s. 261]*

Interpretacja tej definicji w odniesieniu do DSH, których zbiór wyszukiwawczy tworzą charakterystyki wyszukiwawcze dokumentów internetowych, a na wyjściu użytkownik otrzymuje jego wyselekcjonowany podzbiór spełniający kryteria relewancji z dodatkową informacją o lokalizacji (URL) obiektów informacyjnych w Internecie, pozwala na pozytywną weryfikację hipotezy o przynależności tego typu systemów do dokumentacyjnych systemów informacyjno-wyszukiwawczych.

### ***3.3. Struktura dziedzinowego systemu hipertekstowego***

Analizując dziedzinowe systemy hipertekstowe pod kątem budowy oraz procesów, jakie towarzyszą ich projektowaniu i funkcjonowaniu można wyodrębnić elementy, które są ze sobą ściśle powiązane i razem tworzą, o ile są poprawnie zaprojektowane i wdrożone, spójny i efektywny system informacyjno-wyszukiwawczy. Elementy te zostały zidentyfikowane podczas definiowania dziedzinowego systemu hipertekstowego. Zdaniem autora, należy jednak poświęcić temu zagadnieniu więcej uwagi, aby wskazać jakie rozwiązania zastosowano w odniesieniu do konkretnych, funkcjonujących w Internecie dziedzinowych systemów hipertekstowych. Zatem w tym miejscu zostaną scharakteryzowane:

- cel i zakres systemów,
- rozwój i zarządzanie kolekcją,
- charakterystyka treściowa i formalna zasobów,
- dostęp przedmiotowy.

#### **3.3.1. Cel i zakres systemów**

Dziedzinowe systemy hipertekstowe są przeważnie tworzone i zarządzane przez biblioteki oraz instytucje związane ze szkolnictwem, najczęściej na poziomie wyższym.

[Koch 2000] W niektórych przypadkach w tego rodzaju projekty zaangażowane są także instytucje rządowe (np. OCLC Connexion, którego prototypem był OCLC CORC [Hickey 2000]; EdNA Online - Education Network Australia; GEM - Gateway to 21st Skills). Pomimo niewielkich rozbieżności w kwestii rodzaju instytucji sprawczej, grupa docelowa użytkowników tego rodzaju systemu informacyjno-wyszukiwawczego jest w większości przypadków taka sama. Jest to sektor związany z oświatą i szkolnictwem na różnych poziomach, zarówno w skali lokalnej jak i globalnej.

Państwa, w których powstały najbardziej uznane systemy to: Wielka Brytania, Stany Zjednoczone, Finlandia, Szwecja, Niemcy oraz Australia. [Koch 2000] W przypadku systemów wywodzących się z krajów nieanglojęzycznych istnieje wyraźna tendencja do wprowadzania dwujęzyczności (język angielski i narodowy) zarówno interfejsu użytkownika jak i języka indeksowanych zasobów (np. Science Linkhouse - język angielski i fiński, MathGuide - język angielski i niemiecki).

Zakres systemu, a dokładnie jego poprawne sformułowanie jest jednym z warunków spójności jego zbioru informacyjnego. W przypadku dziedzinowych systemów hipertekstowych zakres systemu jest zazwyczaj określony w ogólnodostępnej dokumentacji danego systemu. Często sama nazwa systemu lub rozwinięcie jej akronimu jest źródłem informacji o podstawowym zakresie zbioru informacyjnego. Np.:

- SOSIG - Social Science Information Gateway - nauki społeczne,
- OMNI - Organising Medical Networked Information - medycyna,
- EEVL Internet Guide to Engineering, Mathematics and Computing - inżynieria, matematyka i technika komputerowa.

Zakres zbioru informacyjnego dziedzinowego systemu hipertekstowego wyznaczany jest przez własności treści obiektów informacyjnych go tworzących. Jego pole tematyczne jest projektowane i w założeniu jego zakres ma odpowiadać zakresowi potrzeb informacyjnych grupy użytkowników docelowych. Ze względu na zakres pola tematycznego DSH należą one, w rozumieniu definicji wyrażenia *pole tematyczne* [s. 69 rozprawy lub Słownik encyklopedyczny 2002 s. 197-198], do rodzaju dziedzinowych systemów informacyjno-wyszukiwawczych. Zdefiniowanie zakresu DSH polega przeważnie na zidentyfikowaniu tematów szczegółowych, których suma zakresów składa się na jego pole tematyczne. Np. system VetGate\* z zakresu weterynarii i opieki nad zwierzętami jest określony w jego dokumentacji poprzez wyszczególnienie następujących tematów szczegółowych\*\*:

---

\* obecnie włączony do systemu Intute - <http://www.intute.ac.uk/healthandlifesciences/veterinary/>

\*\* [http://www.intute.ac.uk/healthandlifesciences/Intute\\_HLS\\_Collections\\_Policy.doc](http://www.intute.ac.uk/healthandlifesciences/Intute_HLS_Collections_Policy.doc)

- anatomia zwierząt,
- choroby zwierzęce,
- szpitale dla zwierząt,
- hodowla zwierząt,
- żywienie zwierząt,
- fizjologia zwierząt,
- toksykologia zwierząt,
- opieka nad zwierzętami,
- towarzystwa działające na rzecz zwierząt,
- bakteriologia i wirusologia zwierzęca,
- kształcenie weterynaryjne,
- epidemiologia weterynaryjna,
- genetyka weterynaryjna,
- histologia weterynaryjna,
- immunologia weterynaryjna,
- medycyna weterynaryjna,
- neurologia weterynaryjna,
- patologia weterynaryjna,
- farmakologia weterynaryjna,
- radiologia weterynaryjna,
- chirurgia weterynaryjna.

W przypadku systemów o szerokim zakresie zbioru informacyjnego wykaz tematów szczegółowych może liczyć nawet kilkaset pozycji.

### **3.3.2. Rozwój i zarządzanie kolekcją**

Jednym z warunków definicyjnych dziedzinowych systemów hipertekstowych jest stosowanie opracowanych kryteriów oceny i selekcji zasobów Internetu, które stanowią elementy zbioru informacyjnego danego systemu. Zbiór takich wytycznych zwany jest polityką zarządzania kolekcją (ang. collection policy, scope policy, lub collection management policy). Informacje o polityce zarządzania danym systemem stanowią część jego dokumentacji. Według T. Kocha [Koch 2000] i J. Kirremuira [Kirremuir 1999] polityka zarządzania dziedzinowym systemem hipertekstowym powinna uwzględniać:

- opracowanie mechanizmów regularnego sprawdzania aktualności i poprawności odesłań hipertekstowych do zaindeksowanych zasobów sieciowych,
- kontrolę adekwatności i aktualności informacji zawartych w charakterystykach wyszukiwawczych (monitorowanie zmian w treści obiektów informacyjnych i aktualizacja odpowiadających im ChWD w systemie),
- okresową aktualizację zbioru informacyjnego polegającą na zastępowaniu odesłań do obiektów zawierających informacje nieaktualne (z przyjętego punktu widzenia), informacjami o nowozaindeksowanych źródłach internetowych.

Dokumentacja dziedzinowych systemów hipertekstowych w zakresie polityki zarządzania zawiera najczęściej:

- określenie zakresu systemu,
- określenie grupy użytkowników docelowych,

- zidentyfikowanie rodzaju indeksowanych zasobów WWW,
- opracowanie kryteriów oceny jakości zasobów, których spełnienie decyduje o włączeniu do zbioru informacyjnego.

Dla określenia grupy użytkowników stosuje się często ich szczegółową identyfikację. W Tab. 3.1. zaprezentowano specyfikację grup użytkowników docelowych dla kilku wybranych dziedzinowych systemów hipertekstowych.

Tab. 3.1. Charakterystyka grup użytkowników docelowych w wybranych dziedzinowych systemach hipertekstowych. [Źródło: [http://www.intute.ac.uk/Intute\\_cdfp.doc](http://www.intute.ac.uk/Intute_cdfp.doc)]

Nazwa systemu	Charakterystyka grupy użytkowników docelowych
Organising Medical Networked Information ( <a href="http://www.omni.ac.uk">www.omni.ac.uk</a> ) – medycyna	<ul style="list-style-type: none"> <li>- lekarze, osoby czynne zawodowo z sektora zdrowotnego w Wielkiej Brytanii (szczególnie z publicznej służby zdrowia);</li> <li>- pacjenci oraz wszyscy użytkownicy, którzy potrzebują informacji medycznej;</li> <li>- przedstawiciele przemysłu farmaceutycznego;</li> <li>- naukowcy, nauczyciele akademicki, studenci, związani szeroko pojętą medycyną oraz nauką o zdrowiu;</li> </ul>
VetGate ( <a href="http://www.vetgate.ac.uk">www.vetgate.ac.uk</a> ) - weterynaria	<ul style="list-style-type: none"> <li>- naukowcy, wykładowcy i studenci weterynarii;</li> <li>- weterynarze;</li> <li>- pomoc weterynaryjna oraz inne zawody związane;</li> <li>- użytkownicy zaangażowani w hodowlę zwierząt oraz ich zdrowie;</li> <li>- producenci, dostawcy produktów i usług związanych z weterynarią;</li> </ul>
AgriFor ( <a href="http://www.agrifor.ac.uk">www.agrifor.ac.uk</a> ) - rolnictwo	<ul style="list-style-type: none"> <li>- naukowcy, nauczyciele akademicki, studenci, związani z rolnictwem, leśnictwem oraz nauką o żywieniu;</li> <li>- rolnicy oraz inni, którzy zajmują się hodowlą zwierząt;</li> <li>- grupy związane z produkcją żywności oraz hodowlą trzody chlewnej;</li> <li>- przedstawiciele przemysłu leśnego;</li> <li>- producenci i dostawcy zasobów rolniczych jak również usług informacji z tego zakresu.</li> </ul>

W wielu przypadkach określenie grupy użytkowników docelowych zamieszczone jest w sekcji dokumentacji dotyczącej definiowania zakresu danego systemu.

Opracowanie wykazu typów indeksowanych przez system zasobów WWW ma znaczenie w perspektywie opracowania narzędzi organizacji informacji, w szczególności formatu opisu danych. Polega to na wyszczególnieniu akceptowanych rodzajów obiektów informacyjnych włączanych do zasobów DSH. Najczęściej jest to zbiór wyrażen języka naturalnego przyjętych do określania formy elementów zbioru informacyjnego podczas indeksowania (Tab. 3.2).

Tab. 3.2. Rodzaj zasobów w systemach SOSIG i OMNI. [Źródło: opracowanie własne na podstawie dokumentacji systemów]

SOSIG – z zakresu nauk społecznych	OMNI – z zakresu medycyny i nauki o zdrowiu
- artykuły/raporty – pojedyncze,	- bibliograficzne bazy danych,
- artykuły/raporty – zbiory,	- czasopisma – pełne teksty,
- bibliograficzne bazy danych,	- czasopisma – spisy treści oraz abstrakty,
- bibliografie,	- książki,
- czasopisma – pełne teksty,	- listy mailingowe,
- czasopisma – spisy treści oraz abstrakty,	- materiały dydaktyczne,
- dane,	- niebibliograficzne bazy danych,
- dokumenty – legislacja,	- opinie medyczne,
- dokumenty – porozumienia, układy,	- oprogramowanie,
- dokumenty – raporty prawnicze,	- przeglądy wyników badań (systematic reviews),
- dokumenty – streszczenia,	- przewodniki po zasobach / gateways,
- FAQ,	- raporty, dokumentacje,
- Informatory,	- serwisy informacyjne,
- serwisy instytucji rządowych,	- strony organizacji,
- książki,	- ulotki informacyjne dla pacjentów,
- listy mailingowe,	- wytyczne dot. praktyki klinicznej,
- materiały dydaktyczne,	
- oprogramowanie,	
- organizacje/społeczności,	
- serwisy projektów badawczych,	
- przewodniki po zasobach,	
- publikacje rządowe,	
- serwisy informacyjne,	
- towarzystwa/spółki,	

Przedstawione w tabeli 3.2 porównanie rodzaju indeksowanych zasobów przez dwa DSH wskazuje na różne perspektywy postrzegania potencjalnej elementarnej jednostki opisu. Mamy tutaj do czynienia z zastosowaniem podziału zasobów według ich formy,

przeznaczenia (np. materiały dydaktyczne) oraz specyficzne typy w postaci serwisów instytucji rządowych czy projektów badawczych. W większości przypadków opracowanie rodzaju indeksowanych zasobów sprowadza do stworzenia listy od kilku do kilkunastu wyrażeń stanowiących nazwy dla wyodrębnionych typów.

Kryteria oceny indeksowanych zasobów zarówno pod kątem ich zwartości oraz formy stanowią cechę wyróżniającą te systemy spośród dużej liczby usług informacyjnych dostępnych w środowisku sieciowym. Kryteria te, wykorzystywane są w procesie gromadzenia i selekcji informacji w dziedzinowych systemach hipertekstowych. Zagadnieniu temu poświęcono szczegółowy raport z badań w ramach pierwszego etapu projektu DESIRE. Rozwiązania wypracowane przez autorów tego raportu są szeroko wykorzystywane przez projektantów dziedzinowych systemów hipertekstowych. Prezentowany tam punkt widzenia polega na interpretowaniu kryteriów oceny źródeł internetowych na czterech płaszczyznach [Hofman, Worsfold 1997] :

- kryterium celu systemu (ang. scope criteria) – dotyczące użytkowników;
- kryterium zawartości (ang. content criteria) – ocena zawartości zbioru informacyjnego;
- kryterium formy (ang. form criteria) – ocena medium/nośnika;
- kryterium procesualne (ang. process criteria) – ocena systemu;
- kryterium zarządzania kolekcją (ang. collection management criteria) – dotyczy samego systemu.

Najważniejsze z punktu widzenia gromadzenia i selekcji są kryteria dotyczące zawartości, formy oraz procesu. Analizując obiekty informacyjne pod kątem ich włączenia do zbioru informacyjnego DSH, zdaniem twórców tego raportu, należy uwzględnić:

- **Zakres** (ang. coverage). Czy treść źródła dotyczy zagadnień znajdujących się w polu tematycznym systemu?
- **Poprawność** (ang. validity). Czy prezentowane treści opisują w sposób prawdziwy rzeczywistość? Czy dane, na podstawie których prowadzono badania odznaczają się wiarygodnością i czy wyniki badań są weryfikowalne?
- **Reputacja i znaczenie źródła** (ang. authority and reputation of source). Kim jest autor oraz osoba lub instytucja dostarczająca informacje? Czy prezentowane informacje są recenzowane, jeżeli tak to przez kogo?

- **Substancjonalność\*** (ang. substantiveness). Czy źródło internetowe udostępnia pełne teksty dokumentów? Czy zawiera treści nacechowane komercyjnie? Jaka jest wartość informacyjna źródła?
- **Dokładność** (ang. accuracy). Na jakim poziomie szczegółowości prezentowane są informacje? Czy występują błędy językowe i typograficzne?
- **Wyczerpywalność/wszechstronność** (ang. comprehensiveness). Czy źródło internetowe jest kompletne, czy jest w trakcie tworzenia? Czy porusza wszystkie aspekty zagadnienia?
- **Unikalność** (ang. uniqueness). Unikalność źródła internetowego zależy od ilości informacji w nim zawartych, które są niedostępne w innych źródłach WWW.
- **Kompozycja i organizacja** (ang. composition and organisation). Czy treść jest poprawnie zorganizowana i skomponowana typograficznie? [Hofman, Worsfold 1997]

Formalna ocena internetowego obiektu informacyjnego uwzględnia pięć planów:

- prostotę funkcjonowania systemów nawigacji,
- udostępnianie użytkownikom systemów pomocy,
- stosowanie rozwiązań opartych na funkcjonujących standardach,
- odpowiednie (efektywne) wykorzystanie technologii informacyjnych,
- estetyczne wykonanie interfejsu użytkownika.

Taki sposób interpretacji obiektu informacyjnego utożsamia go z serwisem lub systemem hipertekstowym ponieważ narzuca perspektywę podobną do prezentowanej już architektury informacji. Trudno oceniać funkcjonowanie systemów nawigacji dla np. elektronicznych wersji artykułów z czasopism oraz estetykę interfejsu użytkownika dla e-booków. Kryteria te, które w założeniu miały być zapewne obiektywne, znacznie zawężają perspektywę interpretacji formalnych cech obiektów informacyjnych.

Kryteria procesualne odnoszą się do wymogów, jakie spełniać musi system wobec specyficznego rodzaju dokumentów sieciowych oraz ich treści. Dotyczą one:

- Integralności informacji - stabilność treści dokumentów w czasie. Zależy ona od autora i zarządzającego źródłem internetowym. Należy określić czy informacje muszą być aktualne, jak często należy je aktualizować oraz czy należy je archiwizować.

---

\* Substancjonalność źródła ocenia się biorąc pod uwagę ilość informacji aktualnie udostępnionych [Peereboom 1999]



- Integralności od strony dostępu – określenie, czy gwarantowany jest stały dostęp do zasobów DSH. Należy sprawdzić poprawność i aktualność odsyłaczy hipertekstowych wykorzystanych w zaindeksowanych dokumentach.
- Integralności systemowej – polegającej na permanentnej możliwości dostępu do dokumentów sieciowych i na określeniu poziomu stabilności (pod względem technologicznym) systemów sieciowych udostępniających źródła informacyjne.

### **3.3.3. Charakterystyka treściowa i formalna zasobów. Schematy metadanych.**

Odwzorowanie cech treściowych i formalnych obiektów informacyjnych stanowiących zbiór informacyjny dziedzinowego systemu hipertekstowego jest procesem, tak jak w przypadku katalogowania dokumentów tradycyjnych, czasochłonnym i skomplikowanym. W dziedzinowych systemach hipertekstowych katalogowanie zasobów WWW, spełniających określone wymogi treściowe i formalne, jest poprzedzone opracowaniem odpowiednich narzędzi służących do tego celu. Należą do nich:

- format opisu danych oraz zasady odwzorowania cech treści i formy obiektów informacyjnych,
- źródła leksykalne wykorzystywane w procesie katalogowania przedmiotowego i formalnego (słowniki JIW, kartoteki wzorcowe, listy terminów, itd.).

Zasady sporządzania opisów katalogowych cyfrowych obiektów informacyjnych wykorzystywane przez katalogerów w DSH są zamieszczane w dokumentacji poszczególnych systemów. Rzadko zdarza się sytuacja, w której reguły katalogowania tworzone są specjalnie na potrzeby systemu. W większości przypadków zasady te są oparte na funkcjonujących normach dotyczących sporządzania opisów bibliograficznych dla dokumentów elektronicznych. Zazwyczaj proces katalogowania w DSH jest oparty na zasadach Anglo-American Cataloging Rules 2 (AACR2) [Gorman, Winkler 1978], zasadach tworzenia opisów według Dublin Core Metadata Element Set, a także na zaleceniach normy ISBD(ER):International Standard Bibliographic Description for Electronic Resources. [ISBD(ER) 1997] Poradnik Nancy Olson [Olson 1995] zawierający interpretacje norm i zasad katalogowania dokumentów elektronicznych jest często cytowanym źródłem w dokumentacjach DSH.

Format opisu danych w DSH jest opracowywany na podstawie funkcjonujących schematów metadanych. Najczęściej wykorzystywanym zestawem metadanych jest Dublin Core Metadata Element Set. Uzupełnia się go również o zestaw dodatkowych

atrybutów, odwzorowujących cechy treści i formy obiektów informacyjnych, które mogą wpływać na informatywność ChWD oraz pełnić funkcje cech wyszukiwawczych.

Równie ważną kwestią w przypadku formatu opisu jest określenie sposobu kodowania danych odwzorowywanych podczas indeksowania. W tym celu wykorzystuje się opracowane przez projektantów systemów instrukcje lub wskazuje dokument o charakterze normatywnym rozstrzygającym te kwestie. Dotyczy to przede wszystkim formatu wprowadzanych do systemu dat, nazw własnych, geograficznych, kodów języków oraz innych cech, do których odwzorowania wskazane jest kontrolowanie ich potencjalnych wartości. Do odwzorowywania treści elementów zbioru informacyjnego, co zostanie szerzej omówione w następnym rozdziale, wykorzystuje się słowniki wybranych języków informacyjno-wyszukiwawczych zarówno o notacji paranaturalnej jak i sztucznej, o różnym stopniu kontroli.

Format opisu ma kluczowe znaczenie podczas projektowania i funkcjonowania narzędzi dostępu do kolekcji. Zestaw metadanych, czyli atrybutów elementów zbioru informacyjnego warunkuje możliwości wyszukiwawcze w danym systemie. To właśnie analiza zbioru informacyjnego pod kątem wyodrębniania relewantnych cech dokumentów powinna mieć wpływ na strukturę formatu opisu. Szczegółową charakterystykę formatów opisu danych stosowanych w DSH zawiera raport z projektu badawczego DESIRE. [Dempsey, Heery 1997] W dalszej części rozdziału zostaną scharakteryzowane przykładowe formaty opisu danych w wybranych DSH.

### 3.3.3.1. ROADS/IAFA

Internet Anonymous FTP Archive (IAFA) jest formatem metadanych stworzonych dla archiwów anonimowych FTP. [Nahotko 2004a] Celem twórców tzw. szablonów IAFA (ang. IAFA templates) było opracowanie struktury rekordu stosowanego przez administratorów zasobów dostępnych przy pomocy protokołu przesyłania plików – FTP. [Dempsey 1996] Struktura schematu metadanych dla źródła internetowego wyznacza 34 elementy, którym podczas katalogowania nadawane są określone wartości (Tab. 3.3).

Tab. 3.3. Struktura rekordu IAFA/ROADS dla źródeł internetowych. [Źródło: <http://www.ukoln.ac.uk/metadata/roads/templates/resource.html>]

Element	Charakterystyka
Template-Type	Typ szablonu
Template-Version	Wersja
Handle	Wartość systemowa
Title	Tytuł
Alternative-Title	Tytuł alternatywny

Author-Name-v1	Nazwa autora
Author-Email-v1	Adres e-mail autora
Keywords	Niekontrolowane słowa kluczowe charakteryzujące treść źródła
Subject-(DESCRIPTOR cluster) Subject-Descriptor-v* Subject-Descriptor-Scheme-v*	Charakterystyka treści za pomocą kontrolowanych źródeł leksyki
Description	Charakterystyka słowna
Publisher-Name-v1	Nazwa wydawcy
Publisher-Email-v1	Adres e-mail wydawcy
Contributor-Name-v1	Nazwa osoby lub instytucji która miała intelektualny wkład w powstanie źródła
Contributor-Email-v1	Adres a-mail do współtwórcy
Date-Created	Data utworzenia źródła
Category	Nazwa/nazwy kategorii w narzędziu dostępu przedmiotowego, do której/których przyporządkowano charakterystykę obiektu informacyjnego
Format-v1	Określenie formatu źródła
URI-v1	Ujednoczony identyfikator źródła (URI)
Source	Pochodzenie źródła (jeżeli zostało sporządzone na podstawie innego dokumentu)
Language-v1	Język źródła
Relation	Relacje z innymi źródłami lub obiektami
Coverage	Zasięg źródła (geograficzny oraz chronologiczny)
Copyright	Określenie praw autorskich
Admin-Name-v1	Nazwa osoby lub instytucji mającej prawa do administrowania źródłem
Admin-Email-v1	Adres e-mail administratora
To-Be-Reviewed-Date	Data następnej aktualizacji (przydzielana automatycznie)
Comments	Uwagi
Destination	Przeznaczenie źródła
Record-Created-Date	Data utworzenia rekordu
Record-Created-Email	Adres e-mail autora rekordu
Record-Last-Modified-Date	Data ostatniej modyfikacji rekordu
Record-Last-Modified-Email	Adres e-mail osoby ostatnio modyfikującej rekord
Record-Last-Verified-Date	Data ostatniej weryfikacji rekordu
Record-Last-Verified-Email	Adres e-mail osoby ostatnio weryfikującej rekord

Specyfikacja metadanych IAFA/ROADS przedstawiona w Tab. 3.3 nie określa zasad kodowania informacji w poszczególnych polach oraz nie identyfikuje potencjalnych źródeł słownictwa, np. do odwzorowywania cech treści obiektów informacyjnych. Format opisu danych IAFA/ROADS został zaproponowany w ramach projektu badawczego „Resource Organisation and Discovery in Subject-based Services (ROADS)” realizowanego w ramach programu „eLib Programme”. Jest on uznawany za pierwsze próby opracowania schematu metadanych dla DSH. Systemy powstające pod koniec lat 90-tych bieglego stulecia wykorzystywały ten format opisu nie ze względu na jego duże rozpowszechnienie, lecz z powodów czysto praktycznych. W ramach projektu ROADS opracowano oprogramowanie do zarządzania DSH, w tym również do katalogowania ich zasobów, w którym wspomniany format opisu danych był domyślnie zaimplementowany.

### 3.3.3.2. Intute metadata element set

W systemie Intute ([www.intute.ac.uk](http://www.intute.ac.uk)), jednym z aktualnie największych dziedzinowych systemów hipertekstowych funkcjonujących w Internecie, wykorzystywany jest zestaw metadanych pochodzący z kilku standardów. Skonstruowany i wdrożony zestaw elementów (Tab. 3.4) oparto na wybranych składnikach schematów Dublin Core, RDN/LTSN Learning Object Metadata\*, Research Data Archive (RDA) oraz zasadach katalogowania Anglo-American Cataloguing Rules – wydanie drugie (AACR2). [Intute Cataloging Guidelines 2006]

Tab. 3.4. Zestaw metadanych w systemie Intute.  
[Źródło: [www.intute.ac.uk/intute\\_cataloging\\_guidelines\\_v4\\_1.doc](http://www.intute.ac.uk/intute_cataloging_guidelines_v4_1.doc).]

Element	Charakterystyka
Handle	Nr rekordu nadawany automatycznie przez system
Subject group	Wartość nadawana automatycznie
Subgateway	Grupa przedmiotowa – przydzielanie dokumentu do jednego z podsystemów
Title	Tytuł dokumentu
Alternative titles	Tytuł alternatywny
URL	URL zunifikowany lokalizator zasobu
Language	Określenie języka dokumentu
Description	Charakterystyka słowna dokumentu, abstrakt
Keywords (controlled)	Słowa kluczowe kontrolowane słownikiem JIW
Keywords (uncontrolled)	Słowa kluczowe niekontrolowane
Resource types	Określenie typu zasobów. Format danych: kontrolowana lista wartości
Classifications	Nazwa kategorii tematycznej w strukturze narzędzia dostępu przedmiotowego, do której przyporządkowano dokument
Country of origin	Państwo pochodzenia
Resource creator	Twórca dokumentu: autor, instytucja sprawcza
Resource publisher	Wydawca dokumentu: instytucja odpowiedzialna za opublikowanie dokumentu w Internecie
Date resource created	Data utworzenia dokumentu. Format danych zgodny z ISO 8601:1988 – RRRR-MM-DD
Rights / copyright statement	Adres URL do oświadczenia dotyczącego praw autorskich dokumentu
Administrator name	Nazwa osoby/instytucji administrującej stroną www, gdzie opublikowano dokument
Administrator email	Adres e-mail do osoby/instytucji administrującej stroną www, gdzie opublikowano dokument
Format	Format pliku, w jakim został zapisany dokument. Format danych: kontrolowana lista wartości
Technical requirements	Informacje o wymaganiach technicznych w dostępie do dokumentu (np. konieczność rejestracji, strona zbudowana z tzw. ramek (ang. frames), wymagany program, plugin do odtwarzania dźwięku, itd.)
ISBN	ISBN dla dokumentu (o ile przydzielono)
ISSN	ISSN dla dokumentu (o ile przydzielono)
Period (Coverage)	Zasięg chronologiczny dokumentu
Geographic name	Zasięg geograficzny dokumentu. Format danych: przejmowanie wyrażeń tezaurusu Getty Thesaurus of Geographic Names
Latitude	Szerokość geograficzna – jeżeli dokument dotyczy konkretnego miejsca, obiektu w przestrzeni geograficznej
Longitude	Długość geograficzna – jeżeli dokument dotyczy konkretnego miejsca,

\* RDN/LTSN Learning Object Metadata jest formatem opisu danych dla edukacyjnych obiektów informacyjnych. [Hodgins 2002]

	obiekty w przestrzeni geograficznej
Audience	Określenie grupy użytkowników docelowych. Format danych: kontrolowana lista wartości
Educational level	Określenie poziomu dydaktycznego dokumentu. Format danych: kontrolowana lista wartości
Relationships	Określenie relacji, w jakich pozostaje dokument z innymi obiektami. Format danych: kontrolowana lista wartości
Flag	Wyróżnienie dokumentu. Wartość nadawana arbitralnie przez osoby katalogujące
Record source	Określenie zewnętrznego źródła przejęcia rekordu
Comments	Uwagi
Status	Określenie statusu rekordu, np. kompletny, niekompletny, do usunięcia. Format danych: kontrolowana lista wartości
Date record created	Data utworzenia rekordu
Record creator ID	Identyfikator osoby tworzącej rekord
Date for review	Określenie daty przyszłej aktualizacji rekordu. Format danych: zgodny z ISO 8601:1988, RRRR-MM-DD

W skład Intute wchodzi cztery podsystemy o rozłącznych zakresach przedmiotowych, co sprawia, że zakres całego systemu jest rozległy. Przyjęty zestaw metadanych jest wspólny dla każdego z podsystemów. Niemożliwe jest zatem wyodrębnienie elementów metadanych charakteryzujących bardzo szczegółowe cechy dokumentów, specyficzne dla wybranych kolekcji. Kosztem szczegółowości indeksowania na płaszczyźnie wyodrębniania cech relewantnych dokumentu, osiągnięto spójność formatu opisu danych w systemie. Spójność lokalna – możliwość wymiany metadanych w ramach całego systemu, została osiągnięta kosztem pełnej interoperacyjności formatu opisu. Wymiana metadanych możliwa jest tutaj wyłącznie w zakresie, w jakim udział w schemacie Intute ma Dublin Core.

### 3.3.3.3. Schemat SSG-FI

Sieć DSH Special Subject Guides – Fachinformation (SSG-FI <http://www.sub.uni-goettingen.de/ssgfi/>) skupia cztery systemy z zakresu: matematyki (Mathguide), geologii (GeoGuide), kultury angielskiej i amerykańskiej (Anglo-American Culture) oraz leśnictwa (ForestryGuide). Odzworowanie cech formalnych i treściowych obiektów informacyjnych stanowiących elementy zbioru informacyjnego wymienionych systemów oparte jest na wspólnym formacie opisu danych. [Fisher, Neuroth 2000] Schemat metadanych SSG-FI zbudowano z 15 podstawowych elementów Dublin Core wraz z wyodrębnionymi kwalifikatorami oraz 13 dodatkowych elementów uszczegóławiających charakterystyki indeksowanych zasobów (Tab. 3.5).

Tab. 3.5. Schemat metadanych dla systemów sieci SSG-FI [Źródło: SSG-FI Project Documentary 1997]

Nazwa schematu metadanych	Nazwa	Nazwa oryginalna	Wartość
Dane bibliograficzne			

Wewnętrzny	Numer wewnętrzny pliku	Internal File Number (Internal)	Numer wewnętrzny rekordu automatycznie generowany przez program zarządzający bazą danych
DC.Title	Tytuł	Title	Tytuł dokumentu określony przez autora, twórcę lub wydawcę
DC.Creator	Autor	Author	Osoba lub osoby odpowiedzialne za intelektualną zawartość dokumentu (twórcy stron www nie są traktowani jako autorzy) Format danych: nazwisko i imię, adres, państwo, adres e-mail.
DC.Creator	Edytor	Editor	Osoba lub osoby odpowiedzialne za intelektualną zawartość dokumentu (twórcy stron www nie są traktowani jako autorzy) Format danych: nazwisko i imię, adres, państwo, adres e-mail.
DC.Publisher	Wydawca	Publisher	Wydawca, towarzystwo naukowe, itd. odpowiedzialne za opublikowanie treści dokumentu. Format danych: nazwa, miasto, państwo
DC.Publisher	Dystrybutor	Distributor	Instytucja odpowiedzialna za obsługę (techniczną) serwera, gdzie umieszczony został dokument. Format danych: nazwa, miasto, państwo
DC.Type	Typ źródła	Type	online (WWW), offline (CD-ROM) inne drukowane (np. książka.).
DC.Language	Język	Language	Język zawartości dokumentu. Format danych: zgodny z ISO 639 kody języków, pole powtarzalne
SSG-FI	Państwo	Country	Państwo, w którym zlokalizowano serwer przechowujący dokument. Format danych: zgodny z ISO 3166:1993, German DIN EN 23166:1995
DC.Format	Format	Format	Zasoby online: format danych (np. text/html - plik tekstowy, jpeg – plik obrazu, gif - plik obrazu. Cel - poinformowanie użytkownika o wymaganiach programowych i sprzętowych koniecznych do odczytania dokumentu Zasoby offline: wymagania programowe i sprzętowe Zasoby drukowane: rozmiar dokumentu
DC.Subject	Słowa kluczowe	Keywords	Słowa kluczowe odwzorowujące zawartość dokumentu
DC.Description	Opis	Description	Charakterystyka treści dokumentu wyrażona w języku naturalnym – abstrakt
DC.Identifier	URL	URL	Zasoby online: Uniform Resource Locator: Zasoby offline i drukowane: URL do informacji o dokumencie dostępnej online
DC.Identifier	ISSN/ISBN	ISSN/ISBN	ISSN/ISBN numery identyfikujące zasoby offline i drukowane
DC.Identifier	Sygnatura	Call Number	Tylko zasoby offline i drukowane: sygnatura Niedersächsische Staats- und Universitätsbibliothek Göttingen
DC.Relation	Kopie na serwerach	Mirrors	Tylko zasoby online: URL najbliższej geograficznie kopii strony na serwerze
DC.Relation	Pochodzenie	Derived from	Relacja pomiędzy zasobem online/offline a wersją drukowaną np. ISBN wersji drukowanej dokumentu
DC.Relation	Dokument nadrzędny	Contained in (DC 13)	Zasoby online: adres źródła, którego częścią jest dokument Zasoby offline: zazwyczaj nie stosuje się Zasoby drukowane: np. tytuł serii
DC.Relation	Archiwizacja	Archived by	Instytucja archiwizująca dokument

DC.Type SSG-FI	Typ źródła	Source Type	Określenie typu zasobu oparte na kontrolowanej liście dla tej cechy dokumentów (np. baza danych, serwis informacyjny, oprogramowanie)
Klasyfikacja			
DC.Subject	Symbol MSC/GOK w jęz. naturalnym	MSC/GOKverbal (DC 3)	Określenie przedmiotu dokumentu w postaci odpowiedników słownych symboli Mathematical Subject Classification (MSC) lub Göttinger Online Klassifikation (GOK). Pole powtarzalne
DC.Subject	Symbol MSC/GOK	MSC/GOK	Symbole klasyfikacji MSC lub GOK
DC.Subject	KDD	DDC	Symbole Klasyfikacji Dziesiątej Deweya
DC.Subject	BK	BK	Symbole klasyfikacji Basic Classification of the Joint Library Catalog of Northern Germany Libraries
Data			
DC.Date	Rok publikacji	Year of publication	Rok publikacji dla zasobów offline i drukowanych. Format danych: RRRR-MM-DD
DC.Date	Aktualizacja	Update`	Data ostatniej aktualizacji zasobów online. Format danych: RRRR-MM-DD
SSG-FI	Częstotliwość	Course	Częstotliwość ukazywania się, publikacji (np. kwartalnie, rocznie, itd.)
DC.Coverage	Zakres czasowy	Temporal Coverage	Zasięg czasowy, w którym dokument jest dostępny. Format danych: RRRR-MM-DD
Dostępność			
DC.Rights	Dostęp	Access	Warunki dostępu do zasobów: online: czy dostęp jest darmowy czy też za opłatą; offline i drukowane: czy dostępne w Bibliotece SUB, jaki katalog odnotowuje dany dokument: Verbundkatalog, Deutschen Verbundkatalog, Karlsruher Virtueller Katalog (KVK).
DC.Rights	Ograniczenia	Restriction	Zasoby online: inne ograniczenia w dostępie poza finansowymi, np. autoryzacja, hasło, itd. zasoby offline i drukowane:: zależnie od warunków dostępu stosowanych przez biblioteki
DC.Rights	Uwagi	Remarks	Dodatkowe informacje na temat dostępu do zasobów.
Ocena			
SSG-FI	Zawartość	Contents	Ocena zależna od stopnia relewancji pod względem zakresu przedmiotowego dokumentu oraz sposobu opracowania tematu Ocena: od "*" do "**"
SSG-FI	Klarowność	Clarity	Precyzja w wyrażaniu myśli oraz estetyka pod względem formalnym. Ocena: od "*" do "**"
SSG-FI	Aparat pomocniczy	Index	Ocena aparatu pomocniczego: indeksy, spisy treści, wyszukiwarki serwisów. Ocena: od "*" do "**"
SSG-FI	Odesłania`	Links	Tylko dla zasobów online. Ocena pod kątem liczby odesłań do zasobów Internetu o zbliżonej tematyce. Ocena: od "*" do "**"
SSG-FI	Poziom	Level	Określenie poziomu trudności i przeznaczenia źródła (np. dla szkół podstawowych, dla specjalistów, itd.).
Statystyka			
SSG-FI	Rozmiar	Size	Zasoby online: liczba dokumentów tworzących źródło (jeżeli brak danych – pomijane) Zasoby offline: liczba rekordów w bazie danych Zasoby drukowane: liczba stron
SSG-FI	Odesłania	Backlinks	Tylko dla zasobów online: liczba odesłań

	zwrotne		skierowujących do źródła
SSG-FI	Notatki	Notes	Dodatkowe informacje
SSG-FI	© SSG-FI	© SSG-FI	Data sporządzenia rekordu przez system SSG-FI, data ostatniej aktualizacji rekordu. Format danych: RRRR-MM-DD
Wewnętrzny	Komentarz	Comment	Dodatkowe informacje
Wewnętrzny	Status	Status	Rekord kompletny/niekompletny

W strukturze formatu opisu danych SSG-FI (Tab.3.5) wyodrębniono 6 grup metadanych odwzorowujących poszczególne kategorie cech. Są to:

- dane bibliograficzne,
- klasyfikacja,
- data,
- dostępność,
- ocena,
- statystyka.

Większość cech odwzorowywana jest z wykorzystaniem elementów schematu Dublin Core. Specjalne elementy formatu opisu zostały stworzone dla treściowej i formalnej oceny źródła oraz jego charakterystyki pod względem jego pojemności informacyjnej (liczba dokumentów i/lub rekordów tworzących źródło) oraz wiarygodności i popularności (liczba odesłań skierowujących do źródła). Schemat SSG-FI identyfikuje dla poszczególnych elementów źródła ich wartości oraz zasady kodowania danych.

### 3.3.3.4. Edna metadata standard

Kolejnym DSH wykorzystującym „kombinowaną” strukturę formatu opisu danych jest australijski system z zakresu edukacji Edna Online – Education Network Australia ([www.edna.edu.au](http://www.edna.edu.au)). Gromadzone i udostępniane są tutaj informacje o zasobach WWW, które stanowią potencjalną wartość edukacyjną i mogą być wykorzystane w procesie kształcenia. Specyfiką tego systemu jest charakterystyczne traktowanie internetowych obiektów informacyjnych. Ze względu na cel systemu, traktowany jest on jako tzw. repozytorium obiektów edukacyjnych (ang. learning object repository). [Najjar, Duval 2006] Elementy zbioru wyszukiwawczego, czyli charakterystyki wyszukiwawcze źródeł internetowych interpretowane są jako tzw. metadane obiektów edukacyjnych (ang. learning object metadata). Podstawowa identyfikacja obiektu informacyjnego została oparta na zestawie elementów Dublin Core wraz z odpowiednimi kwalifikatorami. Pozostałe cechy, określające przede wszystkim wartość edukacyjną określono za pomocą dodatkowych



elementów formatu zaprojektowanych lub przejętych z innych schematów (Tab. 3.6)

Tab. 3.6. Zestaw metadanych dla systemu Edna.  
[Źródło: <http://www.edna.edu.au/edna/go/resources/metadata/pid/261>]

Nazwa elementu	Dookreślenie elementu	Definicja	Schemat kodowania, źródło słownictwa
Dublin Core DC.Identifier		Jednoznaczne odesłanie do źródła	URI
DC.Title		Tytuł źródła	
	Alternatywny	Alternatywna forma tytułu pod jaką występuje źródło	
DC.Description		Charakterystyka treści dokumentu wyrażona w języku naturalnym	
	Spis treści	Lista jednostek tematycznych składających się na treść źródła	
	Abstrakt	Abstrakt	
DC.Subject		Przedmiot dokumentu	ASCED (Australian Standard Classification of Education ) APSDEP (Asian and Pacific Skill Development Programme Thesaurus) DDC (Dewey Decimal Classification) edna-klā KBK (Library of Cogress Classification) LCSH (Library of Congress Subject Headings) MeSH (Medical Subject Headings) SCIS (Schools Catalogue Information Service Subject Headings) UDC (Universal Decimal Classification)
DC.Publisher		Jednostka odpowiedzialna za opublikowanie treści dokumentu	
DC.Creator		Jednostka odpowiedzialna za stworzenie zawartości źródła	
DC.Date		Daty związane z wydarzeniem w okresie istnienia źródła	DCMI Period W3C-DTF
	Created	Data utworzenia źródła	
	Valid	Data walidacji źródła	
	Available	Data udostępnienia źródła, lub okres, w którym źródło było dostępne	
	Issued	Data oficjalnego opublikowania źródła	
	Modified	Data ostatniej modyfikacji źródła	
DC.Type		Natura lub kategoria treści źródła. Określenie rodzaju zasobu pod	Źródła słownictwa: DCMI Type Vocabulary edna-dokument

		względem formalnym i rodzaju zawartości	edna-program edna-zdarzenie
DC.Format		Fizyczna lub cyfrowa materializacja źródła	
DC.Language		Język treści intelektualnych źródła	ISO 639-2 RFC 1766
DC.Coverage	Spatial	Zasięg treści źródła	edna-spatial DCMI Point ISO 3166 DCMI Box TGN
DC.Rights		Informacja o prawach własności do źródła	
DC.Relation		Odesłanie do źródła pozostającego w relacji z opisywanym	URI
DC.Contributor		Jednostka odpowiedzialna za współudział w tworzeniu treści źródła.	
DC.Source		Odesłanie do pierwotnego źródła, z którego pochodzi opisywane źródło	URI
Edna			
EDNA.Audience		Kategorie użytkowników, do których skierowana jest treść źródła	Źródło edna-uzytkownicy edna-sektor edna-poziom
EDNA.Approver		Adres e-mail osoby lub organizacji zatwierdzającej włączenie źródła do zasobów EdNA Online	
EDNA.Category Code		Numeryczny kod kategorii w hierarchicznych narzędziach dostępu EdNA Online do jakiej przyporządkowano źródło	
EDNA.Entered		Dane wprowadzane w celu ułatwienia zarządzania rekordami	
EDNA.Indexing		Określenie ziarnistości informacji podczas dalszego indeksowania, której punkt startowy stanowi źródło.	
EDNA.Review		Recenzja źródła lub URL do jego recenzji w Internecie	
EDNA.Reviewer		Nazwisko osoby lub nazwa instytucji autoryzującej recenzję.	
EDNA.Version		Wersja aktualnie stosowanego schematu metadanych	

W przedstawionym formacie opisu, oprócz standardowych elementów schematu Dublin Core, wyodrębniono osiem dodatkowych cech, których wartości uzupełniają

charakterystykę obiektu informacyjnego. Dotyczą one wiarygodności źródła, grupy jego potencjalnych użytkowników oraz lokalizacji ChWD w kategorii/kategoriach tematycznych w narzędziach dostępu przedmiotowego.

### 3.3.3.5. Schemat metadanych Renardus

Europejska sieć akademicka Renardus była projektem badawczym finansowanym w latach 2000-2002 przez Komisję Europejską jako część programu Technologie Społeczeństwa Informacyjnego. [Day, Koch, Neuroth 2004] Zrzeszała 11 europejskich dziedzinowych systemów hipertekstowych o różnych, rozłącznych zakresach przedmiotowych. Celem projektu było opracowanie i wdrożenie wspólnego narzędzia dostępu do zasobów uczestniczących DSH. Efektywne funkcjonowanie sieci musiało więc zostać oparte na osiągnięciu konsensusu na poszczególnych płaszczyznach współpracy, w tym wymiany metadanych. [Heery i in. 2001; Huxley i in. 2003] Format opisu obiektów cyfrowych dla Renardus oparto na strukturze Dublin Core wraz z jego rozszerzeniem w postaci dodatkowych elementów. Każdy z DSH udostępniających swoje zbiory w sieci był zobligowany do zastosowania tzw. formatu przejścia na obowiązujący w sieci. W Tab. 3.7 przedstawiono budowę formatu metadanych dla sieci Renardus.

Tab. 3.7. Schemat metadanych dla sieci Renardus [Źródło: Becker et. al. 2000]

Element	Charakterystyka
<i>Elementy Dublin Core</i>	
Title	Tytuł
Title.Alternative	Tytuł alternatywny
Creator	Twórca
Description	Opis
Subject	Opis rzeczowy
Subject - DC encoding schemes	Opis rzeczowy (kodowanie Dublin Core - LCSH, MeSH, KDD, KBK lub UKD)
Identifier – URI	Identyfikator
Language - ISO639-2	Język (kodowanie – norma ISO 639-2)
Type	Typ
Type – DCMIType	Typ (kodowanie Dublin Core - LCSH, MeSH, DDC, KBK lub UKD)
<i>Elementy Renardus</i>	
Creator – Renardus qualifier	Twórca – kwalifikator Renardus
Subject – Renardus encoding schemes	Opis rzeczowy – kodowanie schematem Renardus
Subject – Renardus DDC	Opis rzeczowy – kodowanie z wykorzystaniem znormalizowanych symboli KDD
Country – ISO 3166-1	Nazwa państwa (kodowanie – norma ISO 3166-1)
<i>Elementy administracyjne</i>	
Full Record URL	URL odesłania hipertekstowego do pełnych danych o opisywanym obiekcie cyfrowym
SBIG ID	Akronim systemu, z którego pochodzi opis obiektu

Dane przedstawione w tabeli 3.7 wskazują na zastosowanie elementów schematu Dublin Core do pełnej identyfikacji obiektu informacyjnego. Dodatkowe elementy schematu Renardus częściowo powielają uzyskiwane dane. Uzupełniają ChWD o metadane administracyjne oraz ujednoliczony opis rzeczowy wyrażony za pomocą symboli KDD. Format ten został opracowany w ramach projektu Research Support Library Programme. [Powell i in. 2000] W 2007 roku zawieszono pracę systemu.

### **3.3.4. Dostęp przedmiotowy\***

Dostęp do zasobów DSH realizowany jest poprzez automatyczne mechanizmy wyszukiwawcze oraz narzędzia dostępu przedmiotowego. Pierwszy tryb dostępu do zbioru informacyjnego DSH polega na udostępnieniu użytkownikowi wyszukiwarki przeszukującej zaindeksowane zasoby systemu. Proces wyszukiwania informacji polega w tym przypadku na zbudowaniu przez użytkownika zapytania informacyjnego w postaci ciągu słów kluczowych i wysłaniu go do mechanizmu przeszukującego. Odpowiedzią systemu jest lista ChWD spełniających kryteria relewancji systemowej. Rezultaty wyszukiwania w DSH wyświetlane są użytkownikowi w trybie prostym, uwzględniającym jedynie podstawowe dane obiektów informacyjnych lub w trybie zaawansowanym, gdzie wyświetla się pełną ChWD.

Dostęp przedmiotowy w DSH polega na grupowaniu charakterystyk wyszukiwawczych obiektów informacyjnych w odpowiednich kategoriach tematycznych zagnieżdżonych w strukturze hipertekstu. Każda kategoria stanowi punkt dostępu do określonej zawartości kolekcji. Mogą one reprezentować wartości wymiaru treści lub formy elementów zbioru informacyjnego. W zależności od przyjętego sposobu organizacji punktów dostępu, mogą one tworzyć w DSH struktury hierarchiczne, ahierarchiczne lub hybrydowe (zob. Rozdz. 6.). Każda kategoria jest węzłem powiązania hipertekstowego, którego aktywacja przez użytkownika powoduje eksplorację takiej struktury. W przypadku hierarchicznych układów kategorii tematycznych powoduje to wyświetlenie użytkownikowi odpowiednich ChWD przyporządkowanych do aktualnie eksplorowanej oraz wyświetlenie kategorii

---

\* Szczegółowa analiza tego typu narzędzi w dziedzinowych systemach hipertekstowych zostanie przedstawiona w kolejnych rozdziałach. W tym miejscu, czyli na etapie określania budowy omawianego rodzaju systemów informacyjnych, zostaną scharakteryzowane ogólne założenia dotyczące wykorzystania zbiorów kategorii tematycznych jako narzędzia dostępu do kolekcji.

bezpośrednio podrzędnych zakresowo. Dla struktur ahierarchicznych (w porządku alfabetycznym) uruchomienie powiązania hipertekstowego dla określonego węzła powoduje jedynie wyświetlenie listy ChWD przyporządkowanych do danej kategorii tematycznej.

Narzędzia dostępu przedmiotowego w DSH swoją budową przypominają strukturę kategorii tematycznych w katalogach internetowych. Charakteryzują się jednak specyficznym wykorzystaniem organizacji pól semantycznych wybranych JIW do tworzenia struktury kategorii tematycznych. Zostanie to szczegółowo omówione w szóstym rozdziale rozprawy. W tym miejscu należy zaznaczyć, że budowa narzędzi dostępu przedmiotowego jest uzależniona w bezpośredni sposób od narzędzi indeksowania wykorzystywanych w danym systemie.

Rezultaty wyszukiwania informacji w DSH zarówno z wykorzystaniem narzędzi dostępu przedmiotowego i mechanizmów wyszukiwawczych można poddawać sortowaniu według przyjętych w danym systemie kryteriów. Są to zazwyczaj cechy formalne, np. tytuł, rodzaj obiektu informacyjnego, data utworzenia. Zastosowanie w niektórych (np. MathGuide) DSH stopniowalnych (niebinarnych) współczynników relewancji pozwala, porządkować rezultaty wyszukiwania według wartości tej zmiennej.

### ***3.4. Typologia dziedzinowych systemów hipertekstowych***

Typologia jest "uporządkowaniem elementów zbioru, według którego każdy z nich zakwalifikowany jest do pewnego typu elementów na podstawie relacji podobieństwa". [Sosińska-Kalata 2002a s. 19] Typologie ułatwiają zrozumienie złożoności zjawisk i zbiorów poddawanych temu procesowi. Twórcą dotychczas jedynej typologii dziedzinowych systemów hipertekstowych funkcjonującej w piśmiennictwie naukowym z zakresu informacji naukowej jest T. Koch. [Koch 2000] Cechą na podstawie, której sporządzono taki podział dla DSH jest zakres systemu. Tabela 3.8 przedstawia typologię T. Kocha w postaci macierzy o wymiarach 3 na 4.

Tab. 3.8. Typologia dziedzinowych systemów hipertekstowych w postaci macierzy. [Źródło: Koch 2000]

	Jeden	Kilka	Wiele/wszystkie
Przedmiot			
Państwo/region			
Język zasobów			

Typ zasobów			
-------------	--	--	--

Na podstawie takiego wielowymiarowego uporządkowania możemy mówić o systemach, których pole tematyczne ograniczone jest do jednego: przedmiotu, regionu geograficznego, języka oraz rodzaju indeksowanych zasobów, jak również o wielojęzycznych, globalnych systemach o charakterze uniwersalnych rejestrujących różne typy zasobów sieciowych.

Najczęściej spotykane typy to: [Koch 2000]

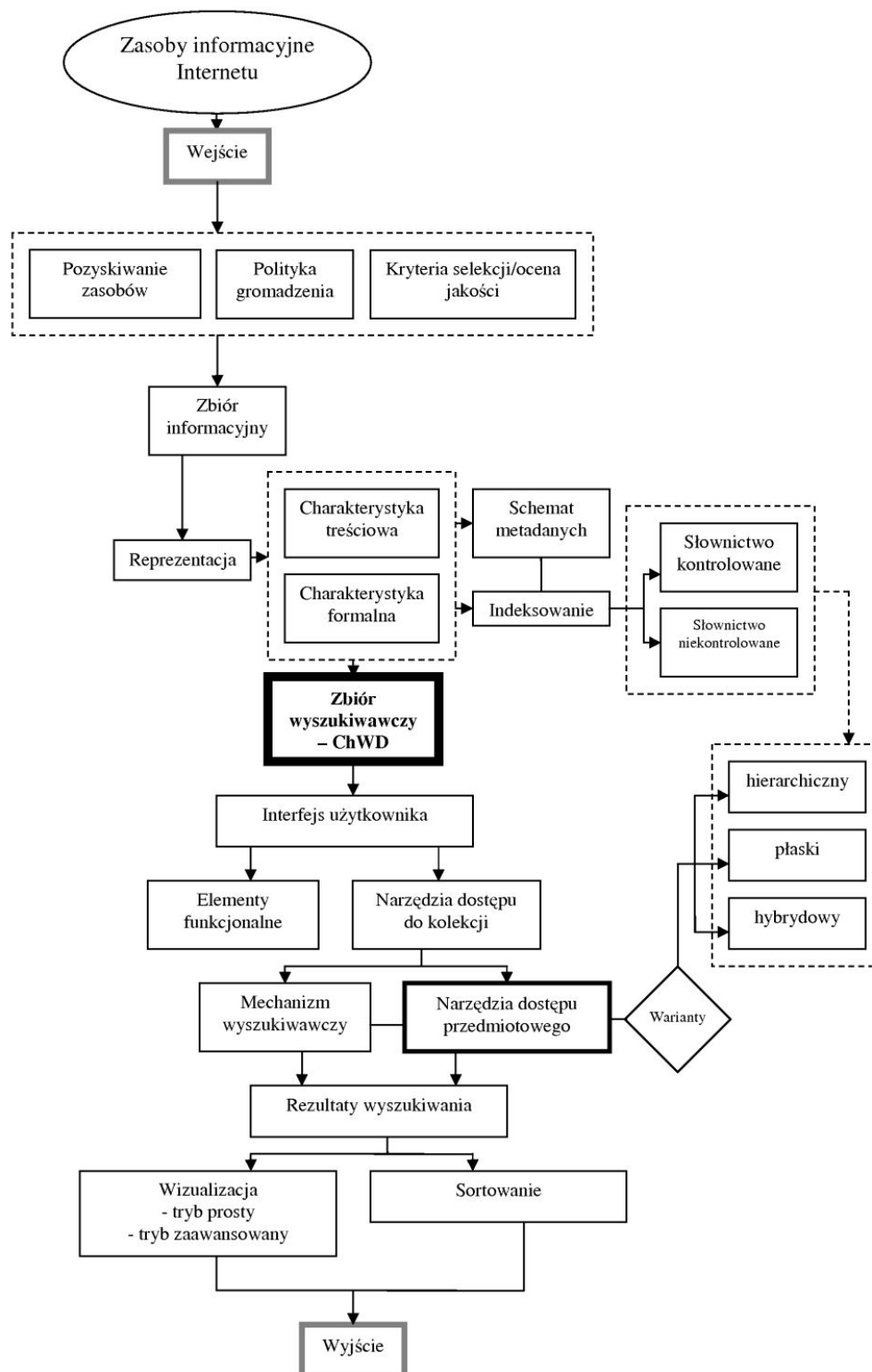
- system o zasięgu globalnym o polu tematycznym ograniczonym do jednej dziedziny wiedzy rejestrującym zasoby wielojęzyczne - np. MathGuide z zakresu matematyki;
- system o zasięgu globalnym o szerokim polu tematycznym, rejestrującym zasoby w jednym języku - np. BUBL indeksujący anglojęzyczne zasoby WWW;
- system o charakterze narodowym, wielotematyczny - np. Science LinkHouse indeksujący zasoby fińskiego internetu;
- system o charakterze narodowym o polu tematycznym ograniczonym do jednej dziedziny, rejestrujący zasoby w jednym języku - np. GEM z zakresu edukacji, AgriGate z zakresu rolnictwa.

Każdy z wymienionych w macierzy wymiarów stanowi podstawę do dalszego podziału, co umożliwia stworzenie bardzo szczegółowej typologii. Podstawę sporządzania typologii stanowi cel, jakiemu ma służyć. Zdaniem autora, w tym przypadku bezzasadne jest wyodrębnianie maksymalnej liczby możliwych typów dziedzinowych systemów hipertekstowych. Istotą przytoczonej typologii jest jedynie zidentyfikowanie potencjalnych kryteriów takiego podziału.

### ***3.5. Model analityczny dziedzinowego systemu hipertekstowego***

Przedstawiona w tym rozdziale charakterystyka dziedzinowego systemu hipertekstowego wyszczególniająca jego elementy strukturalne oraz procesy, które towarzyszą jego funkcjonowaniu pozwala na ekstrakcję z tych treści modelowej wersji tego typu systemów informacyjno-wyszukiwawczych. W tym celu na potrzeby pracy skonstruowano model analityczny dziedzinowego systemu hipertekstowego, co ilustruje Rys. 3.1. Model jako „struktura będąca odwzorowaniem jakiejś rzeczywistości [...] na

zasadzie przyporządkowania jej cech relewantnych, pozwalająca na zasadzie analogii na wyciągnięcie wniosków dotyczących możliwych stanów tej rzeczywistości” [Słownik encyklopedyczny 2002, s. 160] realizuje obok innych przede wszystkim funkcje poznawcze. Jest odwzorowaniem rzeczywistości, ale jest jej jednoczesnym uproszczeniem. Przez to jest łatwiejszy do zrozumienia niż rzeczywistość, którą opisuje. Do opisu modeli analitycznych stosuje się obok języków sformalizowanych również język naturalny, co wykorzystano w rozprawie. Konstrukcja modelu została oparta na analizie dziedzinowego systemu hipertekstowego uwzględniającej warunki wystarczające do zaliczenia systemu do opisywanej klasy systemów informacyjno-wyszukiwawczych. Sam model realizuje przede wszystkim funkcje poznawcze i ma za zadanie wizualizację elementów składowych omawianego typu systemów oraz procesów, jakie w nim zachodzą. W celu uzyskania jak najbardziej przejrzystego obrazu opisywanego fragmentu rzeczywistości nie zidentyfikowano rodzajów relacji oraz typów procesów, jakie zachodzą pomiędzy poszczególnymi elementami. Model ten w zamyśle autora ma stanowić podsumowanie przedstawionych dotychczas rozważań nad dziedzinowymi systemami hipertekstowymi i może zostać wykorzystany do projektowania nowych systemów informacyjnych należących do tej klasy.



Rys. 3.1. Model analityczny dziedzinowego systemu hipertekstowego [Źródło: opracowanie własne]

Na wejściu do systemu mamy do czynienia z rozproszonymi i heterogenicznymi zasobami informacyjnymi World Wide Web. Na podstawie przyjętych w systemie technik pozyskiwania oraz polityki gromadzenia materiałów do kolekcji, dokonuje się ich selekcji oraz oceny jakości zarówno ich zawartości, jak i formy. Obiekty



informacyjne spełniające kryteria włączenia do zasobów systemu tworzą jego zbiór informacyjny. Ten, poddawany jest procesowi reprezentacji informacji, który poprzez charakterystykę treściową i formalną prowadzi do opracowania dla każdego źródła charakterystyki wyszukiwawczej dokumentu internetowego. Proces reprezentacji przeprowadzany jest na podstawie przyjętego schematu metadanych oraz zasad indeksowania, w którym do reprezentacji treści i formy wykorzystuje się zbiory słownictwa kontrolowanego i/lub niekontrolowanego. Sporządzone w ten sposób ChWD tworzą zbiór wyszukiwawczy systemu. Dostęp do zasobów systemu umożliwia interfejs użytkownika, na który obok narzędzi dostępu do zbioru wyszukiwawczego składają się elementy funkcjonalne umożliwiające m.in. interakcję użytkownika z systemem. Bezpośredni dostęp do kolekcji realizowany jest poprzez wykorzystanie mechanizmu wyszukiwawczego, w postaci wyszukiwarki przeszukującej zasoby systemu oraz poprzez narzędzia dostępu przedmiotowego. Te przyjmują formę struktur hierarchicznych, płaskich bądź hybrydowych. Ich funkcjonowanie polega na wykorzystaniu kategorii tematycznych grupujących odesłania do odpowiednich ChWD. Każdy termin stanowiący nazwę elementu takiego narzędzia jest punktem dostępu do zbioru wyszukiwawczego. Posiada status węzła hipertekstowego, którego aktywacja powoduje, w zależności od realizowanego modelu takiego narzędzia, wyświetlenie użytkownikowi rezultatów wyszukiwania w postaci listy relewantnych ChWD oraz powiązanych kategorii tematycznych. Rezultaty wyszukiwania, stanowiące odpowiedź systemu na wysłanie zapytania poprzez mechanizm wyszukiwawczy lub eksplorację narzędzia dostępu przedmiotowego, mogą przyjąć formę wyświetlania w trybie prostym (zawierającym podstawowe informacje o obiekcie informacyjnym) lub zaawansowanym (jego pełna charakterystyka). Tym samym na wyjściu z systemu użytkownik otrzymuje metainformacje w postaci charakterystyk wyszukiwawczych wyselekcjonowanych obiektów informacyjnych opublikowanych w WWW.

### **3.6. Geneza dziedzinowych systemów hipertekstowych**

Rodowód pierwszych DSH bezpośrednio związany jest z projektem badawczym „Access to Network Resources” (ANR) – „Dostęp do zasobów sieciowych”, który jako jeden z wielu, realizowany był w ramach programu badawczego „eLib - The Electronic Libraries Programme”. Funkcjonował on w latach 1994-2001. Podczas jego realizacji powstało około 60 szczegółowych projektów, które dotyczyły [Kirremuir 1998] :

- dostępu do zasobów sieciowych,

- kształcenia użytkowników informacji,
- czasopism elektronicznych,
- digitalizacji,
- dokumentów graficznych,
- elektronicznego dostarczania dokumentów,
- druku na żądanie,
- preprintów i szarej literatury,
- projektów pomocniczych.

Celem twórców projektu ANR było opracowanie modelu dziedzinowego dostępu do zasobów Internetu, a także sposobów finansowania tego rodzaju systemów i inicjatyw. Powstanie projektu oraz realizacja celów, jakie zostały w nim wyznaczone zapoczątkowało rozwój badań oraz wdrażanie tego typu systemów informacyjnych w Internecie. Główną przesłanką do podjęcia prac nad tym zagadnieniem było dostrzeżenie przez twórców tzw. raportu Folleta [Follet 1993] wzrostu znaczenia informacji opublikowanych w Internecie oraz potencjalnych możliwości ich wykorzystania na potrzeby nowych form edukacji. We wspomnianym raporcie stwierdzono, że Internet oferuje dostęp do dużej ilości zasobów o szerokim zakresie, jednak stosunek materiałów relewantnych do dostępnych jest stosunkowo mały. W związku z tym, uznano iż należy wypracować metodyczne podejście do tego zagadnienia. [Whitelaw 2000]

Efektom prowadzonych prac w ramach ANR było dziesięć projektów szczegółowych [The Projects], których efektem było zarówno uruchomienie nowych DSH oraz opracowanie rozwiązań technologicznych i organizacyjnych dla nowopowstających.

Aktualny kierunek rozwoju dziedzinowych systemów hipertekstowych to dążenie do ich funkcjonowania jako elementu kompleksowego systemu brokerskiego lub zrzeszonego w sieć systemów dziedzinowych oferujących dostęp do zasobów Internetu z różnych dziedzin wiedzy. Najlepszym tego przykładem jest kompleksowa platforma informacyjna Intute ([www.intute.ac.uk](http://www.intute.ac.uk)), która powstała w połowie 2006 roku poprzez połączenie kilku samodzielnych DSH. Było to osiem DSH wchodzących w skład sieci „Resource Discovery Network” (RDN). Pracę systemu koordynuje konsorcjum, w skład którego wchodzi przedstawiciele siedmiu uniwersytetów z Wielkiej Brytanii. Kontrolę finansową nad systemem sprawują: Joint Information Systems Committee (JISC), Arts and Humanities Research Council (AHRC) oraz Economic and Social Research Council (ESRC).

Innym przykładem łączenia DSH w sieci jest niemiecka SSG-FI - Special Subject Gateways (<http://www.sub.uni-goettingen.de/ssgfi/>). Biblioteka Uniwersytecka w Göttingen (Niedersächsische Staats- und Universitätsbibliothek Göttingen) realizuje projekt pt. „Specjalne kolekcje przedmiotowe”, który jest finansowany m.in. przez Deutsche Forschungsgemeinschaft, centralną niemiecką fundację do spraw promocji badań naukowych. Trzon projektu tworzą:

- MathGuide ([www.mathguide.de](http://www.mathguide.de)) – z zakresu matematyki,
- GeoGuide ([www.geo-guide.de](http://www.geo-guide.de)) – z zakresu nauki o Ziemi,
- Anglo-American Culture ([www.sub.uni-goettingen.de/ssgfi/anglo-america.html](http://www.sub.uni-goettingen.de/ssgfi/anglo-america.html)) – z zakresu kultury anglo-amerykańskiej,
- ForestryGuide ([www.forestryguide.de/](http://www.forestryguide.de/)) – z zakresu leśnictwa.

Systemy w sieci SSG-FI ustępują liczbą zgromadzonych rekordów dużym systemom jak Intute, lecz opieka merytoryczna Biblioteki Uniwersyteckiej w Göttingen jest gwarantem wysokiej jakości usług informacyjnych oraz stosowania obowiązujących standardów w zakresie katalogowania zasobów internetowych.

Interesującym przykładem współpracy między DSH był wspomniany projekt „Renardus - Reynard: Academic Subject Gateway Service Europe”. Celem projektu było stworzenie sieci europejskich dziedzinowych systemów hipertekstowych, które stanowiłyby wartościowe źródło informacji o dokumentach elektronicznych opublikowanych w Internecie. Zakres systemów miał obejmować możliwie jak najwięcej dziedzin wiedzy. Projekt Renardus skierowany był do użytkowników wywodzących się ze szkolnictwa wyższego. Byli to zarówno studenci, wykładowcy jak i naukowcy. Sieć tworzyło dwanaście instytucji z siedmiu różnych państw. Były to: [Day 2000]:

Biblioteki narodowe:

- Koninklijke Bibliotheek (Holandia) - koordynator projektu oraz partner strategiczny holenderskiego systemu DutchESS ([www.kb.nl/dutchess/](http://www.kb.nl/dutchess/)),
- Bibliotheque Nationale de France (Francja) - patron francuskiego systemu Les Signets (<http://signets.bnf.fr/>),
- Die Deutsche Bibliothek (Niemcy).

Biblioteki uniwersyteckie i instytucje związane:

- Jyväskylä University Library (Finlandia) - odpowiedzialna za Finnish Virtual Library, obecnie Science LinkHouse (<http://www.linkkitalo.fi/>),

- Viikki Science Library, University of Helsinki (Finlandia) - partner NOVAGate, norweskiego systemu z zakresu leśnictwa, weterynarii oraz rolnictwa (<http://novagate.nova-university.org/>),
- Niedersächsische Staats- und Universitätsbibliothek Göttingen (Niemcy) - odpowiedzialne za SSG-FI (<http://www.sub.uni-goettingen.de/ssgfi/>): Anglistik Guide, Geo-Guide, History Guide oraz Math-Guide,
- Zentralstelle für Agrardokumentation und -information, ZADI (Niemcy) - partner systemu DAINet (<http://www.dainet.de/>), niemieckiej sieci informacyjnej z zakresu rolnictwa,
- Institute for Learning and Research Technology, University of Bristol (Wielka Brytania) - odpowiedzialne za SOSIG - the Social Sciences Information Gateway ([www.sosig.ac.uk](http://www.sosig.ac.uk)), który jest częścią Resource Discovery Network (RDN),
- UK Office for Library & Information Networking, University of Bath (Wielka Brytania) - instytucja związana z Resource Discovery Network (RDN).

Zaplecze techniczne:

- Technical Knowledge Centre and Library, DTV (Dania),
- NETLAB, Lund University (Sweden) - odpowiedzialne za system EELS, the Engineering E-library (<http://eels.lub.lu.se/>),
- Centre for Scientific Computing CSC (Finlandia).

DSH uczestniczące w projekcie to [Renardus Participating]:

- Anglistik Guide: Virtual Library of Anglo-American Culture,
- DAINet - German Agricultural information Network,
- Danmarks Elektroniske Forskningsbibliotek (deff),
- DutchESS - Dutch Electronic Subject Service,
- Finnish Virtual Library,
- Geo-Guide: Virtual Library of Earth Sciences, Geography, Thematic Maps, and Mining,
- History Guide:
- Virtual Library of Anglo-American Culture,
- MathGuide,
- NOVAGate - Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences (NOVAGate),
- Online theses, Germany (DNB-Theses),
- Resource Discovery Network (RDN).

System *Renardus* udostępniał wspólny interfejs dla przeszukiwania zawartości zrzeszonych DSH oraz interesującą konstrukcję narzędzia dostępu przedmiotowego zbudowaną na Klasyfikacji Dziesiątej Deweya. Niestety od 2007 roku zawieszono realizowanie projektu i tym samym udostępnianie systemu.

### **3.7. Polskie dziedzinowe systemy hipertekstowe**

W odniesieniu do polskich dziedzinowych systemów hipertekstowych można mówić o siedmiu liczących się systemach. Są to:

- BazTOL – system w trakcie wdrażania. Zakres: architektura i urbanistyka, automatyka i robotyka, biotechnologia, budownictwo, chemia, elektronika i telekomunikacja, elektrotechnika i energetyka, fizyka, geodezja i kartografia, górnictwo i geologia, informatyka, inżynieria i ochrona środowiska, inżynieria materiałowa, oceanologia i oceanotechnika. Powstaje dzięki współpracy Biblioteki Głównej Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy, Biblioteki Głównej Politechniki Poznańskiej oraz Biblioteki Głównej Politechniki Warszawskiej;
- Ekonomia Online (<http://kangur.ae.krakow.pl/Biblioteka/Ekonomia/>) – DSH opracowany i prowadzony przez zespół bibliotekarzy Biblioteki Uniwersytetu Ekonomicznego w Krakowie. Zakres systemu jest ograniczony do nauk ekonomicznych;
- Historicus (<http://historicus.umk.pl/>) – portal historyczny pełniący również funkcję DSH opracowany przez pracowników Zakładu Metod Komputerowych Instytutu Historii i Archiwistyki Uniwersytetu Mikołaja Kopernika w Toruniu;
- InfoRol (<http://platan.cbr.edu.pl:4505/ALEPH/-/start/inrpw>) - powstał z inicjatywy Centralnej Biblioteki Rolniczej im. Michała Oczapowskiego w Warszawie oraz dzięki wsparciu finansowemu Fundacji Współpracy Polsko-Niemieckiej. [Błaszczuk i in. 2001] Zakres systemu obejmuje: rolnictwo, przetwórstwo rolno-spożywcze, ochronę środowiska, sadownictwo i ogrodnictwo, leśnictwo, rybołówstwo i rybactwo, integrację Polski z UE itp. ale także np.: rynki, ubezpieczenia, kredyty, prawodawstwo oraz wszystko, co jest związane z rolnictwem i szeroko pojętą gospodarką żywnościową;
- KINIA Katalog Internetowy Nauki o Informacji (<http://www.kinia.czytelnia.net/>) - jest to inicjatywa pracowników Instytutu Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu Marii Curie-

-Skłodowskiej w Lublinie - Grzegorza Gmitterka, Sebastiana Kotuły, Jarosława Packa oraz Piotra Zielińskiego. System udostępnia informacje o dokumentach internetowych z zakresu szeroko pojętej informacji naukowej i bibliotekoznawstwa. Odesłania do stron internetowych zgrupowano w narzędziu dostępu przedmiotowego stworzonym na podstawie Klasyfikacji Dziesiętnej Deweya;

- e-Przewodnik. Serwis WiMBP im. J. Piłsudskiego w Łodzi (<http://www.wimbp.lodz.pl/informacja/>) – opracowywany przez pracowników Wojewódzkiej i Miejskiej Biblioteki Publicznej im. J. Piłsudskiego w Łodzi jest przykładem DSH o zakresie uniwersalnym. Struktura narzędzia dostępu przedmiotowego jest oparta na Uniwersalnej Klasyfikacji Dziesiętnej;
- TOL – Informator Technika Online - jest projektem Biblioteki Politechniki Poznańskiej, który dedykowany jest studentom I, II i III stopnia studiów oraz pracownikom nauki. Pełni on funkcję przewodnika w dostępie do dziedzinowych źródeł informacji dla środowiska użytkowników politechnik polskich (w obszarach dydaktyka – badania). [Ganińska i in. 2006] Zakres systemu obejmuje wybrane dziedziny nauk technicznych, na które składają się: architektura i urbanistyka, automatyka i robotyka, budownictwo, elektronika i telekomunikacja, elektrotechnika, fizyka techniczna, informatyka, inżynieria, ochrona środowiska, inżynieria materiałowa, mechanika i budowa maszyn, technologia chemiczna, transport, zarządzanie, marketing i inżynieria produkcji;
- ELISAD - europejski portal na temat alkoholu, narkotyków i uzależnień ([http://www.addictionsinfo.eu/startpage.php?select\\_lang=pol](http://www.addictionsinfo.eu/startpage.php?select_lang=pol)). Jest systemem o zasięgu europejskim koordynowanym w Polsce przez przedstawiciela Instytutu Psychiatrii i Neurologii w Warszawie. Udostępnia informacje również o polskich zasobach WWW z tego zakresu tematycznego stosując jednocześnie polski interfejs użytkownika. Narzędzie dostępu przedmiotowego ma formę hierarchicznego układu kategorii, których nazwy przyjmują formę zależną od języka interfejsu (dostępnych jest 17 wersji językowych).

Przedstawione polskie DSH należą do stosunkowo młodych systemów, które aspirują do miana wartościowych źródeł informacji o zasobach polskiego Internetu. Gromadzą również odesłania do dokumentów z jego światowych zasobów, lecz priorytet stanowią dokumenty polskojęzyczne. Najbardziej zaawansowanym koncepcyjnie jest zdaniem autora system BazTOL, który zostanie niebawem

uruchomiony. Opracowanie jego struktury było bowiem przeprowadzane zgodnie z wytycznymi projektu DESIRE, który wyznaczył rozwiązania standardowe w tym zakresie.

### **3.8. Podsumowanie**

Przedstawiona charakterystyka dziedzinowego systemu hipertekstowego pozwala na zakwalifikowanie go do rodzaju dokumentacyjnych systemów informacyjno-wyszukiwawczych. Zasadniczą cechą omawianego rodzaju systemów jest zaspokajanie potrzeb informacyjnych internautów poprzez udostępnianie wysokiej jakości metainformacji o obiektach informacyjnych z danej dziedziny wiedzy czy też działalności ludzkiej opublikowanych w WWW. Selekcja i ocena indeksowanych zasobów WWW dokonywana jest na podstawie opracowanych kryteriów, jakie muszą spełniać źródła, aby zostały włączone do zbioru informacyjnego systemu. Wraz z zasadami katalogowania opartymi na funkcjonujących standardach, stanowią one element polityki zarządzania systemem, która ma bezpośredni wpływ na jego funkcjonalność.

Zagadnienia dziedzinowych systemów hipertekstowych można również odnieść do zjawiska, które określane jest mianem *Bibliografii 2.0*. [zob. np. Pacek 2008] Dotyczy ono zastosowania narzędzi i metod charakterystycznych dla bibliografii, jako specjalności naukowej w nowym środowisku informacyjnym, którym jest Internet. Jest to zarówno aspekt nowego rodzaju narzędzi wykorzystywanych do tworzenia bibliografii, jak również nowego typu obiektów informacyjnych. Tym samym można porównywać do pewnego stopnia funkcjonowanie dziedzinowych systemów hipertekstowych z pojęciem Bibliografia 2.0 czy też Netografia (jako spis źródeł internetowych). Intencją autora, co zostało sprecyzowane we wstępie rozprawy, jest jednak przyjęcie perspektywy zorientowanej systemowo i utożsamianie DSH, jako określonego typu systemów informacyjno-wyszukiwawczych.

## CZEŚĆ DRUGA

Przedmiotem drugiej części rozprawy jest język informacyjno-wyszukiwawczy jako narzędzie organizacji informacji w dziedzinowych systemach hipertekstowych. O ile pierwsza część rozprawy przyniosła podstawy teoretyczne organizacji informacji i zdefiniowanie dziedzinowego systemu hipertekstowego, to w drugiej zostaną przedstawione wyniki badań przeprowadzonych przez autora nad wykorzystaniem JIW we wskazanej klasie systemów informacyjno-wyszukiwawczych. Przyjęta perspektywa badawcza oraz wykorzystana metodologia i terminologia wymagają jednak uzasadnienia. W związku z tym, dalsze rozważania zostaną poprzedzone charakterystyką aspektu metodologicznego rozprawy.

Przedstawiona we wstępie charakterystyka metodologiczna rozprawy zakładała zastosowanie studium analitycznego jako podstawowego sposobu badania dziedzinowego systemu hipertekstowego. Efektem przedstawionych dotychczas analiz było opracowanie modelu analitycznego dziedzinowego systemu hipertekstowego uwzględniającego jego elementy strukturalne. Mając zdefiniowany przedmiot badań, należy dążyć do potwierdzenia postawionej hipotezy o wykorzystaniu JIW w organizacji informacji w tej klasie systemów. Umożliwi to przyjęcie koncepcji organizacji informacji, na którą składają się procesy formatowania jednostki opisu, reprezentacji informacji oraz organizacji dostępu za pomocą odpowiednich mechanizmów. Na podstawie analizy dokumentacji wybranych systemów oraz krytyki piśmiennictwa w tym zakresie, zostanie wskazane zastosowanie JIW we wspomnianych etapach składających się na organizację informacji w dziedzinowych systemach hipertekstowych.

Celem badań nie było jednak wskazanie metod wykorzystania konkretnych JIW w reprezentacji informacji w określonej grupie systemów. Tym samym np. szczegółowe zagadnienia indeksowania z wykorzystaniem JIW zostały jedynie zasygnalizowane. Pominięte kwestie, np. szerokości i głębokości indeksowania z wykorzystaniem JIW w środowisku sieciowym stanowią odrębne i obszernie tematy do dalszych badań. W tym miejscu warto podkreślić, że celem rozprawy było wskazanie udziału JIW w DSH w szerszym kontekście, który wyznacza proces organizacji informacji.

Druga część badań przeprowadzonych przez autora została poświęcona



narzędziom dostępu przedmiotowego w DSH i sposobom zastosowania w ich budowie JIW. Grupę reprezentatywną poddaną szczegółowym badaniom stanowiło 30 wyselekcjonowanych systemów [Zał. 2], które spełniają kryteria włączenia do omawianej klasy systemów oraz wykorzystują język informacyjno-wyszukiwawczy w organizacji informacji. Badania przeprowadzono w okresie od 01-06-2006 r. do 01-03-2007 r. Analiza struktury narzędzi dostępu przedmiotowego w dziedzinowych systemach hipertekstowych miała na celu wskazanie sposobów wykorzystania języków informacyjno-wyszukiwawczych i określenia ich roli w procesie porządkowania i organizowania dostępu do zasobów systemu.

Metodyka badań w pierwszej fazie polegała na ekstrakcji narzędzi dostępu przedmiotowego z interfejsów DSH stanowiących grupę reprezentatywną. W tym celu manualnie przeniesiono struktury tych narzędzi do postaci arkuszy kalkulacyjnych programu Microsoft Excel. Po otrzymaniu wyników o charakterze ilościowym przystąpiono do jakościowej analizy wyodrębnionych struktur. Ze względu na formę organizacji przedmiotowych punktów dostępu wyodrębniono trzy warianty ich porządkowania w narzędziach. Są to: wariant ahierarchiczny, hierarchiczny oraz hybrydowy.

Każde z badanych narzędzi analizowano pod kątem:

- identyfikacji typu języka informacyjno-wyszukiwawczego wykorzystanego do budowy,
- identyfikacji wykorzystanego konkretnego języka informacyjno-wyszukiwawczego,
- sposobu porządkowania punktów dostępu,
- budowy struktur ahierarchicznych,
- budowy struktur hierarchicznych:
  - o łącznej liczby kategorii w schemacie,
  - o liczby kategorii na pierwszym stopniu podziału,
  - o przynależności kategoryjnej działów na pierwszym stopniu podziału,
  - o porządku kategorii,
  - o liczbowego i procentowego rozkładu kategorii na poszczególnych poziomach w strukturze,
  - o pojemności kategorii na poszczególnych poziomach w strukturze (liczba odesłań do ChWD),
  - o średniej pojemności kategorii na poszczególnych poziomach w strukturze,

- dystrybucji ChWD na poszczególnych poziomach w strukturze,
  - średniej liczby kategorii, do których przyporządkowano jedną ChWD,
- budowy struktur hybrydowych.

Identyfikacja takiego zbioru zmiennych poddanych przez autora badaniom również wymaga komentarza. Badania nad budową i efektywnością sieciowych systemów organizacji wiedzy, do jakich należą niewątpliwie narzędzia dostępu przedmiotowego w DSH, nie doczekały się w piśmiennictwie naukowym opracowania spójnej metodyki postępowania. W dokonanym przez autora wyborze cech narzędzi dostępu przedmiotowego które poddano analizie, wykorzystano badania prowadzone przez A. Wheatleya i D. Vizine-Goetz [Wheatley 2000; Vizine-Goetz 2002].\* Zestaw ten został rozbudowany o dodatkowe zmienne, których wartości, zdaniem autora dopełniają obraz struktury narzędzi dostępu przedmiotowego w DSH.

Uzyskane w ten sposób dane pozwoliły na weryfikację hipotezy głównej oraz na opracowanie odpowiedzi na postawione w pracy pytania badawcze dotyczące sposobów wykorzystania języka informacyjno-wyszukiwawczego w organizacji informacji w narzędziach dostępu przedmiotowego w dziedzinowych systemach hipertekstowych.

Poważnym problemem, jaki należało rozwiązać w rozprawie była kwestia zastosowania odpowiedniej terminologii w odniesieniu do narzędzi dostępu przedmiotowego. Są to stosunkowo nowe narzędzia organizacji wiedzy i nadal nie doczekały się w piśmiennictwie naukowym ogólnie akceptowanej definicji i spójnej terminologii. Podczas analizy narzędzi dostępu przedmiotowego na określenie punktu dostępu przyjęto termin *kategoria tematyczna*. [por. Sosińska-Kalata 2002; 2002a] Celowo odrzucono termin *klasa*, który jest elementem struktury języków informacyjnych typu klasyfikacji, i który utożsamia się z podziałem logicznym. Podział kategorii tematycznych w hierarchicznych narzędziach dostępu nie posiadał takiej cechy, w związku z tym zastosowanie terminu *klasa* byłoby nadużyciem. Analiza struktury tego typu narzędzi została jednak przeprowadzona z zastosowaniem innych terminów występujących w teorii klasyfikacji (np. łańcuch i ogniwo). Celem takiego zabiegu było wskazanie na podobieństwa w budowie tych dwóch rodzajów struktur porządkujących.

Dodatkowym problemem, który pojawił się podczas zbierania danych przez autora była permanentna modyfikacja narzędzi dostępu przedmiotowego

---

\* zob. Rozdz. 2.7.3.2.2.

w dziedzinowych systemach hipertekstowych. Spowodowane jest to przyrostem elementów zbioru wyszukiwawczego i tym samym pojawianiem się nowych kategorii tematycznych dla zaindeksowanych obiektów informacyjnych. Tym samym zmianom ulegają również struktury narzędzi, których elementami są kategorie tematyczne. W związku z tym, przeprowadzone przez autora badania miały charakter synchroniczny, czyli reprezentują strukturę tego typu narzędzi, ale wyłącznie w momencie badania. Ich wyniki mogą stanowić podstawę do badań porównawczych dotyczących np. tendencji w zmianach w budowie tego typu narzędzi w DSH.

## **4. Wykorzystanie języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych**

Dostęp do zasobów systemu jest kluczowym elementem realizacji procesów informacyjnych przez użytkownika podczas jego interakcji z systemem. Pozyskiwanie informacji z zasobów dziedzinowych systemów hipertekstowych odbywa się poprzez wykorzystanie wyszukiwarki i/lub narzędzi dostępu przedmiotowego. Efektem procesu wyszukiwania informacji w tego rodzaju systemach informacyjnych jest zbiór relewantnych charakterystyk wyszukiwawczych dokumentów. W tej części pracy zostanie scharakteryzowana organizacja punktów dostępu do zasobów dziedzinowego systemu hipertekstowego jako jednego z elementów organizacji informacji oraz rola języka informacyjno-wyszukiwawczego w tym procesie. Scharakteryzowana w pierwszym rozdziale koncepcja punktów dostępu zostanie w tej części pracy odniesiona do organizacji informacji w DSH. Jednocześnie w centrum rozważań będzie pozostawał JIW jako narzędzie służące do jej realizacji.

### ***4.1. Miejsce i rola języka informacyjno-wyszukiwawczego w dziedzinowych systemach hipertekstowych***

Zakres wykorzystania JIW w dziedzinowych systemach hipertekstowych jest wyznaczany poprzez jego udział w procesach realizowanych w ramach organizacji informacji w systemie. Tym samym miejsce JIW w tego rodzaju systemach można zidentyfikować poprzez wskazanie jego udziału w:

- reprezentacji informacji,

- identyfikacji cech wyszukiwawczych/punktów dostępu oraz
- organizacji tych punktów w narzędziach dostępu.

Pierwsze dwa sposoby zastosowania JIW wynikają z realizacji jego podstawowych funkcji: metainformacyjnej i wyszukiwawczej. Organizacja punktów dostępu z wykorzystaniem JIW w narzędziach dostępu do zbioru informacyjnego jest jednym z nowych sposobów adaptacji tego typu języków sztucznych w środowisku cyfrowym. Element nowatorski polega tutaj wyłącznie na przeniesieniu funkcji heurystycznej oraz organizującej JIW do środowiska sieciowego.

Jedną z charakterystycznych cech DSH w kontekście dostępu do ich zasobów jest funkcjonowanie narzędzi dostępu przedmiotowego jako komplementarnego w stosunku do mechanizmów wyszukiwawczych. Rola JIW w organizacji i dostępie do informacji w dziedzinowych systemach hipertekstowych zostanie przedstawiona tutaj w kontekście rozważań nad reprezentacją informacji, identyfikacją cech wyszukiwawczych oraz budową narzędzi dostępu przedmiotowego.

## ***4.2. Reprezentacja informacji***

Efektem prac nad DSH w ramach projektu badawczego DESIRE było m.in. opublikowanie raportu pt. „Information Gateways Handbook”. [Belcher i in. 1999] Zawiera on szereg wytycznych, jakimi powinni kierować się twórcy DSH. Wśród problemów związanych z zarządzaniem DSH są również zagadnienia katalogowania zasobów Internetu oraz wykorzystania w tym procesie języków informacyjnych. Na podstawie wytycznych wspomnianego raportu oraz analizy dokumentacji wybranych DSH dokonano charakterystyki reprezentacji informacji z zastosowaniem JIW oraz zbiorów słownictwa o różnym stopniu kontroli. Jej zadaniem było wskazanie na rzeczywiste wykorzystanie JIW w dziedzinowych systemach hipertekstowych w reprezentacji informacji. Przyjęta perspektywa zakłada identyfikację i charakterystykę JIW wykorzystywanych w DSH do reprezentacji poszczególnych typów atrybutów obiektów informacyjnych. Są to: nazwy osobowe i korporatywne, typ zasobów, język zasobów, daty, nazwy geograficzne, treść oraz dodatkowe cechy.

### **4.2.1. Nazwy osobowe i korporatywne**

Dziedzinowe systemy hipertekstowe gromadzące odesłania i charakterystyki

wyszukiwawcze zasobów WWW wykorzystują najczęściej reguły tworzeniu haseł osobowych zamieszczone w drugim wydaniu Anglo-American Cataloguing Rules (AACR2). Jest to charakterystyczne dla systemów wykorzystujących język angielski, jako język interfejsu użytkownika oraz język dokumentów tworzących zbiór informacyjny. Zastosowanie JIW w reprezentacji tego typu metainformacji polega na wykorzystaniu istniejących kartotek wzorcowych w tym zakresie. Do największych i najpopularniejszych źródeł leksykalnych tego typu stosowanych w DSH należą Library of Congress Name Authority File (LCNAF) oraz British Library Name Authority File (BLNAF). Kartoteki wzorcowe nazw osobowych są stosowane w DSH jako źródło leksyki do odwzorowywania przede wszystkim tematów osobowych i korporacyjnych, a w drugiej kolejności do kontrolowania formy nazwisk autorów publikacji internetowych i nazw instytucji związanych z ich powstaniem i rozpowszechnianiem. Podstawowym źródłem słownictwa w tym zakresie stosowanym w dużej liczbie anglojęzycznych DSH jest Library of Congress Name Authority File (LCNAF). Ta kartoteka jest wykorzystywana w obecnie największym DSH – Intute (<http://www.intute.ac.uk/>) (oraz w wielu grupach tematycznych składających się na jego kolekcję). LCNAF jest wykorzystywane również w DSH funkcjonujących w państwach spoza Ameryki Północnej oraz Wielkiej Brytanii. Taka sytuacja ma miejsce w sieci niemieckich systemów Special Subject Guides / SSG-Fachinformation (<http://www.sub.uni-goettingen.de/ssgfi/index.html>). We wszystkich DSH działających w ramach tego projektu, czyli:

- MathGuide (matematyka) - <http://www.mathguide.de/>,
- GeoGuide (geologia) - <http://www.geo-guide.de/>,
- Anglo-American Culture (kultura anglo-amerykańska) - <http://www.sub.uni-goettingen.de/ssgfi/anglo-americana.html>,
- ForestryGuide (leśnictwo)- <http://www.forestryguide.de/>,

wykorzystuje się Kartotekę Wzorcową Nazw Osobowych Biblioteki Kongresu (LCNAF). Służy ona za źródło leksyki do odwzorowywania atrybutów dokumentów internetowych typu: twórca, współtwórca oraz hasło osobowe. [SSG-FI Project Documentary 1999] W systemie ADAM\*, z zakresu sztuki, architektury oraz mediów, do reprezentacji nazw osobowych zarówno w postaci oznaczenia odpowiedzialności, jak i tematów osobowych, wykorzystuje się kartotekę ULAN: Union List of Artist

---

\* System ADAM funkcjonuje obecnie w ramach grupy tematyczne Intute: Arts and Humanities - <http://www.intute.ac.uk/artsandhumanities/>

Names, która dostępna jest również poprzez serwer Fundacji J.P. Getty'ego.\*\* System ADAM w 2008 roku został włączony w strukturę systemu Intute i tym samym wykorzystuje się tam dwie wspomniane kartoteki wzorcowe nazw osobowych.

Duża liczba DSH nie wykorzystuje zasobów leksykalnych kartotek wzorcowych do reprezentacji wartości dla wskazanych atrybutów, pozostawiając jednocześnie ich formę poza kontrolą. Przyjęte w tych systemach zasady budowy tego rodzaju metadanych opierają się zazwyczaj na funkcjonujących w danych państwach standardach dotyczących katalogowania. Przeważająca liczba DSH wykorzystuje jednak AACR2 oraz zamieszczone tam zasady tworzenia nazw osobowych. Innym przykładem kontrolowania formy nazw osobowych jest konsorcjum Australian Subject Gateways Forum zrzeszające australijskie DSH. Cztery z dziewiętnastu działających tam systemów wykorzystują do odwzorowania nazw osobowych „Międzynarodowy standard archiwalnych haseł wzorcowych ciał zbiorowych, osób i rodzin” – „International Standard Archival Authority Record For Corporate Bodies, Persons and Families ISAAR(CFP)”. [ISAAR(CFP) 2006]

Wspominany wcześniej podręcznik pt. „Information Gateways” zaleca stosowanie w tym zakresie narodowych lub międzynarodowych kartotek wzorcowych nazw osobowych wyróżniając przede wszystkim Library of Congress Name Authority File, British Library Name Authority File oraz kartoteki niemieckiej biblioteki narodowej nazw osobowych (Personennamendatei) oraz nazw ciał zbiorowych (Gemeinsame Körperschaftsdatei). [Belcher 1999, s.48]

#### 4.2.2. Typ zasobów

Reprezentacja wartości dla cechy *typ zasobów* polega na identyfikowaniu formy utrwalenia treści obiektu informacyjnego. W tym celu w DSH wykorzystuje się zbiory słownictwa kontrolowanego w postaci list terminów. Jedną z pierwszych list tego typu wartości był wykaz terminów opracowany dla systemów działających jeszcze w sieci Resource Discovery Network (RDN), czyli m.in. SOSIG oraz BIOME. [Cliff 2002] Zawierał on 14 wartości dla wspomnianej cechy (Tab. 4.1)

Tab. 4.1 Typ zasobów dla systemów sieci RDN. [Źródło: Cliff 2002]

Nazwa wartości	Opis
Journal Full Text	Serwis udostępniający pełne teksty artykułów z czasopism
Journal Abstract Contents	Serwis udostępniający spisy treści czasopism i/lub abstrakty

\*\* [http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)

	artykułów, a także dostęp do wybranych pełnych tekstów
FAQ	Zbiory najczęściej zadawanych pytań przez użytkowników i odpowiedzi na nie
Government Publication	Publikacje rządowe
Document	Pojedynczy dokument
Organization	Serwisy informacyjne organizacji
Learning Material Courseware	Serwisy udostępniające dokumenty o potencjalnym wykorzystaniu edukacyjnym
Mailing List	Listy, grupy i fora dyskusyjne
News	Aktualności informacyjne (np. portali, agencji informacyjnych)
Event Announcement	Kalendaria wydarzeń
Project	Serwisy informacyjne projektów naukowych i edukacyjnych
Database	Bazy danych
Reference	Informatory (bezpośrednie i pośrednie)
Resource Guide	Katalogi internetowe

Przedstawiona w Tab. 4.1 lista wartości dla cechy *typ zasobu* jest jedynie przykładem stosowania zbiorów słownictwa kontrolowanego w DSH. Zasób leksykalny tego typu list jest zróżnicowany. We wspomnianej niemieckiej sieci DSH – SSG-FI wykorzystuje się listę wartości dla tego atrybutu liczącą ponad 40 elementów. Podstawą tworzenia zbioru wartości i jego dalszej rozbudowy może być również DCMI Type Vocabulary jako specyfikacja słownictwa w schemacie Dublin Core. Zawiera ona wykaz 12 wartości, jakie może przyjąć ta cecha w odniesieniu do zasobów cyfrowych. Są to:

- kolekcja,
- zbiór danych,
- wydarzenie,
- ruchome obrazy,
- zasób interakcyjny,
- dokument wizualny,
- obiekt fizyczny,
- usługa,
- oprogramowanie,
- dokument dźwiękowy,
- obraz statyczny,
- tekst.[<http://dublincore.org/documents/dcmi-type-vocabulary/>]

Zakresy niektórych z wymienionych wyrażen zachodzą na siebie, dlatego też dla osiągnięcia większej precyzji stosuje się własne listy wartości opracowywane na potrzeby DSH. W tym celu wykorzystuje się słowniki terminologiczne z zakresu dziedzin zbieżnych z zakresem danego systemu. Np. w systemie GEM – Gateway to

21st Skills (<http://www.thegateway.org/>) z zakresu edukacji wykorzystuje się słowniki terminologiczne oraz informatory bezpośrednie z zakresu szeroko rozumianej edukacji.

Interesującym źródłem gotowych zbiorów słownictwa kontrolowanego odwzorowującego ten typ atrybutu jest publikacja będąca efektem prac ósmych warsztatów Dublin Core w 2000 roku w Ottawie. [Authority 2000] Zawiera ona zbiory słownictwa stanowiące wyciągi m.in. z Tezaurusu ERIC, słownictwa kontrolowanego systemu GEM, wspomnianego zestawu wyrażań Dublin Core, słownictwa stosowanego w Alexandria Digital Library, a także słownictwa kontrolowanego dla cechy *typ zasobu* zalecanego przez schemat metadanych dla obiektów edukacyjnych - Learning Object Metadata. Jest to gotowe źródło leksyki do zastosowania w DSH w celu reprezentacji omawianego atrybutu.

### **4.2.3. Język zasobów**

Język, w jakim wyrażone są treści internetowych obiektów informacyjnych jest kolejną z cech, której reprezentacja odbywa się z wykorzystaniem zbiorów słownictwa kontrolowanego. W DSH odwzorowanie tego atrybutu wiąże się najczęściej z zastosowaniem norm regulujących formy skrótów nazw języków etnicznych. W DSH język dokumentu jest reprezentowany przez trzyliterowy skrót, którego forma jest pobierana z normy ISO 639-2 „Codes for the representation of names of languages lub ze starszych źródeł np. RFC 1766 - Tags for the Identification of Languages”. Systemy, w których nie stosuje się skrótów nazw języka dokumentu internetowego do reprezentacji tej cechy wykorzystują formę nazwy wyrażoną w języku naturalnym, zależną od języka etnicznego interfejsu użytkownika danego systemu.

### **4.2.4. Daty**

W przypadku metainformacji, które muszą zostać wyrażone w postaci dat, w DSH kontroluje się jedynie format ich zapisu. Najczęściej źródłem regulacji w tym zakresie jest norma ISO 8601:1988 „Data elements and interchange formats - Information interchange - Representation of dates and times”. Wykorzystuje się również dokument opracowany przez World Wide Web Consortium [Newman 1997], który odnosi się do zapisu dat dla obiektów cyfrowych i którego podstawą sporządzenia jest wspomniana norma. Zalecana forma dat wyrażana jest poprzez zestawienie RRRR-



MM-DD, które oznaczają kolejno rok, miesiąc oraz dzień. Taka specyfikacja formy omawianego atrybutu zamieszczona jest również w jego charakterystyce w schemacie Dublin Core – DCMI-Period (<http://dublincore.org/documents/dcmi-period/>).

#### **4.2.5. Nazwy geograficzne**

Reprezentacja wartości dla atrybutów dotyczących zasięgu geograficznego internetowych obiektów informacyjnych najczęściej nie jest kontrolowana w DSH i przyjmuje formę odpowiednich wyrażen języka naturalnego. W sytuacji, kiedy twórcy systemu decydują się na nadzór formalny jednostek leksykalnych, najczęściej wykorzystywanym źródłem słownictwa jest Getty Thesaurus of Geographic Names, który jest stosowany m.in. w systemach skupionych w Intute czy też australijskim Edna Online. Innym wariantem kontroli leksykalnej w tym zakresie jest oznaczanie nazw państw, które wyznaczają zasięg geograficzny za pomocą dwu- lub trzyliterowych kodów pochodzących z normy ISO 3166-1 „Kody nazw krajów i ich jednostek administracyjnych - Część 1: Kody krajów” wraz z ich tłumaczeniami na poszczególne języki etniczne.\* Zalecenia cytowanego już podręcznika pt. „Information Gateways Handbook” wskazują również na odwzorowywanie szerokości i długości geograficznej obiektów będących przedmiotem dokumentów internetowych. W tym przypadku cytuje się postanowienia schematu Dublin Core: Dcmi-point (<http://dublincore.org/documents/dcmi-point/>)

#### **4.2.6. Treść. Polireprezentacja wiedzy**

Reprezentacja treści dokumentów internetowych w DSH charakteryzuje się dużą heterogenicznością przede wszystkim pod względem rodzajów narzędzi używanych do tego celu oraz sposobów ich wykorzystania. Stosuje się tutaj różne typy języków informacyjno-wyszukiwawczych, ale istnieje jednak wyraźna tendencja do wykorzystywania JIW o notacji paranaturalnej.

Wyniki badań przeprowadzonych przez autora na grupie reprezentatywnej trzydziestu DSH (Załącznik 2.) pokazują, że odwzorowanie cech treści w tych systemach charakteryzuje się jednoczesnym wykorzystaniem kilku JIW. To zjawisko

---

\* Np. w języku niemieckim: DIN EN 23166:1995 stosowane w systemach sieci SSG-FI (<http://www.transpatent.com/landverz.html>).

znane jest pod nazwą *polireprezentacji wiedzy* i zostanie zanalizowane w tym rozdziale w odpowiednim miejscu. Zastosowanie wielu JIW w reprezentacji wiedzy w DSH powoduje trudności w bezpośrednim określeniu udziału ilościowego poszczególnych typów języków w tym procesie. Dane przedstawione w Tab. 4.2. uwzględniają rzeczywiste wskazania ilościowe poszczególnych typów JIW w analizowanej grupie reprezentatywnej DSH.

Tab. 4.2. Udział JIW w reprezentacji wiedzy w DSH [Źródło: badania własne]

	Język haseł przedmiotowych	Język deskryptorowy	Klasyfikacja	Język słów kluczowych	
				kontrolowane	niekontrolowane
Liczba wskazań	14	18	19	7	9
Procentowo	47%	60%	63%	23%	30%

Podczas badania odnotowano łącznie 83 wystąpienia JIW, które przyporządkowano do czterech typów JIW:

- języka haseł przedmiotowych,
- języka informacyjno-wyszukiwawczego typu klasyfikacji,
- języka deskryptorowego (a bezpośrednio jego słownika, czyli tezaurusa),
- języka słów kluczowych w postaci
  - o kontrolowanej oraz
  - o niekontrolowanej.

Grupa reprezentatywna liczyła 30 DSH, a zidentyfikowano w niej użycie 83 JIW do reprezentacji informacji. Zatem średnio na jeden analizowany DSH przypadły ponad dwa JIW użyte do reprezentacji treści.

Największy udział JIW w reprezentacji wiedzy odnotowano w stosunku do klasyfikacji. Może się to wydawać zastanawiające ze względu na szerokie rozpowszechnienie w Internecie języka słów kluczowych oraz na trudności w recepcji zakresu znaczeniowego symboli klasyfikacyjnych jako systemu semiotycznego o notacji sztucznej. W DSH mamy do czynienia jednak ze specyficznym użyciem tego języka. Dane w tabeli 4.2. wskazują na zbliżony poziom udziału zarówno klasyfikacji, jak i języków deskryptorowych w reprezentacji wiedzy. Wynosi on około 60% wskazań, natomiast 53% wystąpień JIW to zastosowanie języka słów kluczowych. Szczegółowa analiza wykorzystania tego typu JIW wykazała, że w 23% wystąpień był to system leksykalny w postaci kontrolowanej, a w 30% był to język swobodnych słów kluczowych. W 47% wskazań odnotowano zastosowanie języka haseł przedmiotowych.

Pomimo ponad 50% wskazań wykorzystania języka słów kluczowych

w reprezentacji wiedzy w DSH jest on językiem, który stanowi zazwyczaj drugi typ narzędzia w polireprezentacji wiedzy. Forma wyrażenia tego języka przyjmuje postać najczęściej jedno- lub dwuwyrzowych jednostek leksykalnych. Gramatyka swobodna (niepozycyjna) jest charakterystyczną regułą tworzenia tekstów w tym języku informacyjnym. Jego wykorzystanie w odwzorowaniu treści elementów zbioru informacyjnego DSH polega na sporządzeniu komunikatu za pomocą jednostek leksykalnych języka słów kluczowych, które pozostają wobec siebie równoważne, i które bezpośrednio identyfikują przedmiot (lub przedmioty) internetowego obiektu informacyjnego oraz nazwy dla dodatkowych cech jego treści.

Obok języka swobodnych słów kluczowych w DSH występują również kontrolowane systemy leksykalne tego rodzaju. Wykorzystuje się tutaj m.in.:

- ICOM International Committee for the Museums and Collections of Costume Vocabulary – system Artifact,
- język słów kluczowych International Bibliography of the Social Sciences (IBSS) - system Intute Social Sciences,
- Military Science Index – system AERADE,
- Moving Image Genre-Form Guide – system Artifact,
- Social Care Institute for Excellence Vocabulary, Statistics and Research Methods Vocabulary - system Intute Social Sciences,
- Schools Key Learning Areas – system Edna Online.

Zastosowanie języka haseł przedmiotowych do odwzorowania treści zasobów DSH cechują zapożyczenia syntaktyczne z języka słów kluczowych. Polega to na redukcji hasła przedmiotowego do postaci tematu, któremu rzadko towarzyszy grupa określników. Tym samym zdania języka haseł przedmiotowych reprezentujące treść obiektów informacyjnych w DSH przyjmują postać wykazu tematów reprezentujących ich przedmiot. Kiedy dochodzi do wykorzystania jednostek synsyntaktycznych w postaci określników, hasło przedmiotowe ma postać najczęściej dwuelementową (temat + określnik). Funkcję elementu niesamodzielnego pełnią najczęściej określniki klasowe\*. Stosuje się również określniki rzeczowo-treściowe oznaczające ogólne właściwości wspólne dla wielu tematów, np. w systemie MedHist (<http://www.intute.ac.uk/healthandlifesciences/medhist/>) z zakresu historii medycyny, gdzie do indeksowania wykorzystuje się język haseł przedmiotowych MeSH.

---

\* „Mają one postać nazw dziedzin wiedzy lub działalności. Wyrażają [...] najczęściej punkt widzenia, ujęcie lub reprezentują jakieś cechy i własności przedmiotów uogólnione do nazwy dziedziny”. [Sadowska, Turowska 1990, s. 62].

Przykładowe hasła przedmiotowe z jednym określnikiem przyjmują formę:

*Mental Disorders / rehabilitation*

*Malaria / epidemiology*

Zachodzi tutaj pełna zgodność w planie wyrażania jednostek z ich odpowiednikami w słowniku MeSH. Jednym z nielicznych DSH, które stosują tak rozbudowaną grupę określnikową jest system Infomine (<http://infomine.ucr.edu/>). Wykorzystywany jest w nim m.in. język haseł przedmiotowych LCSH. Teksty tego języka mają postać zdań - rozwiniętych haseł przedmiotowych, np.

*Abnormalities, Human -- Research -- Periodicals*

*Teratology -- Periodicals*

Szyk grupy określnikowej jest zgodny z regułami gramatycznymi łączenia jednostek leksykalnych LCSH.

Kategoria leksykalna określnika chronologicznego, geograficznego oraz formalnego jest bardzo rzadko wykorzystywana w budowie zdań z wykorzystaniem tego typu JIW. Status formalny dokumentu internetowego odwzorowuje się w specjalnym polu ChWD, któremu odpowiada atrybut *typ zasobu*. Jednym z nielicznych wyjątków od tej reguły jest wspomniany system Infomine. Stosuje się tutaj rozwinięte hasła przedmiotowe, w których grupę określników tworzą różne rodzaje określników rzeczowych, a także określniki geograficzne i formalne. Nie odnotowano wystąpień określników chronologicznych. Format opisu oraz jego wizualizacja w postaci ChWD w tym systemie zawierają dodatkowo odpowiedni atrybut w postaci cechy *rodzaj zasobu* wraz z jego wartością nadaną podczas indeksowania. Dochodzi tym samym do redundancji informacji, ale z jednoczesnym zachowaniem spójności składniowej hasła przedmiotowego.

Udział języka haseł przedmiotowych w badanej grupie DSH jest zdominowany poprzez dwa najbardziej popularne języki z tej klasy. Są to Medical Subject Headings oraz Library of Congress Subject Headings.

Zastosowanie języka deskryptorowego w reprezentacji wiedzy w DSH to przede wszystkim wykorzystanie tezaurusów jako źródła leksyki do indeksowania. Odzworowanie treści za pomocą tego JIW polega na wykorzystaniu określonych deskryptorów w układzie równoważnych jednostek leksykalnych, tworzących zdania

tego języka. W DSH nie stosuje się dodatkowych wskaźników np. roli, więzi czy wagi do odwzorowania zależności semantycznych pomiędzy wyrażeniami pochodzącymi z tezaury. Forma jednostek leksykalnych jest zgodna z ich odpowiednikami w słowniku wykorzystanego JIW.

Wykorzystanie w DSH języków deskryptorowych sprowadza się najczęściej do wykorzystania Tezaury CAB (systemy: Vetgate / Intute Veterinary, Intute Health and Life Science, Intute Social Science, Agrifor) oraz Humanities and Social Science Electronic Thesaurus - HASSETT (systemy: Altis, Intute Social Science, Artifact). Dodatkowo odnotowano użycie Arts and Architecture Thesaurus (system ADAM), British Educational Thesaurus (system GEES), Research Methods & Tools – SRM Thesaurus (system Intute Social Science), Royal College of Nursing Thesaurus, Research Methods & Tools (system Intute Health and Life Sciences, Nursing, Midwifery and Allied Health), Thesaurus for Graphic Materials (system Artifact), UKAT UK Archival Thesaurus (system Artifact) oraz Thesaurus Social Care Online (system Intute Social Sciences).

Zastosowanie języka informacyjno-wyszukiwawczego typu klasyfikacji w DSH do odwzorowania treści obiektów informacyjnych charakteryzuje się specyficznym sposobem użycia jego leksyki. Pierwszy sposób zakłada reprezentację wiedzy zapisanej w dokumentach za pomocą odpowiednich symboli zbudowanych z zasobu leksykalnego danej klasyfikacji. Jest to standardowe wykorzystanie tego typu JIW. W ChWD teksty wyrażone w wybranych JIW typu klasyfikacji przyjmują postać jednego lub dwóch symboli klasyfikacyjnych wskazujących na zakres tematyczny obiektu informacyjnego. W DSH sprowadza się to do opracowania odpowiednich symboli prostych i/lub rozwiniętych\*, których zakres odpowiada zakresowi treściowemu oznaczanego dokumentu. Nie stosuje się wyznaczania zakresu dla tematu poprzez iloczyn czy sumę zakresów symboli tworzących symbol złożony (Rys. 4.1).

---

\* Symbol prosty – symbol klasyfikacyjny utworzony z jednego symbolu głównego [...]. Symbol rozwinięty – symbol klasyfikacyjny utworzony z jednego symbolu głównego, do którego przyłączono symbol lub symbole pomocnicze [...] [Słownik encyklopedyczny 2002, s. 256]

### Against the Grain

Against the Grain disseminates news about libraries, publishers, book jobbers, and subscription agents, with a collection of reports on the issues, literature, and people that impact the world of books and journals. ATG is published six times a year.

Author: ATG

→ DeweyClass: 020

### BBC News: Election 2005

Offers news stories and other information resources relating to the UK General Election campaign of 2005.

Author: BBC

→ DeweyClass: 324

### Biblio Tech Review

Information on key technologies affecting the information industry.

Author: Biblio Tech

→ DeweyClass: 025.04

### El Profesional de la Informacion

A Spanish language journal focussing on information technology. Abstracts free. Subscription required for full-text access.

Author: Swets & Zeitlinger

→ DeweyClass: 020

Rys. 4.1. Użycie Klasyfikacji Dziesiątej Deweya do odwzorowania treści w systemie BUBL.

[Źródło: <http://bubl.ac.uk/>]

W przedstawionych na Rys. 4.1 ChWD w systemie BUBL, treść dokumentów jest reprezentowana poprzez odpowiednie symbole Klasyfikacji Dziesiątej Deweya. Odpowiednia etykieta pola rekordu, tutaj *DeweyClass*, identyfikuje atrybut, który podczas indeksowania przyjmuje wartość określonego symbolu klasyfikacyjnego. Analogiczna sytuacja ma miejsce w odniesieniu do systemu GeoGuide (Rys. 4.2)

### Bentonite & Tonstein Bibliography

#### BIBLIOGRAPHIC DATA

<b>Description</b>	This homepage offers a bibliography for bentonites and tonsteins.
<b>URL</b>	<a href="http://www.st-and.ac.uk/~www_sgg/bentrefs.html">http://www.st-and.ac.uk/~www_sgg/bentrefs.html</a>
<b>Publisher</b>	School of Geography & Geosciences - University of St. Andrews: St. Andrews, GB
<b>Distributor</b>	University of St. Andrews: St. Andrews, GB
<b>Language</b>	English
<b>Country</b>	United Kingdom
<b>Format</b>	text/html

#### Search for similar sites by:

##### Keywords

- bentonite
- tonstein
- bibliography
- sedimentary petrology

##### Source Type

- Current Bibliographic and Abstracting Series

##### Subject Class

- Sedimentary Petrology
- Mineralogy

##### GOK

- VKB 300
- VF 000

#### ADDITIONAL INFORMATION

<b>Access</b>	free; no restriction
<b>Level</b>	popular; undergraduate

#### STATISTICS

<b>Timestamp</b>	2001-08-28	<b>Source</b>	SUB
------------------	------------	---------------	-----

Rys. 4.2. Użycie Göttinger Online Klassifikation do odwzorowania treści w systemie GeoGuide.

[Źródło: <http://www.geo-guide.de/>]

Charakterystyki wyszukiwawcze dokumentów w systemie GeoGuide (Rys. 4.2) zawierają pole o etykiecie *GOK*, co jest akronimem dla nazwy *Göttinger Online*

*Klassifikation*. Klasyfikacja ta została zastosowana w systemie do reprezentacji cech wymiaru treści elementów zbioru informacyjnego systemu. Zatem wartości dla tej cechy we wspomnianym polu przyjmują formę odpowiednich symboli klasyfikacji GOK.

Drugim sposobem wykorzystania klasyfikacji do indeksowania elementów zbioru informacyjnego DSH jest dołączanie do odpowiednich symboli klasyfikacyjnych ich odpowiedników słownych lub zupełne zastępowanie ich wyrażeniami języka naturalnego. W pierwszym przypadku polega to na utworzeniu charakterystyki treściowej dokumentu poprzez dołączenie do wykorzystanego symbolu jego odpowiednika słownego pochodzącego z tablic danej klasyfikacji (Rys. 4.3).

The screenshot shows a search result for 'CERN' in the ViFaPhys system. The record includes the following fields:

- Subject:** CERN
- Category:** www.cern.ch [HTML/Javascript]
- Description:** This main website of CERN, the European Organization for Nuclear Research, provides everybody interested in CERN research with information about CERN: general information, the CERN visitors programme, educational resources, as well as press and media information, recruitment and technology transfer issues. As part of the website, the CERN users' pages provide information for researchers using or intending to use CERN facilities. At CERN scientists study the building blocks of matter and the forces that hold them together. CERN provides scientists with the necessary tools: accelerators and detectors.
- Publisher:** European Organization for Nuclear Research
- Classification:** PACS 2003 10. The Physics of Elementary Particles and Fields  
PACS 2003 29. Experimental methods and instrumentation for elementary-particle and nuclear physics
- Category:** Academic Departments and Research Institutes
- Language:** English
- Access:** free
- Rights:**
- Date:** 2006-11-10

Rys. 4.3. Łączenie symboli klasyfikacyjnych z ich odpowiednikami słownymi w systemie ViFaPhys.

[Źródło: <http://vifaphys.tib.uni-hannover.de/>]

Przedstawiony na Rys. 4.3 rekord serwisu internetowego organizacji CERN pochodzi ze zbioru wyszukiwawczego systemu ViFaPhys. Do reprezentacji cech wymiaru treści zastosowano tutaj klasyfikację przedmiotową Physics and Astronomy Classification Scheme (PACS) w wydaniu z 2003 roku. W ChWD atrybut ten posiada identyfikator *Classification*, którego wartość jest wyrażana poprzez połączenie odpowiedniego symbolu z jego odpowiednikiem słownym zamieszczonym w tablicach. W tablicach PACS 2003 można odnaleźć:

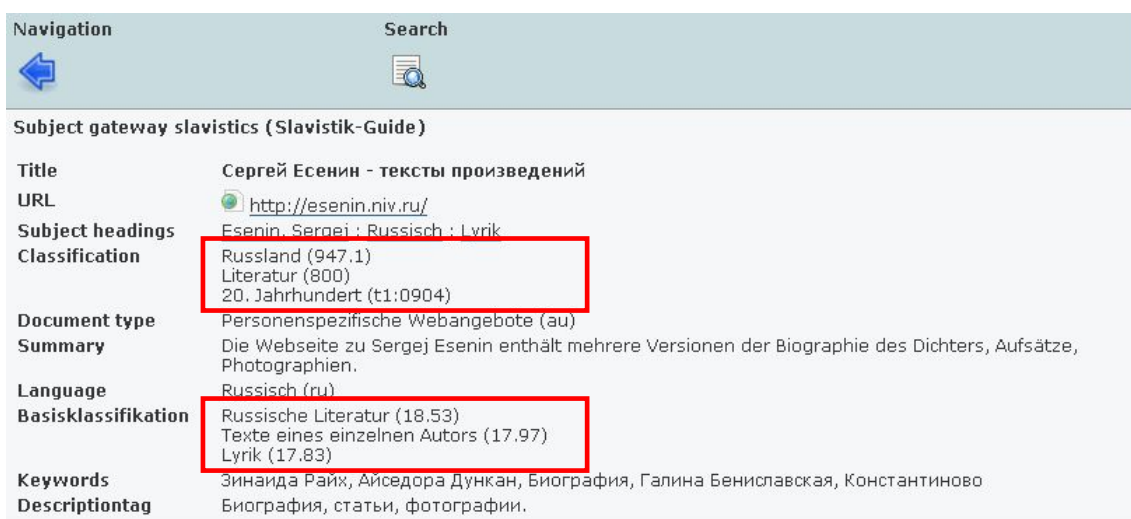
*10 - The Physics Of Elementary Particles and Fields*

oraz

29 - *Experimental methods and instrumentation for elementary-particle and nuclear physics*

Zachodzi tutaj pełna zgodność w formie przejmowanych symboli oraz wyrażen języka naturalnego stanowiącego charakterystykę słowną symboli.

Ten sposób adaptacji klasyfikacji na potrzeby indeksowania w DSH ma miejsce również w niemieckim systemie Slavistik-Portal (Rys. 4.4).



Navigation		Search	
Subject gateway slavistics (Slavistik-Guide)			
Title	Сергей Есенин - тексты произведений		
URL	<a href="http://esenin.niv.ru/">http://esenin.niv.ru/</a>		
Subject headings	Esenin, Sergej : Russisch : Lyrik		
Classification	Russland (947.1) Literatur (800) 20. Jahrhundert (t1:0904)		
Document type	Personenspezifische Webangebote (au)		
Summary	Die Webseite zu Sergej Esenin enthält mehrere Versionen der Biographie des Dichters, Aufsätze, Photographien.		
Language	Russisch (ru)		
Basisklassifikation	Russische Literatur (18.53) Texte eines einzelnen Autors (17.97) Lyrik (17.83)		
Keywords	Зинаида Райх, Айседора Дункан, Биография, Галина Бениславская, Константиново		
Descriptiontag	Биография, статьи, фотографии.		

Rys. 4.4. Technika łączenia symboli klasyfikacyjnych z odpowiednikami słownymi w systemie Slavistik-Portal. [Źródło: <http://www.slavistik-portal.de/>]

W systemie tym stosuje się Klasyfikację Dziesiątą Deweya oraz niemiecką klasyfikację uniwersalną Basisklassifikation. Tutaj również dochodzi do łączenia symbolu klasyfikacyjnego z jego odpowiednikiem słownym zamieszczonym w tablicach, w tym przypadku w wydaniu niemieckim. Elementem podstawowym, identyfikującym wartość dla cech wymiaru treści jest wyrażenie języka naturalnego, opracowane dla odpowiedniego symbolu klasyfikacyjnego (Rys. 4.4). Interesującym rozwiązaniem we wspomnianym systemie jest jednoczesne zastosowanie kilku JIW do indeksowania. W tym miejscu warto jednak podkreślić zastosowanie zasobu leksykalnego języka niemieckiego jako podstawy do tworzenia i przejmowania odpowiedników słownych symboli stosowanych klasyfikacji (etykieta pola: *Classification*) oraz haseł przedmiotowych (etykieta pola: *Subject Headings*), a także języka rosyjskiego do tworzenia niekontrolowanych słów kluczowych (etykieta pola: *Keywords*). Omawiany system udostępnia interfejs użytkownika w językach: niemieckim (domyślnie), angielskim i rosyjskim, ale to w języku niemieckim tworzona jest charakterystyka wyszukiwawcza dokumentów. Wykorzystanie języka rosyjskiego do tworzenia słów kluczowych ma charakter komplementarnej charakterystyki



treściowej i podyktowane jest zakresem systemu, który obejmuje szeroko pojmowaną sławistykę.

Nietypowym przykładem wykorzystania klasyfikacji do indeksowania dokumentów w DSH był system NMAP - Nursing, Midwifery and Allied Health\* z zakresu pielęgniarstwa, położnictwa i opieki społecznej. W systemie tym do odwzorowania treści elementów zbioru informacyjnego zastosowano National Library of Medicine Classification. Polegało to na indeksowaniu dokumentów, ale wyłącznie za pomocą odpowiedników słownych adekwatnych symboli klasyfikacyjnych. W systemie tym, po włączeniu do grupy tematycznej Intute: Health and Life Science zastąpiono tę klasyfikację językiem haseł przedmiotowych MeSH oraz tezauresem The Royal College of Nursing Thesaurus.

Zastosowanie w DSH języka informacyjno-wyszukiwawczego typu klasyfikacji do reprezentacji treści obiektów informacyjnych w DSH charakteryzuje się znacznymi uproszczeniami w wyznaczaniu ich zakresów tematycznych, które znakuje się określonymi symbolami klasyfikacyjnymi. Polega to na uogólnianiu tego typu reprezentacji poprzez rezygnację z kategorii symboli pomocniczych. Odstąpienie z odwzorowania odniesienia przestrzennego i chronologicznego przedmiotu oraz formy dokumentu w symbolu klasyfikacyjnym spowodowane jest obawą o nadmiarowość informacji, jaką niewątpliwie zawierałyby wówczas ChWD. Stosowane w DSH schematy metadanych już na poziomie formatu opisu identyfikują autonomiczność tego rodzaju atrybutów. W związku z tym, nie ma potrzeby ich odwzorowywania dodatkowo w symbolu klasyfikacyjnym. Ta uwaga odnosi się również do struktury haseł przedmiotowych i stosowania w nich grup określnikowych. Symbole klasyfikacyjne jako komunikaty języka informacyjnego o notacji sztucznej, mają małą wartość informacyjną dla użytkowników bez odpowiedniego przygotowania. W związku z tym, wykorzystanie ich charakterystyk słownych ma na celu bezpośrednio określenie ich zakresu znaczeniowego za pomocą języka naturalnego.

Interesującą i charakterystyczną cechą dla dużej liczby DSH jest polireprezentacja wiedzy. Jej celem jest konstruowanie zachodzących na siebie zakresowo komunikatów reprezentujących informacje o obiekcie, które zakładają różne perspektywy poznawcze. „Ma ona charakter intencjonalnej redundancji w reprezentacji dokumentów (oraz potrzeb informacyjnych) na różne, komplementarne sposoby”. [Sparck-Jones 1990 za Larsen, Ingwersen 2005 s. 45] W dokumentacyjnych systemach

---

\* Obecnie kolekcja systemu tworzy grupę tematyczną Intute: medycyna i nauki o życiu (<http://www.intute.ac.uk/healthandlifesciences/>)

informacyjno-wyszukiwawczych, do których zalicza się DSH (por. Rozdz. 3.4), przez polireprezentację wiedzy rozumie się taką „reprezentację wiedzy, w której wykorzystano więcej niż jedną metodę reprezentacji dokumentów, np. opis w postaci tekstów jednego lub równolegle stosowanych kilku języków informacyjno-wyszukiwawczych oraz opis w postaci tekstu języka naturalnego”. [Słownik encyklopedyczny 2002, s. 198]

W DSH zjawisko to polega na jednoczesnym zastosowaniu obok charakterystyki słownej, dodatkowo kilku języków informacyjno-wyszukiwawczych do odwzorowania treści obiektu informacyjnego. Tym samym odpowiednie elementy strukturalne ChWD zawierają teksty różnych JIW zastosowanych do reprezentacji tej samej cechy. Duża liczba systemów stosuje polireprezentację wiedzy, która ma swoje odzwierciedlenie w strukturze ChWD (Rys. 4.5)

BIBLIOGRAPHIC DATA	
<b>Title</b>	Abstract and Applied Analysis
<b>Editor</b>	[Editor in Chief] Kartsatos, Athanassios G. <hermes@tarski.math.usf.edu> (University of South Florida, US)
<b>Publisher</b>	Mancorp Publishing, Inc
<b>Distributor</b>	University of South Florida - Department of Mathematics: Tampa, US (FL) < <a href="http://www.math.usf.edu/">http://www.math.usf.edu/</a> >
<b>Language</b>	English
<b>Country (State)</b>	United States of America; United States of America (FL)
<b>Format of data</b>	text/html
→ <b>Keywords</b>	abstract analysis; applied analysis; functional analysis; nonlinear functional analysis; ordinary differential equations; partial differential equations; optimization theory; control theory
→ <b>Description</b>	Abstract and Applied Analysis (AAA) is a printed mathematical journal devoted exclusively to the publication of high quality papers in the fields of Abstract and Applied Analysis. Emphasis is placed on important developments in Linear and Nonlinear Functional Analysis, Ordinary and Partial Differential Equations, Optimization Theory and Control Theory. This site offers only scarce information about the printed journal and tables of contents of the 3 volumes published up to now. Additionally there's a list of links to related internet resources.
<b>URL</b>	<a href="http://www.hindawi.com/journals/aaa/">http://www.hindawi.com/journals/aaa/</a>
<b>ISSN/ISBN</b>	1085-3375 printed journal
CLASSIFICATION	
<b>Source Type</b>	Journals
→ <b>MSCverbal</b>	fourier analysis; abstract harmonic analysis; global analysis; analysis on manifolds; ordinary differential equations; partial differential equations; functional analysis; optimal control; optimization
→ <b>MSC</b>	42-XX; 43-XX; 58-XX; 34-XX; 35-XX; 46-XX; 49-XX

Rys. 4.5 Polireprezentacja wiedzy w systemie MathGuide. [Źródło: <http://mathguide.de>]

Przedstawiona na Rys. 4.5 charakterystyka wyszukiwawcza dla dokumentu internetowego pt. „Abstract and Applied Analysis” w systemie MathGuide (<http://www.mathguide.de>) zawiera pola:

- Keywords - słowa kluczowe niekontrolowane,
- Description - opis słowny,
- MSCverbal - odpowiednik słowny symbolu klasyfikacji Mathematics Subject Classification,
- MSC - symbol klasyfikacji Mathematics Subject Classification.

Są one odpowiedzialne za reprezentację wiedzy skodyfikowanej w obiekcie informacyjnym z użyciem odpowiednio:

- języka swobodnych słów kluczowych,
- charakterystyki słownej w języku naturalnym,
- odpowiednika słownego klasyfikacji Mathematics Subject Classification
- symbolu klasyfikacji Mathematics Subject Classification.

W DSH polireprezentacja wiedzy przyjmuje formę jawną oraz ukrytą. Polireprezentacja jawna zachodzi najczęściej z wykorzystaniem niekontrolowanego języka słów kluczowych, jednego lub kilku wybranych JIW oraz charakterystyki słownej, przy czym w strukturze ChWD są one zidentyfikowane za pomocą odpowiednich etykiet wskazujących na typ reprezentacji lub nazwę konkretnego JIW. Ma ona miejsce we wspomnianym systemie MathGuide, ale również w systemie Edna Online. Tutaj również odpowiednie etykiety pól w ChWD wskazują na sposób i narzędzie reprezentacji wiedzy. (Rys. 4.6)

### Search results

<b>IDENTIFIER</b>	http://www.webquest.org/questgarden/lessons/33820-060918183104/
<b>TITLE</b>	Bully Quest
<b>DESCRIPTION</b>	This bullying focussed Webquest is designed for students in Year One (age 5-6) and would also be suitable for years two and three. It addresses the social challenges of bullying and making friends. The initial task is to form small groups where students each adopt a specific role in researching the issue of bullying. Ultimately, they develop a story for a short video and act it out.
<b>KEYWORDS</b>	violence, making friends, leaving someone out, web quest, bully
<b>SUBJECT (ScOT)</b>	Bullying
<b>PUBLISHER</b>	QuestGarden
<b>PUBLISHER</b>	http://www.webquest.org/
<b>CREATOR</b>	Bruce, Glenn
<b>SECTOR</b>	preschool,school
<b>AUDIENCE</b>	student,teacher/lecturer
<b>USER LEVEL</b>	Early Childhood,1,2
<b>COVERAGE</b>	Australia
<b>RIGHTS</b>	Creative Commons
<b>TYPE</b>	text
<b>LANGUAGE</b>	en
<b>CATEGORY</b>	Early Childhood Learning Quests
<b>CATEGORY</b>	Cross Curriculum Learning Quests


Rys. 4.6 Polireprezentacja wiedzy w systemie EdnaOnline

Przedstawiona na Rys. 4.6 ChWD zawiera wartości dla cech wymiaru treści, które zostały odwzorowane odpowiednio:

- Description (opis) – charakterystyka słowna,
- Keywords (słowa kluczowe) – niekontrolowane słowa kluczowe,
- Subject (SCOT) (przedmiot) – tezaurus Schools Online Thesaurus,
- Audience (grupa użytkowników docelowych) – kontrolowana lista wyrażzeń,
- User level (poziom kształcenia) - kontrolowana lista wyrażzeń,
- Category (kategoria tematyczna w dostępie przedmiotowym) – odpowiednia kategoria tematyczna ze struktury narzędzia dostępu przedmiotowego.

Celem polireprezentacji wiedzy w DSH jest dążenie do maksymalizacji poziomu informatywności ChWD w kontekście wskazywania i nazywania przedmiotu/przedmiotów dokumentów internetowych oraz ich dodatkowych cech treści. Zakres zasobu leksykalnego danego JIW, ale również jego siła semantyczna do adekwatnej identyfikacji obiektów rzeczywistości dokumentacyjnej decyduje w tym przypadku o jego zastosowaniu w procesie indeksowania.

W systemach, gdzie do katalogowania zasobów WWW stosuje się zbiór słownictwa kontrolowanego, którego źródłem są słowniki kilku wybranych JIW, a taka sytuacja w odniesieniu do DSH ma miejsce często, zjawisko polireprezentacji ma formę ukrytą. Jeśli bazowy zbiór słownictwa stosowany do indeksowania jest derywowany ze słowników różnych JIW, to przynależność określonego wyrażenia do słownictwa konkretnego JIW nie jest identyfikowana odpowiednią etykietą w CHWD. Ma to miejsce np. w systemie Intute Social Sciences z zakresu nauk społecznych. Tutaj polireprezentacja wiedzy w ChWD ma formę ukrytą. (Rys. 4.7)

<b>Title</b>	<b>Descent of Man</b> <span style="background-color: yellow;">Editor's Choice</span>
<b>Description</b>	This is the full-text of Charles Darwin's book 'The Descent of Man', which contains a number of observations (related to Darwinian natural selection) that Darwin did not include in his original 'On The Origin of Species' (short title) in case he inflamed its already heated reception. These observations relate to the evolution of mankind. The book is hosted by the Online Literature Library, itself sponsored by Knowledge Matters Ltd (a UK-based company).
 <b>Keywords - controlled</b>	<a href="#">evolution</a>
<b>Keywords - uncontrolled</b>	<a href="#">human evolution</a> ; <a href="#">genetics</a>
<b>Type</b>	E-books
<b>URL</b>	<a href="http://www.literature.org/authors/darwin-charles/the-descent-of-man/">http://www.literature.org/authors/darwin-charles/the-descent-of-man/</a> [English]
<b>Classification</b>	<a href="#">Animal Experimental and Comparative Psychology</a>
<b>Country of origin</b>	United Kingdom
<b>Date added</b>	2004-01-07

Rys. 4.7 Ukryta polireprezentacja wiedzy w grupie tematycznej Social Science systemu Intute

W tej grupie tematycznej do indeksowania stosuje się kilka JIW. Są to:

- język słów kluczowych International Bibliography of the Social Sciences,

- język słów kluczowych Social Care Institute for Excellence Vocabulary, Statistics and Research Methods Vocabulary,
- tezaurus CAB,
- tezaurus HASSET,
- tezaurus Research Methods & Tools,
- tezaurus Social Care Online.

Prezentacja zawartości ChWD uwzględnia wyłącznie etykietę pola *keywords controlled* (kontrolowane słowa kluczowe) jako wskaźnik zastosowania heterogenicznego zbioru słownictwa kontrolowanego, z którego pochodzą określone wyrażenia. Zatem charakterystyczną cechą tego rodzaju ukrytej polireprezentacji jest brak elementów identyfikujących wykorzystanie konkretnych JIW.

Wiele DSH praktykuje rozwiązanie, w którym elementy ChWD reprezentujące treść dokumentów internetowych mają wysoką wartość informacyjną, lecz pod względem wymogów strukturalnych narzucanych przez gramatyki poszczególnych JIW są mocno zredukowane. Autor nie odnotował zależności w sposobach i zasadach łączenia zasobów leksykalnych JIW różnych typów w ukrytej polireprezentacji wiedzy w DSH.

#### **4.2.7. Dodatkowe cechy**

Do dodatkowych cech obiektów informacyjnych, identyfikowanych podczas indeksowania a niewymienionych dotychczas, należą przede wszystkim atrybuty wskazujące na ich potencjalne wartości edukacyjne. DSH są systemami informacyjno-wyszukiwawczymi pełniącymi funkcje źródeł informacji, przede wszystkim dla użytkowników z sektora nauki i szkolnictwa na różnych poziomach. Funkcja ta przekłada się na perspektywę analizy i reprezentacji elementów zbioru informacyjnego i tym samym na przyjęty schemat opisu danych. Z przedstawionej w trzecim rozdziale rozprawy analizy schematów metadanych stosowanych w DSH wynikało, że do dodatkowych atrybutów wyodrębnianych w formacie opisu należą najczęściej te dotyczące wartości edukacyjnej obiektu informacyjnego. Są to np. przeznaczenie obiektu informacyjnego dla kształcenia na określonym poziomie, przeznaczenie dla określonej grupy użytkowników, czy też stopień trudności prezentowanych treści. Tego typu atrybuty odwzorowuje się najczęściej za pomocą opracowanych na potrzeby systemu list słownictwa kontrolowanego. Źródłem słownictwa w tym przypadku są np. normy lub słowniki terminologiczne. W systemie GEM: Gateway to 21st Century Skills

z zakresu szeroko rozumianej edukacji stosuje się również dodatkowy atrybut - *metody nauczania*, do odwzorowania którego wykorzystuje się listę słownictwa kontrolowanego sporządzoną na podstawie słowników terminologicznych z tego zakresu oraz odpowiednich deskryptorów tezaurya Thesaurus of ERIC Descriptors.

Udział języka informacyjno-wyszukiwawczego w reprezentacji wiedzy w DSH to przede wszystkim priorytetowe traktowanie jego zasobu leksykalnego do identyfikacji tematów dokumentów z jednoczesnym pomniejszeniem jego roli w odwzorowaniu relacji semantycznych pomiędzy obiektami reprezentowanej rzeczywistości dokumentacyjnej. Mamy tutaj do czynienia z redukcją tekstów JIW do postaci równoważnych wypowiedzeń złożonych z elementów autosyntaktycznych i/lub elementów autosyntaktycznych z ograniczoną liczbą grup wyrażenń synsyntaktycznych. Tego rodzaju zabieg można interpretować jako dostosowanie struktury charakterystyki treściowej do postkoordynacyjnej metody realizacji funkcji wyszukiwawczych w systemach.\* Rzeczywiście tego typu redukcja gramatyczna jest potrzebna w postkoordynacji, ale ta w DSH nie do końca pełni swoją funkcję. Jej realizacja w mechanizmach wyszukiwawczych polega na wyszukiwaniu wystąpień wyrażenń, ale domyślnie w całej ChWD, nie zaś w polach do tego przeznaczonych. Szerzej zostanie to scharakteryzowane w kolejnym podrozdziale.

Warunek definicyjny DSH stanowi, że proces indeksowania elementów zbioru informacyjnego przeprowadzany jest manualnie przez specjalistów dziedzinowych lub katalogerów do tego upoważnionych. W tego typu systemach stosuje się również mechanizmy automatycznego indeksowania. Omówienie tego zagadnienia wykracza jednak poza ramy pracy, ale z pewnością jest interesujące z badawczego punktu widzenia. W tym miejscu warto jednak odnotować fakt, że w systemach gdzie wykorzystuje się mechanizmy automatycznej charakterystyki treściowej dokumentów, tego typu opisy pełnią jedynie funkcję uzupełniającą ChWD, a odpowiednio wygenerowane symbole klasyfikacyjne czy wyrażenia języka naturalnego są graficznie wyróżniane jako pochodzące z automatycznego nadania.

---

\* Postkoordynacja jest to „cecha systemu informacyjno-wyszukiwawczego polegająca na tym, że każde wyrażenie charakterystyki wyszukiwawczej dokumentu samodzielnie może pełnić funkcję wyszukiwawczą, umożliwiając bezpośredni dostęp do zbioru wyszukiwawczego”. [Słownik encyklopedyczny 2002, s. 200]

### ***4.3. Identyfikacja cech wyszukiwawczych***

Formaty opisów danych, jakie stosuje się w dziedzinowych systemach hipertekstowych wyznaczają zbiór cech identyfikujących elementarną jednostkę opisu. Przedstawiona w rozdziale trzecim charakterystyka zestawów metadanych, stosowanych w wybranych dziedzinowych systemach hipertekstowych, ujawniła powtarzający się zestaw atrybutów, najczęściej w postaci elementów obowiązkowych schematu Dublin Core. Jest on często uzupełniany o dodatkowe atrybuty, które uznano za istotne dla realizacji celów wyszukiwawczych. Dodatkowe elementy odwzorowują zarówno cechy:

- formalne (np. dostępność obiektu, umiejscowienie przedmiotu w przestrzeni geograficznej i czasie),
- jakościowe obiektu (np. poziom trudności informacji, stopień ich szczegółowości),
- treściowe (np. symbol klasyfikacji specjalistycznej, nazwa kategorii tematycznej, do której przyporządkowano charakterystykę obiektu w narzędziu dostępu przedmiotowego),
- przeznaczenie (np. w postaci identyfikacji grupy użytkowników docelowych, poziomu kształcenia, wymaganych kwalifikacji).

Format opisu danych wyznacza semantyczną strukturę jednostki opisu identyfikując jej atrybuty. Jednocześnie stanowi on o potencjale wyszukiwawczym systemu poprzez możliwość nadania wybranemu atrybutowi statusu cechy wyszukiwawczej. Te, z kolei wyznaczają zestaw kryteriów wyszukiwawczych w mechanizmach wyszukiwania informacji oraz są podstawą do budowy narzędzi dostępu przedmiotowego.

Dziedzinowe systemy hipertekstowe charakteryzują się stałym zestawem punktów dostępu identyfikowanych w dwóch rodzajach narzędzi dostępu do zasobów: wyszukiwarkach i narzędziach dostępu przedmiotowego. Rozróżnienie na punkty dostępu aktywowane poprzez wyszukiwarki oraz narzędzia dostępu przedmiotowego (ang. searchable / browsable) spowodowane jest odmiennymi technikami ich wykorzystania w procesie pozyskiwania informacji.

Do stałych punktów dostępu w dziedzinowych systemach hipertekstowych należą:

I. w wyszukiwarkach:

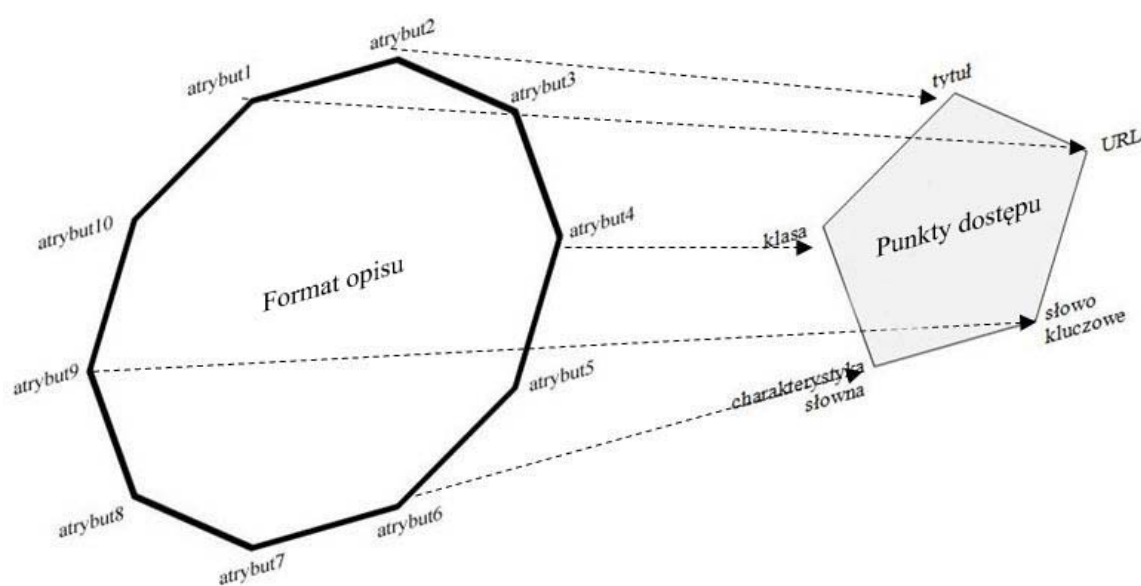
- tytuł dokumentu internetowego,
- adres URL dokumentu internetowego,

- słowo kluczowe,
- słowo w całej charakterystyce wyszukiwawczej.

II w narzędziach dostępu przedmiotowego:

- kategoria tematyczna, reprezentująca wymiar treści i/lub formy dokumentu internetowego.

Tym samym na podstawie cech strukturalnych elementarnej jednostki opisu równoznacznej ze schematem metadanych, dokonuje się identyfikacji tych atrybutów, które będą pełniły funkcję punktów dostępu (Rys. 4.8).



Rys. 4.8 Identyfikacja cech wyszukiwawczych w dziedzinowych systemach hipertekstowych.

[Źródło: opracowanie własne]

Przedstawione na Rys. 4.8 atrybuty obiektów informacyjnych, które pełnią funkcję cech wyszukiwawczych w DSH są typowymi, lecz nie jedynymi punktami dostępu dla tej klasy systemów. Zestaw punktów dostępu jest różny w zależności od konkretnego systemu. Liczba zidentyfikowanych punktów dostępu w danym systemie określa jego możliwości wyszukiwawcze poprzez wielowymiarową charakterystykę jednostki zbioru wyszukiwawczego. Budowa dwóch rodzajów narzędzi dostępu (wyszukiwarka i narzędzia dostępu przedmiotowego) determinuje sposoby włączania w nie określonych punktów dostępu.

W przypadku wyszukiwarek, włączenie zestawu punktów dostępu powoduje możliwość zastosowania określonych kryteriów wyszukiwawczych. Ma to swoje przełożenie na budowę interfejsu mechanizmu wyszukiwawczego. Kategorie tematyczne tworzące narzędzie dostępu przedmiotowego pełnią funkcję przedmiotowych punktów dostępu poprzez grupowanie ChWD według cech ich treści,



oraz formy. W Tab. 4.3 przedstawiono rodzaje punktów dostępu w przykładowych DSH.

Tab. 4.3 Punkty dostępu w wybranych DSH. [Źródło: opracowanie własne]

Nazwa systemu	Mechanizm wyszukiwawczy	Narzędzia dostępu przedmiotowego
Intute ( <a href="http://www.intute.ac.uk/">www.intute.ac.uk/</a> )	tytuł słowo kluczowe charakterystyka słowna	przedmiot rodzaj zasobów
Mathguide ( <a href="http://www.mathguide.de">www.mathguide.de</a> )	tytuł URL dowolny tekst w opisie słowo kluczowe typ dokumentu symbol klasyfikacji	przedmiot rodzaj zasobów
History Guide ( <a href="http://www.historyguide.de">www.historyguide.de</a> )	tytuł słowo kluczowe URL symbol klasyfikacji	region okres historyczny przedmiot
Aerade ( <a href="http://aerade.cranfield.ac.uk">http://aerade.cranfield.ac.uk</a> )	tytuł opis dowolny słowo w całym rekordzie	przedmiot
Infomine ( <a href="http://infomine.ucr.edu/">http://infomine.ucr.edu/</a> )	autor tytuł słowo kluczowe przedmiot opis pełny tekst pole systemowe	przedmiot słowo kluczowe autor tytuł rodzaj zasobów

Rola JIW w procesie identyfikacji punktów dostępu wynika z możliwości języka opisu bibliograficznego do identyfikacji atrybutów obiektów i nadania wybranym statusu cechy wyszukiwawczej. O ile o wyborze konkretnych atrybutów do pełnienia funkcji cech wyszukiwawczych decydują względy praktyczne i kierowanie się efektywnością procesów wyszukiwawczych, to w przypadku reprezentacji wartości punktów dostępu rola JIW ogranicza się do jego podstawowej funkcji – metainformacyjnej.

#### **4.4. Organizacja punktów dostępu**

Organizacja punktów dostępu do zasobów informacyjnych DSH polega na dystrybucji wejść do systemu za pomocą dostępnych narzędzi realizujących funkcje wyszukiwawcze. W DSH mamy do czynienia z dwoma rodzajami takich narzędzi. Są to wyszukiwarki oraz narzędzia dostępu przedmiotowego. Pomimo wyraźnych

rozbieżności w sposobach ich funkcjonowania należy je traktować jako komplementarne mechanizmy dostępu do kolekcji DSH. W tym miejscu zostaną one scharakteryzowane w kontekście realizacji funkcji wyszukiwawczych oraz rozmieszczenia i aktywizowania punktów dostępu.

#### **4.4.1. Wyszukiwarki**

Zastosowanie mechanizmów wyszukiwawczych w postaci wyszukiwarek w DSH, polega na wykorzystaniu aplikacji przeszukujących zawartość pól i podpól rekordów stanowiących elementy zbioru wyszukiwawczego. Zadaniem użytkownika jest zbudowanie zapytania informacyjnego za pomocą narzędzi udostępnionych przez system, czyli w tym przypadku formularza wyszukiwarki wraz z jego funkcjami zawężania i poszerzania zakresu zapytania informacyjnego. Rezultaty wyszukiwania, czyli listę rekordów spełniających kryteria relewancji systemowej, można poddawać modyfikacjom polegającym na wyborze przez użytkownika sposobu ich wyświetlania. Większość zbadanych przez autora systemów oferuje dwa tryby realizacji sesji wyszukiwawczych, tzw. tryb prosty (ang. simple search) oraz zaawansowany (ang. advanced search, option search).

Interfejs użytkownika modułu wyszukiwania informacji w tym trybie składa się formularza do wprowadzania zapytania informacyjnego oraz dodatkowych narzędzi służących do uszczegółowienia zapytania. Rys. 4.9 przedstawia wygląd typowego modułu wyszukiwawczego w trybie prostym.

**AERADE services**

Search and browse:

- [Aerospace & defence resources](#)
- [DEVISE](#)
- [ESDU Series](#)
- [Reports Archive](#)
- [Shrivenham Index](#)

Inform:

- [ConferenceBrief](#)
- [JobsBrief UK](#)
- [NewsBrief](#)

**AERADE news**

Keep up to date with the latest developments and news in your field. [More...](#)

**AERADE friends**

- [Books Express](#)
- [Intute](#)
- [IEE Aerospace Network](#)
- [Internet for Aeronautical Engineering tutorial](#)

Welcome to DEVISE

DEVISE is a special collection of quality military and defence sources focussing directly on the information needs of the armed forces. It forms a tailored subset of the main aerospace and defence resources database.

**DEVISE** DEFENCE VIRTUAL INFORMATION SERVICE

Quick Search | [Full search options](#)

Enter search terms:

To search for DEVISE resources, enter one or more terms. You can use the expressions AND, OR and NOT to specify your search more precisely.

**Browse DEVISE**

- [Alphabetically](#)
- [By subject heading](#)

**AERADE Survey 2007**

We are keen that AERADE reflects your evolving interests and needs, and welcome your feedback to help us to continue to improve the service we offer you. We would be grateful if you could take five minutes to [answer a few questions](#) for us about your current use of AERADE and your views on it.

Rys. 4.9 Tryb prosty modułu wyszukiwawczego w systemie Aerade. [Źródło: [http://aerade.cranfield.ac.uk/devise\\_index.html](http://aerade.cranfield.ac.uk/devise_index.html)]

Narzędzie to oprócz możliwości wprowadzenia zapytania informacyjnego nie udostępnia użytkownikowi żadnej dodatkowej opcji jego uszczegółowienia. Efektywność wyszukiwania w tym przypadku zależy zarówno od poprawnego sformułowania zapytania informacyjnego oraz posiadanej przez użytkownika wiedzy o systemie. W modułach wyszukiwania informacji w DSH, do formułowania zapytania informacyjnego wykorzystuje się język swobodnych słów kluczowych. Niektóre dziedzinowe systemy hipertekstowe, w trybie prostym wyszukiwania informacji oferują więcej opcji. Rys. 4.10 przedstawia interfejs użytkownika modułu wyszukiwania informacji w systemie MathGuide.

**MATHGUIDE**  
SEARCH ENGINE

---

**Category:**

**Display:**

**Word(s):**

**Sort:**

We are experimenting with the incorporation of the **MathNet Links** into the MathGuide. They will be enhanced co-operation with the Centrum voor Wiskunde en Informatica (CWI) in the Netherlands. Since they don't have a MathGuide record yet, they need a custom display form, which is not yet ready. But you can search in the

Search in MathGuide  Search in CWI data  Search in MathNetLinks

There is also an advanced search available.

Rys. 4.10 Tryb prosty modułu wyszukiwawczego w systemie MathGuide [Źródło: [www.mathguide.de](http://www.mathguide.de)]

W tym przypadku użytkownik ma możliwość uszczegółowienia zapytania oraz modyfikacji kryteriów wyświetlania rezultatów wyszukiwania podczas interakcji z formularzem wyszukiwawczym. Interfejs mechanizmu wyszukiwawczego tego systemu umożliwia:

- sformułowanie zapytania do postaci ciągu niekontrolowanych słów kluczowych – pole *Word(s)*;
- określenie (wybór z listy) pola lub podpola w strukturze rekordu bazy danych, które mają być przeszukiwane – pole *Category*.

W przypadku drugiej opcji pole wyboru o nazwie *Category* zawiera następujące wartości:

- *Text* – wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w całej strukturze rekordu.
- *Title* - wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w polu *Tytuł*.
- *URL* - wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w polu *adres URL*.
- *Keywords* - wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w polu *Słowa kluczowe*.
- *Source Type* - wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w polu *Rodzaj źródła*.
- *MSC* - wyszukiwanie wystąpień ciągu słów kluczowych z pola *Word(s)* w polu przeznaczonym dla symbolu klasyfikacji *Mathematic Subject Classification*.

Użytkownicy korzystający z modułu wyszukiwania informacji w DSH mogą opcjonalnie skorzystać również z trybu zaawansowanego, który oferuje znacznie więcej funkcji i umożliwia konstruowanie złożonych zapytań. Większa liczba funkcji, w stosunku do trybu prostego, jest odzwierciedlona w złożonej budowie interfejsu użytkownika. Na rys. 4.11 przedstawiono interfejs użytkownika w trybie zaawansowanym w tym samym systemie MathGuide.

Advanced Search

Category:  Word(s):

AND  Category:  Word(s):

AND  Category:  Word(s):

AND  Source Type:

Display:  Sort:

Rys. 4.11 Tryb zaawansowany modułu wyszukiwawczego w systemie MathGuide  
[Źródło: www.mathguide.de]

Budowa tego interfejsu została oparta na logicznej rozbudowie interfejsu dla trybu prostego. W tym przypadku z wielokrotności pól *Word(s)* oraz *Category*, które zostały powiązane za pomocą operatorów algebry Boole'a (AND, OR, NOT). W trybie zaawansowanym pole *Category* zawiera dziesięć wartości, nie zaś sześć jak to było w przypadku trybu prostego. Dodatkowe pola to:

- *Language* – wyszukiwanie wystąpień ciągu słów kluczowych w postaci dwuznakowych kodów nazw języków (na podstawie normy ISO 639-1 [Codes for representation 2002]) w polu przeznaczonym do określania języka dokumentu.
- *Country* – wyszukiwanie wystąpień ciągu słów kluczowych w postaci dwuznakowych kodów nazw państw (na podstawie normy ISO 3166 [Codes for the representation of names of countries and their subdivisions 1999]) w polu przeznaczonym do określenia zasięgu geograficznego dokumentu lub związanego z powstaniem dokumentu lub jego treścią.
- *Last update* - wyszukiwanie wystąpień ciągu słów kluczowych w formie daty dziennej (rrrr-mm-dd) lub jej elementów w polu przeznaczonym do określania daty ostatniej aktualizacji dokumentu.
- *Timestamp* – wyszukiwanie wystąpień ciągu słów kluczowych w formie daty dziennej w polu przeznaczonym do określania daty dodania do zbioru wyszukiwawczego.

Kolejnym sposobem uszczegóławiania zapytania informacyjnego jest możliwość wyboru kryterium rodzaju obiektu informacyjnego. Zawężanie zapytania z wykorzystaniem tej funkcji polega na wyborze z listy wyłącznie jednej wartości

odpowiadającej formie obiektu informacyjnego. W systemie MathGuide użytkownik może wybrać jedną z czterdziestu ośmiu kategorii wyodrębnionych dla cechy *typ zasobów*. Są to m.in.:

- serwisy informacyjne bibliotek,
- serwisy informacyjne muzeów,
- serwisy informacyjne instytucji badawczych,
- słowniki, encyklopedie online,
- normy, standardy,
- biblioteki wirtualne,
- bibliografie dostępne w Internecie,
- czasopisma elektroniczne,
- oprogramowanie, i inne.

W omawianym systemie MathGuide można zauważyć pewną niekonsekwencję w wykorzystaniu tej funkcji. Użytkownik może bowiem przeprowadzić wyszukiwanie informacji według typu dokumentów wykorzystując zamkniętą listę terminów określających rodzaj indeksowanych przez system źródeł elektronicznych, ale również samodzielnie wpisać tę wartość w formularzu wyszukiwawczym wybierając z pola *Category* cechę *Typ zasobów* (resource type). Druga możliwość narażona jest na ryzyko popełnienia przez użytkownika błędu w formie wyrażenia (tutaj skrótu) użytego przez indeksatorów do odwzorowania tej cechy.

Kolejną funkcją, jaką oferują zaawansowane moduły wyszukiwania informacji to możliwość zawężania zapytania według cech treści obiektów informacyjnych. Polega to na umożliwieniu użytkownikowi wyboru z listy wyrażenia określającego zakres przedmiotowy i przeprowadzenia wyszukiwania informacji wśród ChWD przypisanych do danej kategorii tematycznej. Wyrażenia te charakteryzują się dużym stopniem ogólności, a suma ich zakresów równa jest zakresowi danego systemu. Są one zazwyczaj projekcją najwyższego poziomu schematów kategoryalnych zastosowanych w narzędziu dostępu przedmiotowego. Rys. 4.12 przedstawia sposób realizacji tej funkcji w systemie *ForestryGuide* z zakresu leśnictwa.

**Advanced Search**

**Category:** Text  **Word(s):**   
 **Category:** Keywords  **Word(s):**   
 **Subject:**    
 **Source Type:**    
**Display:**

Any

Forestry, Forest Products and Agroforestry (General)

Forests and Forest Trees (Biology and Ecology)

Silviculture and Forest Management

Forest Mensuration and Management

Forest Fires

Protection Forestry

Ornamental and Amenity Trees

Forest Products and Industries (General)

Wood Properties, Damage and Preservation

Logging and Wood Processing

Wood Utilization and Engineered Wood Products

Chemical and Biological Processing of Wood

Non-Wood Forestry Products

Agroforestry, Farm and Social Forestry

There is also a sin

:

This service can only be used from  
 To search in the categories Subjed  
 Except for Keywords you can use  
 You can also use **question marks**  
 words.

enter the respect  
 search with **and**  
 question marks at


Rys. 4.12 Zawężanie zapytania informacyjnego w systemie ForestryGuide  
 [Źródło: <http://www.forestryguide.de>]

Wyszukiwanie wspomagane słownikiem wykorzystanego w systemie języka informacyjno-wyszukiwawczego jest stosunkowo rzadko stosowanym rozwiązaniem w omawianym sposobie dostępu do zasobów. Najczęściej do określania wartości pewnych cech dokumentów wykorzystuje się zamknięte listy terminów. W interfejsach wyszukiwania informacji stosuje się zatem np.:

- dla cechy *język dokumentu* odesłanie do listy skrótów określających nazwę języka (najczęściej na podstawie normy ISO 639-1),
- dla cechy *nazwa państwa* odesłanie do listy skrótów określających nazwę państw (najczęściej na podstawie normy ISO 3166),
- dla cechy *rodzaj zasobów* zamkniętą listę wyrażen określających formę indeksowanych przez system obiektów informacyjnych.

Sprzężenie słownika stosowanego w systemie JIW z mechanizmem wyszukiwawczym polega na umożliwieniu użytkownikowi pobierania wyrażen z takiego zasobu do budowy zapytania informacyjnego. Takie rozwiązanie wprowadzono np. w systemie Intute: Social Sciences (Rys. 4.13).

[Intute](#) > [Social Sciences](#) > Thesaurus engine

<b>Subject links</b>	<b>Thesaurus engine</b>
<a href="#">About us</a>	Use the thesaurus engine to provide alternative terms that will generate hits within the Intute Catalogue. Enter a few letters and the engine will retrieve all matching terms which contain the text. e.g. search for econ to retrieve economics, economic policy, etc.
<a href="#">A-Z of Services</a>	
<b>Internet catalogue</b>	<input type="text"/> 
<a href="#">Search</a>	<input type="radio"/> General social science (HASSET) <input type="radio"/> Government, politics and anthropology (IBSS) <input type="radio"/> Social work and welfare (SCIE) <input type="radio"/> Research methods & tools (SRM) <input type="radio"/> CAB thesaurus (CABI)
<a href="#">Browse</a>	
<a href="#">New resources</a>	
<a href="#">Suggest a site</a>	
<b>Internet training</b>	
<a href="#">Virtual Training Suite</a>	
<a href="#">Support materials</a>	
<a href="#">Intute events</a>	
<b>Additional services</b>	
<a href="#">Conferences &amp; events</a>	
<a href="#">Departments</a>	
<a href="#">Subject news</a>	

Rys. 4.13 Wyszukiwanie informacji wspomagane tezaurusami w Intute: Social Sciences  
 [Źródło: <http://www.intute.ac.uk/socialsciences/thesaurus.html>]

Wyszukiwanie wspomagane słownictwem kontrolowanym umieszczono tutaj jako opcję zaawansowanego trybu wyszukiwania informacji. Użytkownik przez wysłaniem zapytania do systemu, odnajduje za pomocą modułu zaimplementowanego tezausa formy jednostek leksykalnych reprezentujących jego zapytanie informacyjne. W tym przypadku, przed wpisaniem poszukiwanego wyrażenia należy również określić jeden z pięciu zakresów tematycznych, którym odpowiadają następujące zbiory słownictwa kontrolowanego:

- Nauki społeczne – ogólnie (General Social Science) – tezaurus HASSET (Humanities And Social Sciences Electronic Thesaurus),
- Administracja, polityka i antropologia (Government, Politics and Anthropology) – słownik derywowany z listy deskryptorów tematycznych pochodzących z International Bibliography of the Social Sciences (IBSS),
- Opieka społeczna (Social Work and Welfare) – zbiór leksyki opracowany przez Social Care Institute for Excellence (SCIE) na podstawie zasobów Social Care Online (<http://www.scie-socialcareonline.org.uk/>),
- Metodologia badań społecznych – Social Research Methodology Thesaurus,
- CAB Thesaurus.

Po wysłaniu zapytania użytkownik otrzymuje artykuł hasłowy dla wyszukiwanego terminu (Rys. 4.14).



**Subject links**

About us

A-Z of Services

**Internet catalogue**

Search

Browse

New resources

Suggest a site

**Internet training**

Virtual Training Suite

Support materials

Intute events

**Additional services**

Conferences & events

Departments

Subject news

Timelines

### Search Results

Use the thesaurus engine to provide alternative terms that will generate hits within the Intute Catalogue. Enter a few letters and the engine will retrieve all matching terms which contain the text.

public international law

- General social science (HASSET)
- General social science (HASSET)
- Government, politics and anthropology (IBSS)
- Social work and welfare (SCIE)
- Research methods & tools (SRM)
- CAB thesaurus (CABI)

Suggested matches for **public international law** from **hasset**. Click on a term to view broader/narrower terms and any related terms available. If you wish to choose a term to search the Intute catalogue, click the checkbox and press "Search Intute".

You may select more than 1 term and the search will retrieve records containing one or more of the selected terms.

Current term		
<input type="checkbox"/> <a href="#">public international law</a> [16]		
Broader terms	Narrower terms	Related terms
<input type="checkbox"/> <a href="#">international law</a> [507]	<input type="checkbox"/> <a href="#">international agreements</a> [95]	<input type="checkbox"/> <a href="#">human rights</a> [1591]
	<input type="checkbox"/> <a href="#">international disputes</a> [4]	<input type="checkbox"/> <a href="#">humanitarian law</a> [63]
	<input type="checkbox"/> <a href="#">rights of states</a> [0]	

Rys. 4.14 Artykuł hasłowy w tezausie HASSET [Źródło:www.intute.ac.uk]

Przykładowy artykuł hasłowy dla wyrażenia *public international law* (międzynarodowe prawo publiczne) zbudowany jest z:

- wyrazu hasłowego, czyli jednostki leksykalnej stanowiącej podstawę do jego sporządzenia – *public international law* (międzynarodowe prawo publiczne),
- jednostek leksykalnych powiązanych z wyrażeniem wyróżnionymi graficznie relacjami paradygmatycznymi,
  - o relacją nadrzędności zakresowej – wykładnik *broader terms*,
  - o relacją podrzędności zakresowej – wykładnik *narrower terms*,
  - o relacją asocjacyjną – wykładnik *related terms*.

Budowa zapytania informacyjnego z wykorzystaniem zaimplementowanego tezausa polega w tej sytuacji na wyborze wyrażenia lub wyrażen poprzez ich zaznaczenie w odpowiednim polu w artykule hasłowym oraz wysłaniu zapytania do systemu. Dodatkową opcją systemu Intute jest możliwość wyboru z tezausa wyrażenia stanowiącego podstawę do wyszukiwania informacji wraz ze wszystkimi terminami powiązanymi relacją podrzędności lub nadrzędności zakresowej z wybranym wyrażeniem.

Omówione przykłady interfejsów użytkownika dla modułów wyszukiwania informacji w trybie prostym i zaawansowanym są typowymi rozwiązaniami dla

dziedzinowych systemów hipertekstowych. Budowa interfejsów użytkownika dla wskazanych trybów zakłada:

- możliwość wysłania przez użytkownika zapytania w postaci ciągu niekontrolowanych słów kluczowych lub pobrania jednostek leksykalnych, w takiej formie, w jakiej występują w słowniku stosowanego języka informacyjno-wyszukiwawczego,
- uszczegółowienie instrukcji poprzez:
  - o wybór pola, które będzie przeszukiwane,
  - o określenie rodzaju źródła, czyli typu obiektu informacyjnego,
  - o wyodrębnienie podzbioru przeszukiwanych obiektów na podstawie ich zakresu przedmiotowego,
- wpływ użytkownika na wizualizację rezultatów wyszukiwania.

Podczas wyszukiwania informacji z wykorzystaniem jednostek leksykalnych pobieranych z tezaury, realizacja tego procesu jest mało efektywna. Spowodowane jest to faktem, że mechanizm wyszukiwawczy wyszukuje jednostki leksykalne równokształtne z pobranymi np. z tezaury, ale w całej strukturze rekordu, nie zaś w polach przeznaczonych do odwzorowania treści dokumentów. Rys. 4.15. przedstawia rezultaty wyszukiwania w systemie SOSIG dla pobranego z tezaury HASSET wyrażenia *global warming* (globalne ocieplenie).

The screenshot shows a search interface with the following elements:

- Search Results** header.
- Breadcrumb: You are here : [Home](#) > Search Results
- Search input field containing "global AND warming" and a "go" button.
- Text: Use the **Thesaurus** from [Advanced Search](#) to find more specific terms
- Links: [Advanced Search](#) and [Thesauri](#)
- Results for: global AND warming**
- Match count: 1 - 10 from 18 matches
- Navigation: 1 2 [Next](#)
- Up in Smoke? Threats from, and Responses to, the Impact of Global Warming on Human Development**
- Description:** This 2004 report was published by the New Economics Foundation. Its message is that [global warming](#) could prevent us reaching the Millennium Development Goals. It is a call for urgent action by governments. Measures like cuts in emission of greenhouse gases by industrialised countries are needed. The report contains a range of case studies, examining subjects like why women suffer most from [global warming](#), trade and climate, the movement of people and learning from history. [+ Full Record](#)
- Keywords:** targets, pollution, government policy, [global warming](#), environmental planning policy, environment, economic and social development
- Subject Section(s):** [Environmental Economics](#) [Development Economics](#) [Environmental Impact of Business](#) [Pollution Control](#) [Global Warming](#)
- Resource Type:** Papers/Reports/Articles (individual)
- [http://www.neweconomics.org/gen/z\\_sys\\_PublicationDetail.aspx?pid=196](http://www.neweconomics.org/gen/z_sys_PublicationDetail.aspx?pid=196)

Rys. 4.15 Rezultaty wyszukiwania z wykorzystaniem tezaury [Źródło:www.sosig.ac.uk].

Wyróżnione przez system elementy charakterystyki wyszukiwawczej są równokształtne z wyrażeniami tworzącymi zapytanie informacyjne. Zatem na podstawie kryteriów relewancji systemowej dokument ten został wybrany jako odpowiedź systemu na tak postawione zapytanie informacyjne. Należy zauważyć, że ciąg znaków w postaci wyrażenia *global warming* pojawia się tutaj w kilku polach rekordu: w tytule, charakterystyce słownej dokumentu, w polu przeznaczonym dla niekontrolowanych słów kluczowych (*Keywords*) oraz w polu kontrolowanych jednostek leksykalnych tezaurya HASSET (*Subject Section(s)*). Mechanizm wyszukiwawczy systemu rozbił wyrażenie *global warming* wybrane z tezaurya przez użytkownika na dwa terminy *global* + *warming* i zastosował koniunkcję jako działanie na zbiorze. Warto zaznaczyć, że stopień koordynacji w tezaurysie HASSET dla wyrażenia *global warming* zakłada, że jest to termin złożony i w taki sposób powinien być traktowany. Można zatem uznać, że taka adaptacja tezaurya na potrzeby modułu wyszukiwania informacji w systemie sprawia, że funkcja wyszukiwawcza tego narzędzia jest realizowana nieefektywnie. Jest to przykład na wspomnianą w tym rozdziale „falszywą” postkoordynację. Polega ona z jednej strony na rozdzielaniu elementów składowych terminu wielowyrazowego, a z drugiej na odszukiwaniu wystąpień terminów w całej strukturze ChWD nie zaś w polach przeznaczonych do reprezentacji tego rodzaju atrybutów.

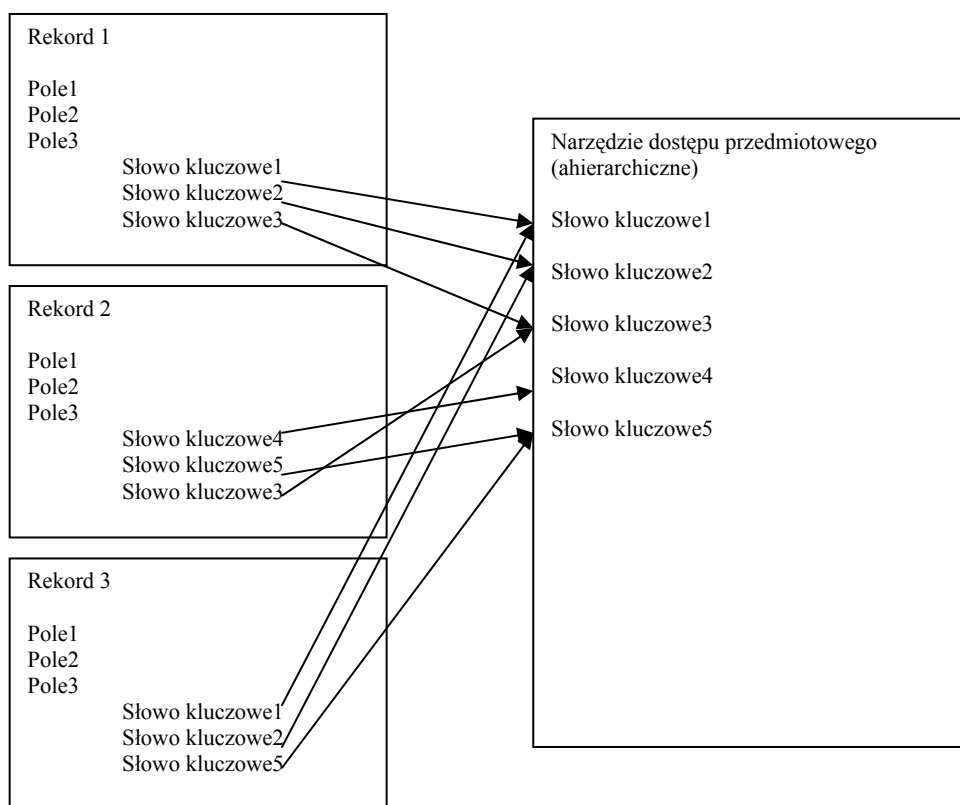
#### **4.4.2. Narzędzia dostępu przedmiotowego**

Dostęp przedmiotowy w dziedzinowych systemach hipertekstowych realizowany jest poprzez odpowiednie narzędzia wykorzystujące technikę przeglądania jako podstawowy sposób organizacji dostępu do zasobów systemów. Pomimo stwierdzenia M. Bates, że przeglądanie jest często jedynym sposobem lokalizowania informacji i zasobów, które nie mogą być bezpośrednio scharakteryzowane za pomocą haseł indeksowych [Bates 2007], to właśnie ten sposób jest podstawą pozyskiwania informacji w narzędziach dostępu przedmiotowego w omawianej klasie systemów informacyjnych.

Chociaż użyta nazwa *narzędzia dostępu przedmiotowego* konotuje dostęp do zasobów na podstawie ich cech treściowych, to narzędzia te mają również szersze wykorzystanie w postaci odsyłania do ChWD na podstawie wybranych cech formalnych. Trzon tego rodzaju narzędzi tworzą wykazy jednostek leksykalnych pełniących funkcje klas grupujących ChWD. Reprezentują one określone kategorie cech elementów zbioru informacyjnego zorganizowane w postaci struktur realizujących dostęp do kolekcji. Struktury te są generowane lub tworzone na podstawie

charakterystyki treściowej tworzonej z wykorzystaniem JIW. Tym samym możliwa jest interpretacja narzędzi dostępu przedmiotowego jako indeksów lub katalogów, w zależności od ich struktury – alfabetycznych lub systematycznych. Podstawową funkcją tego rodzaju narzędzi jest skierowanie użytkownika do podzbioru dokumentów (bezpośrednio w postaci ich ChWD), które zostały zaindeksowane za pomocą określonego wyrażenia. Każde z wyrażen użytych w takim narzędziu posiada status punktu dostępu do zasobów systemu.

W narzędziach dostępu przedmiotowego wykorzystujących alfabetyczną organizację punktów dostępu mamy do czynienia z procesem generowania tego rodzaju układów. Odpowiednie wartości atrybutu przyjętego do reprezentacji cech treści dokumentu są ekstrahowane z rekordu, przenoszone do środowiska hipertekstowego i porządkowane w układzie alfabetycznym (Rys. 4.16 oraz 4.17).



Rys. 4.16 Ahierarchiczne narzędzia dostępu przedmiotowego wykorzystujące słowa kluczowe do reprezentacji treści dokumentów. [Źródło: opracowanie własne]

**AERADE**  
Your quality portal to aerospace & defence resources on the Internet

Cranfield UNIVERSITY

Home | About | AERADE@yoursite | Contact us | Help

**AERADE services**

Search and browse:

- [Aerospace & defence resources](#)
- [DEVISE](#)
- [ESDU Series](#)
- [Reports Archive](#)
- [Shrivenham Index](#)

Inform:

- [ConferenceBrief](#)
- [JobsBrief UK](#)
- [NewsBrief](#)

**AERADE newsletter**

Keep up to date with the latest developments and news in your field. [More...](#)

**AERADE friends**

- [Books Express](#)
- [Intute](#)
- [IEE PN - Aerospace](#)
- [Internet Aviator](#)

**Browse DEVISE alphabetically**

Jump to...

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

**A**

Academic journals

- [Full text academic journals](#)
- [Abstracts/index/partial full text](#)
- [Full text academic journal articles](#)

[Aerospace consultants and consultancy services](#)

[Afghanistan](#)

[Air Forces](#)

[Air Forces - United Kingdom](#)

[Air Forces - NATO](#)

[Air defence systems](#)

[Air delivered weapons](#)

[Air-to-air-missiles](#)

[Air-to-surface missiles](#)

[Air warfare](#)

[Ammunition technology](#)

[Amphibious warfare](#)

[Antennas](#)

[Anti-submarine warfare](#)

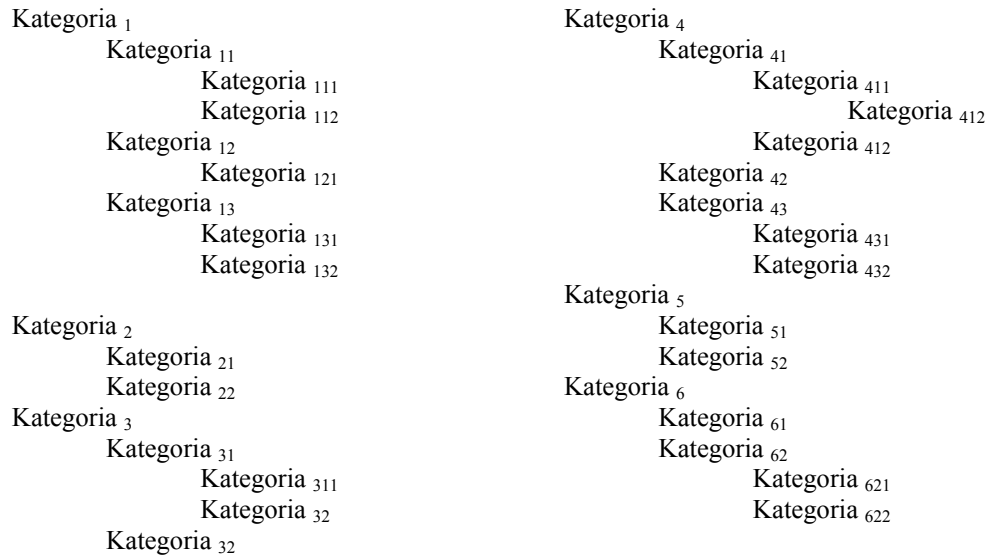
[Armed Forces \(Combined Forces\)](#)

[Armed Forces \(Combined Forces\) - NATO](#)

Rys. 4.17 Ahierarchiczne narzędzia dostępu przedmiotowego w systemie AERADE [Źródło: [http://aerade.cranfield.ac.uk/devise\\_browsea.html](http://aerade.cranfield.ac.uk/devise_browsea.html)]

Funkcjonowanie tworzonych w ten sposób narzędzi dostępu przedmiotowego polega na identyfikacji przez użytkownika wyrażenia, które jest nazwą dla kategorii tematycznej grupującej odesłania do relewantnych ChWD. Zagnieżdżenie takiej struktury w środowisku hipertekstowym umożliwia traktowanie każdego z wyrażen jako węzła powiązania hipertekstowego, którego aktywacja powoduje wyświetlenie użytkownikowi zbioru ChWD zaindeksowanych za jego pomocą. Identyfikacja wyrażenia oraz wykorzystanie jego funkcji jako punktu dostępowego powoduje reorganizację zbioru wyszukiwawczego w postaci wyodrębnienia jego podzbioru spełniającego tak sformułowane kryteria relewancji.

Narzędzia dostępu przedmiotowego wykorzystujące hierarchiczny sposób porządkowania punktów dostępu są bardziej złożone. W tym przypadku zbiór wyrażen reprezentujących określone kategorie cech dokumentów zostaje zorganizowany do postaci struktury drzewiastej (Rys. 4.18 oraz 4.19).



Rys. 4.18 Narzędzie dostępu przedmiotowego w postaci struktury drzewiastej.

[Źródło: opracowanie własne]



Rys 4.19 Narzędzie dostępu przedmiotowego w postaci struktury drzewiastej w systemie BizigGate

[Źródło: <http://www.bizigate.dk/?lang=en>]

Organizacja pola tematycznego wyznaczonego przez sumę zakresów znaczeniowych poszczególnych wyrażen, może być oparta na zaprojektowanym na potrzeby systemu własnym schemacie lub na organizacji pola semantycznego

wybranego języka informacyjno-wyszukiwawczego. Tak skonstruowane narzędzie umożliwia użytkownikowi eksplorację struktury drzewiastej (z wykorzystaniem techniki przeglądania) w poszukiwaniu wyrażenia stanowiącego nazwę dla kategorii grupującej określone ChWD. Wykorzystanie do tego celu węzłów i powiązań hipertekstowych umożliwia manipulowanie tym narzędziem poprzez wyznaczenie ścieżki dostępu, rozpoczynając od kategorii o najszerszym zakresie, przez wyodrębnione podkategorie, aż do identyfikacji i eksploracji kategorii tematycznej poszukiwanej przez użytkownika. W tym przypadku aktywowanie powiązania hipertekstowego dla danej kategorii powoduje wyświetlenie użytkownikowi oprócz rezultatów wyszukiwania, w postaci listy relewantnych ChWD, również wyodrębnionych kategorii bezpośrednio podrzędnych, dla których przewidziano analogiczne powiązania hipertekstowe.

Podstawą do tworzenia tego rodzaju struktur hierarchicznych, pełniących funkcje dostępu do zasobów jest zbiór zidentyfikowanych w polu tematycznym systemu tematów, które są reprezentowane w ChWD przez jednostki leksykalne stosowanego JIW. Duża liczba DSH opracowuje na własne potrzeby indywidualne schematy kategorialne.

Przedstawione dwa sposoby (płaski i hierarchiczny) organizowania dostępu do zasobów systemu z wykorzystaniem narzędzi dostępu przedmiotowego bywają również łączone tworząc postać narzędzia kombinowanego, czy też hybrydowego. Polega to np. na wykorzystaniu struktur hierarchicznych do organizowania wyrażen reprezentujących wyłącznie cechy treści dokumentów, zaś alfabetycznych wykazów wyrażen jako punktów dostępu do zbioru reprezentującego wymiar formy dokumentów (np. wykaz typów dokumentów). Szczegółowa analiza budowy narzędzi dostępu przedmiotowego w wybranych DSH zostanie przedstawiona w rozdziale piątym, a rozwiązania modelowe w rozdziale szóstym.

#### ***4.5. Punkty dostępu w rezultatach wyszukiwania***

Odpowiedzią dziedzinowego systemu hipertekstowego na wysłanie przez użytkownika zapytania informacyjnego (za pomocą wyszukiwarki i/lub narzędzi dostępu przedmiotowego) jest wyświetlenie użytkownikowi charakterystyk wyszukiwawczych spełniających stosowane w systemie kryteria relewancji. Ilość informacji o obiekcie informacyjnym wyświetlana w rezultatach wyszukiwania powinna być optymalna do podjęcia przez użytkownika decyzji o jego wyborze lub

porzuceniu. Pomimo spełniania kryteriów relewancji systemowej, ChWD wyświetlane na wyjściu są ostatecznie analizowane przez użytkownika, który ocenia rzeczywisty poziom pertynencji informacji na wyjściu z systemu.


#### 4.5.1. Tryby wyświetlania rezultatów wyszukiwania

W dziedzinowych systemach hipertekstowych można zidentyfikować dwa tryby wyświetlania rezultatów wyszukiwania – tryb prosty i zaawansowany. Różni je ilość informacji o obiekcie informacyjnym wyświetlanych jako odpowiedź systemu na zapytanie użytkownika. Nie wszystkie dziedzinowe systemy hipertekstowe wykorzystują obydwa tryby, jednak obecność trybu prostego jest domyślnym sposobem prezentacji odpowiedzi systemu na zapytanie informacyjne.

Ilość informacji wyświetlana użytkownikowi w trybie prostym jest zredukowana zazwyczaj do:

- tytułu obiektu informacyjnego,
- URL obiektu informacyjnego,
- charakterystyki słownej obiektu informacyjnego.

Najczęściej stosowanym rozwiązaniem jest wykorzystanie tytułu obiektu informacyjnego jako węzła dla powiązania hipertekstowego, którego uruchomienie spowoduje przekierowanie użytkownika do jego rzeczywistej lokalizacji w Internecie. (Rys. 4.20)



The screenshot shows the AERADE website header with the logo and tagline "Your quality portal to aerospace & defence resources on the Internet". Below the header are navigation links: Home, About, AERADE@yoursite, Contact us, and Help. The main content area is divided into two columns. The left column contains sections for "AERADE services" (with links to Aerospace & defence resources, DEVISE, ESDU Series, Reports Archive, and Shrivenham Index), "AERADE newsletter" (with a link to More...), and "AERADE friends" (with links to Books Express, Intute, IEE PN - Aerospace, and Internet Aviator). The right column features a section titled "Air Forces - NATO" with a sub-section "Air Force Link : Fact Sheets" and a paragraph describing the content provided by the United States Air Force. Below this is another section "Air Force Times.Com" with a paragraph describing it as a weekly US publication.

Rys. 4.20 Tryb prosty wyświetlania rezultatów wyszukiwania w systemie Aerade (<http://aerade.cranfield.ac.uk/>)



Takie podejście umożliwia wyświetlenie większej liczby ChWD na ekranie monitora użytkownika, ale konieczność analizy przez niego krótkich streszczeń dokumentów jest procesem czasochłonnym w większym stopniu niż interpretacja jego reprezentacji treściowej np. w postaci słów kluczowych. Tym samym znaczna liczba DSH w trybie prostym prezentuje rozszerzony zestaw metainformacji. Są to np. reprezentacje treści obiektu informacyjnego wyrażone za pomocą systemów leksykalnych o słownictwie paranaturalnym (np. słowa kluczowe, hasła przedmiotowe) lub sztucznym (symbole klasyfikacyjne) i/lub identyfikacja typu dokumentu (Rys. 4.21, 4.22).

**MathGuide:**  
**Algebraic geometry (30 records)**

Results: 1-20 | 21-30

<b>1. Algebraic Surfaces</b>			
Subject Class	Algebraic geometry		
Source Type	Virtual Exhibitions		
URL	<a href="http://www.mathematik.uni-kl.de/~hunt/drawings.html">http://www.mathematik.uni-kl.de/~hunt/drawings.html</a>		
Contents ★★	Clarity —	Index —	Links —
<b>2. Aluffi, Paolo - Florida State University</b>			
Subject Class	Algebraic geometry; Global analysis, analysis on manifolds; Algebraic topology		
Source Type	Personal Homepage		
URL	<a href="http://www.math.fsu.edu/~aluffi/">http://www.math.fsu.edu/~aluffi/</a>		
Contents ★★	Clarity —	Index —	Links —
<b>3. Annales de l'Institut Fourier</b>			
Subject Class	Differential geometry; Algebraic geometry		
Source Type	Journals		
URL	<a href="http://aif.cedram.org/">http://aif.cedram.org/</a>		
Contents ★★★	Clarity —	Index —	Links —
<b>4. Balcan Journal of Geometry and Its Applications</b>			
Subject Class	Algebraic geometry; Differential geometry		
Source Type	Journals		
URL	<a href="http://www.emis.de/journals/BJGA/">http://www.emis.de/journals/BJGA/</a>		
Contents ★★★	Clarity —	Index —	Links —

Rys. 4.21 Tryb prosty wyświetlania rezultatów wyszukiwania w systemie MathGuide.  
 [Źródło: <http://www.mathguide.de>]

1. **ACT Early Childhood Education - Our shared understandings**

URL: [http://www.decs.act.gov.au/services/shared\\_understandings.htm](http://www.decs.act.gov.au/services/shared_understandings.htm)

Category: [Publications](#) [ECE curriculum documents](#)

From the ACT Department of Education and Training 'Our Shared Understandings' documents the principles of early childhood education. It also describes the early childhood curriculum, the early childhood educator and the early childhood education e... [Full record](#)

2. **First Steps**

URL: <http://www.ecurl.com.au/src/firststep-frameset.html>

Category: [ECE curriculum documents](#) [Supporting Learning in the Home](#) [Literacy & Numeracy](#)

First Steps is an English language K-7 literacy resource, developed by Edith Cowan University for the WA Department of Education, that helps individual teachers, schools and education systems achieve targeted literacy outcomes and standards for th... [Full record](#)

3. **Indigenous preschool profile**

URL: [http://www.dest.gov.au/publications\\_resources/preschool\\_profile/](http://www.dest.gov.au/publications_resources/preschool_profile/)

Category: [National policy](#) [ECE curriculum documents](#)

The Indigenous Preschool Profile is a key tool in the monitoring process of educational outcomes and progress for students in their preschool years in preparation for school. The Profile and an explanatory booklet are available to download. [Full record](#)

Rys. 4.22 Tryb prosty wyświetlania rezultatów wyszukiwania w systemie Edna Online  
[Źródło: <http://www.edna.edu.au/>]

W przypadku systemu MathGuide (Rys. 4.21) tryb prosty wyświetlania rezultatów wyszukiwania obejmuje wartości atrybutów dla treści (ang. subject class) i formy dokumentów (ang. source type), ale również zestaw wskaźników i ich wartości w postaci mniej formalnej. Są to cechy: zawartość (ang. contents), klarowność treści (ang. clarity), materiał pomocniczy (ang. index) oraz odesłania (ang. links), którym w skali od jednego do trzech, w postaci cyfry lub specjalnego symbolu graficznego przypisuje się podczas indeksowania określoną wartość. Wyświetlanie tych elementów w trybie prostym ma na celu raczej intuicyjne określenie poziomu jakości informacji w obiekcie internetowym.

Tryb zaawansowany wyświetlania rezultatów wyszukiwania daje możliwość prezentacji użytkownikowi zbioru pełnych ChWD, wyodrębnionego jako odpowiedź na zapytanie informacyjne. Uruchomienie tego trybu następuje najczęściej poprzez aktywację określonego powiązania hipertekstowego, którego węzeł przybiera postać symbolu graficznego lub etykiet, np.: *pełny rekord* (ang. full record), *metadane* (ang. metadata), *szczegóły* (ang. details), itp. Nie wszystkie zbadane przez autora, dziedzinowe systemy hipertekstowe wykorzystują ten sposób prezentacji wyników wyszukiwania. Wśród systemów, które realizują prezentację wyników wyszukiwania w omawianym trybie można zaobserwować dwa sposoby aktywacji wyświetlania

pełnych danych o dokumencie internetowym. W pierwszym, aktywacja określonego powiązania hipertekstowego powoduje wyświetlenie pełnego rekordu dla wybranego dokumentu w tym samym oknie przeglądarki internetowej. W drugim, polega to na wyświetleniu tych danych w nowym, dodatkowym oknie typu *pop-up*<sup>\*</sup>, tak by użytkownik nie stracił z pola widzenia dotychczasowych rezultatów wyszukiwania.

Zawansowany tryb wyświetlania rezultatów wyszukiwania to wizualizacja pełnej struktury rekordu bazy danych dla danego obiektu informacyjnego. Oczywiście, nie wszystkie informacje o dokumencie są wyświetlane użytkownikowi w tym trybie. Pomijane są m.in. metadane systemowe, czy też nazwisko lub inicjały autora rekordu oraz pozostałe informacje, które nie mają wpływu na ocenę przez użytkownika przydatności obiektu informacyjnego. Tym samym zestaw metadanych wyświetlanych w trybie zaawansowanym jest zróżnicowany dla danego systemu (Rys. 4.23, 4.24).

<b>Title</b>	<b>Astrochemistry : The Astrobiology Web</b>
<b>Description</b>	This site from the Astrobiology Web focuses on astrochemistry. It provides links to research agencies, which contain experimental data and references to the projects under way. The site also has a search engine and a newsboard. There is a subscription facility to receive updates by email. The page also contains links to several articles on recent research into chemical reactions in interstellar space.
<b>Keywords - uncontrolled</b>	<a href="#">astrochemistry</a> ; <a href="#">origins of life</a> ; <a href="#">stars</a> ; <a href="#">planets</a> ; <a href="#">comets</a> ; <a href="#">dust clouds</a>
<b>Type</b>	Resource guides and directories
<b>URL</b>	<a href="http://www.astrobiology.com/astrochemistry.html">http://www.astrobiology.com/astrochemistry.html</a> [English]
<b>Classification</b>	<a href="#">Astronomy</a> > <a href="#">Astrochemistry and Astrobiology</a> <a href="#">Chemistry</a> > <a href="#">Chemistry of Materials and Environments</a>
<b>Publisher</b>	SpaceRef Interactive Inc
<b>Format</b>	html

Rys. 4.23 Tryb zaawansowany wyświetlania rezultatów wyszukiwania w systemie Intute

[Źródło: <http://www.intute.ac.uk/>]

---

<sup>\*</sup> Okna typu *pop-up* (tzw. wyskakujące okna) to wyświetlanie informacji w nowym oknie przeglądarki, w taki sposób, aby użytkownik nie tracił z pola widzenia informacji aktualnie prezentowanych.

#### BIBLIOGRAPHIC DATA

<b>Title</b>	Algebraic Surfaces
<b>Author</b>	Hunt, Bruce
<b>Publisher</b>	Universität Kaiserslautern - Fachbereich Mathematik: Kaiserslautern, DE
<b>Distributor</b>	Universität Kaiserslautern - Fachbereich Mathematik: Kaiserslautern, DE
<b>Language</b>	English
<b>Country (State)</b>	Germany
<b>Format of data</b>	text/html
<b>Keywords</b>	University of Kaiserslautern; algebraic geometry; pictures; computer graphics; vort; knot singularities; imag
<b>Description</b>	This page contains a gallery of algebraic surfaces which are visualized with vort (very ordinary ray tracing). algebraic surfaces and to give a rough understanding in the theory of knot singularities for algebraic surface the symmetry group and some further explanations are downloadable.
<b>URL</b>	<a href="http://www.mathematik.uni-kl.de/~hunt/drawings.html">http://www.mathematik.uni-kl.de/~hunt/drawings.html</a>
<b>Contained in</b>	<a href="http://www.mathematik.uni-kl.de/">http://www.mathematik.uni-kl.de/</a>

#### CLASSIFICATION

<b>Source Type</b>	Virtual Exhibitions
<b>MSCverbal</b>	Singularities of surfaces
<b>MSC</b>	14J17

#### ADDITIONAL INFORMATION

<b>Temporal Coverage</b>	1993 -
<b>Access</b>	free
<b>Contents ★★</b>	<b>Level</b> popular - professional

#### STATISTICS

<b>Backlinks</b>	100
<b>© SSG-FI</b>	1998-05-20

Rys. 4.24 Tryb zaawansowany wyświetlania rezultatów wyszukiwania w systemie MathGuide  
[Źródło: <http://www.mathguide.de>]

Zestaw danych, których wartości są wyświetlane w tym trybie obejmuje zazwyczaj:

- tytuł obiektu informacyjnego,
- nazwisko i imię autora i/lub nazwę instytucji sprawczej,
- adres URL,
- charakterystykę treściową wyrażoną w języku naturalnym - abstrakt, streszczenie,
- charakterystykę treściową - wyrażenia pochodzące ze zbioru słownictwa o różnym stopniu kontroli (słowa kluczowe, hasła przedmiotowe, deskryptory, symbole klasyfikacyjne, itp.)
- dodatkowe informacje:
  - o typ obiektu informacyjnego,
  - o język obiektu informacyjnego,
  - o data utworzenia/aktualizacji obiektu informacyjnego,
  - o data utworzenia rekordu,
  - o forma dostępu do obiektu informacyjnego.

W prezentacji wyników wyszukiwania w trybie prostym i zaawansowanym w dziedzinowych systemach hipertekstowych wykorzystuje się również sieć relacji intertekstualnych. Polega to na przyjęciu założenia, że za zbiór tekstów, w którym wyznacza się określone rodzaje relacji intertekstualnych uważa się zestaw wszystkich charakterystyk wyszukiwawczych obiektów informacyjnych przechowywanych w bazie systemu. Kolejnym krokiem jest wykorzystanie hipertekstu jako środowiska prezentacji informacji do generowania powiązań pomiędzy tekstami ChWD oraz identyfikacja wykładników tego typu relacji w strukturze ChWD. W przypadku DSH aktualizacja powiązań o charakterze intertekstualnym jest możliwa podczas wspomnianych dwóch trybów wyświetlania rezultatów wyszukiwania. Najczęstszym sposobem wykorzystywania tego rodzaju relacji jest ustanowienie jej wykładnika w postaci elementów opisu rzeczowego obiektu informacyjnego. Tym samym wartości tych atrybutów wyświetlane użytkownikowi w ChWD jako rezultaty wyszukiwania mają status węzła hipertekstowego (Rys. 4.25).

**intute : health & life sciences** [Intute Home](#) [Contacts](#) [Help desk](#) [Site map](#) [Survey](#)

Quick links

[Medicine](#) | [Nursing, Midwifery and Allied Health](#) | [Veterinary](#) | [Bioresearch](#) | [Natural History](#) | [Agriculture, Food and Forestry](#)  
[BioethicsWeb](#) | [MedHist](#) | [Psci-com](#)

Intute > [Health and Life Sciences](#) > Full record display [\[Back to record list\]](#)

<b>Subject links</b>	<b>Title</b>	<b>Public health interventions for prevention and control of avian influenza</b>
<a href="#">About us</a>	<b>Description</b>	This manual for improving biosecurity in the food chain has been produced by the South East Asian Regional Office of the World Health Organization. Focusing on live animal markets and those who work in them, the manual outlines simple but effective public health measures aimed at enhancing biosecurity during food production, transportation marketing and consumption, to ensure that the potential for transmission the of avian influenza (AI) virus from animals to humans is reduced. The manual includes sections on factors influencing the spread of AI, methods of enhancing biosecurity throughout the food chain, the protection of humans, healthy food markets, and key public health messages. The document was issued in March 2006.
<b>Internet catalogue</b>	<b>Keywords - controlled</b>	<a href="#">biosafety</a> , <a href="#">health hazards</a> , <a href="#">avian influenzavirus</a> , <a href="#">influenza in Birds / prevention &amp; control</a> , <a href="#">food security</a> , <a href="#">Public Health / education</a>
<a href="#">Advanced search</a>	<b>Type</b>	Papers/reports/articles (individual)
<a href="#">Browse</a>	<b>URL</b>	<a href="http://www.searo.who.int/LinkFiles/Publications_and_Documents_WHO_WetMarket">http://www.searo.who.int/LinkFiles/Publications_and_Documents_WHO_WetMarket</a> [English]
<a href="#">New resources</a>	<b>Format</b>	PDF
<a href="#">Suggest a site</a>	<b>Technical requirements</b>	Adobe Acrobat (PDF)
<b>Internet training</b>	<b>Gateway</b>	agrifor; omni; vetgate
<a href="#">Virtual Training Suite</a>		
<a href="#">Support materials</a>		
<a href="#">Intute events</a>		
<b>Additional services</b>		
<a href="#">Blog</a>		
<a href="#">Hot topics</a>		

Rys. 4.25 Relacje intertekstualne w systemie Intute. [Źródło: <http://intute.ac.uk/>]

Elementy ChWD reprezentujące treść obiektów informacyjnych i wyrażone za pomocą środków leksykalnych stosowanego w systemie JIW są argumentami relacji intertekstualnych. Relacja ta pozwala na wyodrębnienie zbioru charakterystyk wyszukiwawczych dokumentów opisanych za pomocą tego samego hasła, słowa

kluczowego, itp., czyli dokumentów na ten sam temat. Związek ten zachodzi więc bezpośrednio pomiędzy takimi dwoma ChWD, które w określonym polu służącym do odwzorowania treści dokumentów za pomocą przyjętego JIW, stosują wyrażenie o tej samej, określonej formie. Pośrednio zachodzi pomiędzy każdymi dwoma obiektami informacyjnymi, których charakterystyki wyszukiwawcze przechowuje system, i które spełniają wymienione warunki. Np. aktywacja powiązania dla wyrażenia *avian influenza virus (wirus ptasiej grypy)* (Rys. 4.25) w rezultatach wyszukiwania spowoduje wyodrębnienie i wyświetlenie przez system ChWD, w których to wrażenie występuje jako wartość pola *Słowa kluczowe – kontrolowane*.

Wykorzystanie relacji intertekstualnych wkomponowanych w środowisko hipertekstowe w takiej postaci z pewnością wpływa korzystnie na funkcjonalność całego systemu. Charakterystyka treści obiektu informacyjnego zamieszczona w ChWD w postaci jednostek leksykalnych przyjętego JIW pełni więc jednocześnie funkcje metainformacyjną i wyszukiwawczą.

Nie wszystkie dziedzinowe systemy hipertekstowe stosują ten sposób tworzenia dodatkowych powiązań pomiędzy elementami zbioru wyszukiwawczego. W systemach, które wykorzystują ten rodzaj relacji, jest to ograniczone wyłącznie do elementów opisu treści. Nie stosuje się powiązania np. dla obiektów informacyjnych posiadających wspólnego autora, czy też dla innej cechy opisu formalnego.

#### **4.5.2. Porządkowanie rezultatów wyszukiwania**

Najczęściej wykorzystywanym kryterium porządkowania listy rezultatów wyszukiwania w DSH jest zastosowanie porządku alfabetycznego. Lista relewantnych ChWD wyświetlana przez system jest najczęściej uporządkowana alfabetycznie, według tytułów obiektów informacyjnych. W zależności od konkretnego systemu, użytkownik ma możliwość zmiany sposobu sortowania rezultatów wyszukiwania. Istnieje tutaj duża rozbieżność w oferowanych przez systemy kryteriach porządkowania rezultatów wyszukiwania. Obok najczęściej stosowanego porządku alfabetycznego ChWD według tytułów dokumentów, można spotkać także sortowanie według współczynnika relewancji (np. systemy Intute, MathGuide, ForestryGuide). Wartość wskaźnika, jak również algorytm jego obliczania nie ma charakteru jawnego. Intute, jeden z największych DSH, stosuje dodatkowo możliwość uporządkowania rezultatów wyszukiwania z podziałem na typy obiektów informacyjnych, których charakterystyki wyszukiwawcze przechowuje i udostępnia. System MathGuide reprezentuje najbardziej

zaawansowany sposób porządkowania rezultatów wyszukiwania wśród zbadanych przez autora systemów. Udostępnia aż osiem kryteriów sortowania wyników wyszukiwania. Są to:

- tytuł obiektu informacyjnego,
- symbol klasy Mathematical Subject Classification,
- typ obiektu informacyjnego,
- nazwa państwa (zakres geograficzny),
- język,
- data utworzenia,
- data wprowadzenia do systemu,
- współczynnik relewancji.

Systemy wykorzystujące więcej niż jedno kryterium porządkowania rezultatów wyszukiwania nie oferują możliwości łączenia zasad sortowania (np. posortowanie wyników wyszukiwania według tytułów dokumentów, a następnie wydzielenie z nich grup określonych typów dokumentów).

#### ***4.6. Podsumowanie***

Przedstawiona w rozdziale charakterystyka organizacji informacji w DSH miała na celu przede wszystkim wskazanie na udział w tym procesie języka informacyjno-wyszukiwawczego. Jego rola została przedstawiona w kontekście reprezentacji informacji, identyfikacji punktów dostępu i ich organizacji w mechanizmach wyszukiwawczych i narzędziach dostępu przedmiotowego. Rola JIW w reprezentacji informacji w DSH polega na wykorzystaniu jego zasobu leksykalnego do odwzorowania wartości dla wybranych atrybutów. Reprezentacja wiedzy w DSH charakteryzuje się bezpośrednim oznaczaniem tematów i dodatkowych cech wymiaru treści z jednoczesną redukcją gramatyczną zdań JIW do postaci jednostek autosyntaktycznych i ograniczeniem roli jednostek pomocniczych. W przypadku cech pozat treściowych mamy do czynienia z zastosowaniem licznych kartotek wzorcowych oraz zbiorów słownictwa kontrolowanego.

Przeprowadzone przez autora badania ilościowe nad wykorzystaniem JIW w reprezentacji wiedzy w DSH wskazały, że największy udział w tym procesie ma JIW typu klasyfikacji. Do oznaczania zakresu tematycznego obiektów informacyjnych za pomocą tego typu JIW stosuje się symbole klasyfikacyjne oraz ich odpowiedniki słowne. Powszechnym zjawiskiem w odwzorowywaniu treści w DSH jest polireprezentacja wiedzy.

Identyfikacja cech wyszukiwawczych w DSH polega na wyznaczeniu w formacie opisu elementów, które mają zostać włączone w mechanizmy realizujące dostęp do kolekcji. W DSH można wyodrębnić zestaw stałych punktów dostępu. Są to

- tytuł obiektu informacyjnego,
- adres URL obiektu informacyjnego,
- słowo kluczowe,
- słowo w całej charakterystyce wyszukiwawczej oraz
- kategoria tematyczna reprezentująca wymiar treści i/lub formy obiektu informacyjnego w narzędziach dostępu przedmiotowego.

Przedstawiona charakterystyka mechanizmów wyszukiwawczych wskazała na sposoby tworzenia zapytań informacyjnych również z wykorzystaniem JIW jako narzędzia pomocniczego. Budowa narzędzi dostępu przedmiotowego wykazuje zastosowanie w niej elementów strukturalnych JIW. Zagadnieniom tym poświęcone zostaną kolejne rozdziały rozprawy.



## 5. Język informacyjno-wyszukiwawczy w narzędziach dostępu przedmiotowego wybranych dziedzinowych systemów hipertekstowych

W tym miejscu rozprawy zostanie przedstawiona charakterystyka narzędzi dostępu przedmiotowego w pięciu wybranych dziedzinowych systemach hipertekstowych. Są to następujące systemy:

- BUBL ([www.bubl.ac.uk](http://www.bubl.ac.uk)),
- CISMef ([www.cismef.org/](http://www.cismef.org/)),
- GEM ([www.thegateway.org](http://www.thegateway.org)),
- Intute ([www.intute.ac.uk](http://www.intute.ac.uk)),
- MathGuide ([www.mathguide.de](http://www.mathguide.de)).

Każdy z wymienionych systemów spełnia warunki definicyjne dziedzinowego systemu hipertekstowego, które przedstawiono w trzecim rozdziale pracy. Motywacją do prezentacji tego rodzaju treści była zdaniem autora konieczność ukazania szczegółowej charakterystyki struktury tego typu narzędzi z uwzględnieniem funkcji, jaką pełnią w nich konkretne języki informacyjno-wyszukiwawcze. Wybór pięciu systemów podyktowany był odmiennymi sposobami budowy narzędzi dostępu przedmiotowego oraz wykorzystaniem w nich elementów JIW. Jednocześnie wskazane systemy wchodziły w skład grupy reprezentatywnej 30 dziedzinowych systemów hipertekstowych, które autor poddał szczegółowym badaniom. Zatem treści przedstawione tutaj stanowią jedynie wycinek z wyników badań przeprowadzonych przez autora i zarazem poprzedzają ich wnioski, które zostaną przedstawione w kolejnym rozdziale rozprawy.

Przyjęta perspektywa badawcza zakłada analizę wskazanego typu narzędzi pod kątem:

- zastosowania w ich budowie języka informacyjno-wyszukiwawczego,
- identyfikacji podstawowych cech strukturalnych narzędzi,
- ustalenia zależności ilościowych i jakościowych pomiędzy kategoriami tematycznymi.

Analiza poszczególnych systemów została poprzedzona ich krótką charakterystyką, w której uwzględniono zakres i wielkość zbioru wyszukiwawczego, kryteria włączania obiektów informacyjnych do kolekcji oraz identyfikację grupy użytkowników docelowych.

## 5.1. *BUBL Information Service*

System BUBL powstał w 1993 roku w Wielkiej Brytanii. Jego nazwa była akronimem dla Bulletin Board for Libraries. Obecnie pełna nazwa systemu to BUBL LINK Catalogue of Internet Resource (BUBL LINK Katalog Zasobów Internetu) Jest to pierwszy narodowy system o zakresie ogólnym gromadzący informacje o dokumentach opublikowanych w Internecie. [Gold 1995] Zastosowane w nim narzędzie dostępu przedmiotowego nazwano wówczas *drzewem przedmiotowym* (ang. *subject tree*), co spowodowało stosowanie tej nazwy w odniesieniu do systemów powstających później. Zakres systemu obejmuje całe uniwersum wiedzy ze szczególnym naciskiem na zagadnienia bibliotekoznawstwa i informacji naukowej. [Dawson, Simpson 1997] Grupę użytkowników docelowych stanowi przede wszystkim środowisko akademickie. Instytucją, która sprawuje nadzór merytoryczny oraz zarządza systemem jest Centrum Badań nad Biblioteką Cyfrową (Centre for Digital Library Research) przy szkockim Uniwersytecie w Strathclyde w Glasgow. [Vizine-Goetz, Beall 2004] Oprócz narzędzi dostępu przedmiotowego system posiada mechanizm wyszukiwawczy umożliwiający wysyłanie zapytania w postaci ciągu niekontrolowanych słów kluczowych. Uznano to za ważną, chociaż niesatysfakcjonującą funkcję w kontekście efektywności wyszukiwania informacji. [Koch 1996] Dlatego też przedmiotem zainteresowania jego twórców było stworzenie efektywnego narzędzia dostępu przedmiotowego. Dotychczasową strukturę hierarchiczną kategorii tematycznych opartą na Uniwersalnej Klasyfikacji Dziesiątej zastąpiono w 1996 roku nową, nazwaną LINK (Libraries of Networked Knowledge), a obecnie nazywa się po prostu *Dewey*. [Hofman, Worsfold 1997] Powstało ono na bazie systematyzacji klas w Klasyfikacji Dziesiątej Deweya (KDD), co było możliwe dzięki otrzymaniu odpowiedniej licencji z OCLC. System gromadzi około 12 tysięcy charakterystyk wyszukiwawczych obiektów informacyjnych opublikowanych w Internecie (stan na 02-01-2007 r.), które spełniają kryteria jakościowe opracowane w raporcie z przytaczanego w rozprawie projektu DESIRE. [Hofman, Worsfold 1997]

Dostęp przedmiotowy do zasobów systemu jest realizowany przez następujące narzędzia: (Rys. 5.1)

- alfabetyczny wykaz kategorii grupujących ChWD według treści oraz formy:  
*subject menus*,

- hierarchiczna struktura kategorii tematycznych zbudowanych na podstawie Klasyfikacji Dziesiętnej Deweya oraz ich odpowiedników słownych w postaci wyrażen języka naturalnego: *Dewey*
- alfabetyczny wykaz kategorii grupujących ChWD według nazw państw: *countries*,
- alfabetyczny wykaz kategorii grupujących ChWD według cech formalnych (typ dokumentu): *types*.



**BUBL LINK** Catalogue of Internet Resources

[Dewey](#) | [Search](#) | [Subject Menus](#) | [Countries](#) | [Types](#) | [BUBL UK](#) | [BUBL Archive](#)

Selected Internet resources covering all academic subject areas

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

**000 Generalities**

*Includes:* computing, Internet, libraries, information science

**100 Philosophy and psychology**

*Includes:* ethics, paranormal phenomena

**200 Religion**

*Includes:* bibles, religions of the world

**300 Social sciences**

*Includes:* sociology, politics, economics, law, education

**400 Language**

*Includes:* linguistics, language learning, specific languages

**500 Science and mathematics**

*Includes:* physics, chemistry, earth sciences, biology, zoology

**600 Technology**

*Includes:* medicine, engineering, agriculture, management

**700 The arts**

*Includes:* art, planning, architecture, music, sport

**800 Literature and rhetoric**

*Includes:* literature of specific languages

**900 Geography and history**

*Includes:* travel, genealogy, archaeology

[E-LIS](#) | [CDLR Projects](#) | [Contacts and Credits](#)

Rys. 5.1. Narzędzia dostępu do zasobów systemu BUBL. [Źródło: <http://bubl.ac.uk/>]

Głównym narzędziem dostępu przedmiotowego jest wielopoziomowa struktura kategorii tematycznych zbudowana z wykorzystaniem Klasyfikacji Dziesiętnej Deweya. Na pierwszym stopniu podziału wyodrębniono 10 kategorii tematycznych odpowiadających w planie treści i wyrażania klasom głównym KDD:

- 000 Dział ogólny (Generalities),
- 100 Filozofia i psychologia (Philosophy and psychology),
- 200 Religia (Religion),
- 300 Nauki społeczne (Social sciences),
- 400 Język (Language),
- 500 Nauki ścisłe i matematyka (Science and mathematics),
- 600 Nauki techniczne (Technology),
- 700 Sztuka (The arts),
- 800 Literatura i retoryka (Literature and rhetoric),
- 900 Geografia i historia (Geography and history).

Nazwa kategorii tematycznej stanowi węzeł dla odesłania hipertekstowego, którego uruchomienie powoduje wizualizację rozbudowy danej klasy macierzystej oraz wyświetlenie listy ChWD przyporządkowanych do aktualnie eksplorowanej. Struktura hierarchiczna BUBL zbudowana jest z 1362 kategorii tematycznych zgrupowanych na siedmiu poziomach. Dystrybucję kategorii tematycznych w tym układzie prezentuje Tab. 5.1.

Tab. 5.1 Rozkład kategorii tematycznych w strukturze hierarchicznej BUBL.

[Źródło: opracowanie własne]

Poziom	Liczba kategorii tematycznych	Procentowy rozkład kategorii tematycznych
Poziom 1	10	0,73%
Poziom 2	130	9,54%
Poziom 3	467	34,29%
Poziom 4	388	28,49%
Poziom 5	273	20,05%
Poziom 6	84	6,17%
Poziom 7	10	0,73%
Suma	1362	100,00%

Przedstawione w Tab. 5.1 dane wskazują na duży udział kategorii tematycznych wyodrębnionych na poziomach od trzeciego do piątego. Ponad 80% łącznej liczby kategorii zostało wyróżnionych na wskazanych poziomach. Interesującym rozwiązaniem w BUBL jest grupowanie ChWD wyłącznie w kategoriach tematycznych stanowiących ostatnie ogniwo łańcucha. Tym samym odnalezienie ChWD wiąże się z koniecznością identyfikacji kategorii tematycznej, w której ją zgrupowano poprzez eksplorację całego łańcucha, rozpoczynając od kategorii tematycznej najwyższego poziomu. Przejmowanie kategorii tematycznych ze struktury KDD oparto na analizie przedmiotowej kolekcji i wyodrębnieniu tematów dla obiektów informacyjnych. Dla każdego elementu zbioru informacyjnego odszukuje się kategorię tematyczną o odpowiednim zakresie i przyporządkowuje się go do niej za pomocą odesłania hipertekstowego. W Tab. 5.2 przedstawiono przykładową organizację kategorii tematycznej 020 Bibliotekoznawstwo i informacja naukowa, jaką zastosowano w BUBL z wykorzystaniem KDD.

Tab. 5.2 Organizacja kategorii tematycznej 020 Bibliotekoznawstwo i informacja naukowa w BUBL.

[Źródło: opracowanie własne]

020 Library and information sciences
020 Library and information science: general resources
020 Library and information science: journals
020 Library and information science: news and discussion lists
020 Library and information studies departments in the UK
020 Library and information studies departments worldwide
020.6 Library organisations
020.6 Library organisations in the UK

- 020.6 Library organisations in the USA
- 020.6 Library organisations worldwide
- 021 Cooperation and resource sharing
- 022 Administration of the physical plant
- 023 Personnel administration in libraries
- 025 Operations of libraries and information centres
  - 025.0 Library systems and information technology in libraries
    - 025.0 Library systems: general resources
    - 025.0 Library systems: journals and magazines
    - 025.0 Library systems: projects
    - 025.0 Library systems: societies
    - 025.0 Library systems: suppliers
    - 025.0 Library systems: Z39.50
  - 025.1 Library administration and management
  - 025.2 Acquisitions (AcqLink)
    - 025.2 Acquisitions: general resources
    - 025.2 Acquisitions: discussion lists
    - 025.2 Acquisitions: electronic order forms
    - 025.2 Acquisitions: journals
    - 025.2 Acquisitions: services
    - 025.2 Library suppliers
    - 025.2 Censorship
    - 025.2 News
  - 025.3 Bibliographic analysis and control
    - 025.3 Bibliographic analysis and control: general resources
    - 025.3 Cataloguing of Internet and multimedia resources
    - 025.3 Indexing
    - 025.3 Metadata standards and research
    - 025.3 Online public access catalogues
    - 025.3 National Library OPACs in the UK
    - 025.3 Subject-specific and research library OPACs in the UK
    - 025.3 University OPACs in England
    - 025.3 College OPACs in England
    - 025.3 London University OPACs
    - 025.3 Other university and college OPACs in London
    - 025.3 University and college OPACs in Scotland
    - 025.3 University and college OPACs in Wales
    - 025.3 University and college OPACs in Northern Ireland
    - 025.3 University and college OPACs in Ireland
    - 025.3 Worldwide OPACs
  - 025.4 Subject analysis and control
  - 025.5 Services to library users
    - 025.5 Services to library users: general resources
    - 025.5 Services to library users: journals and magazines
    - 025.5 Services to library users: CD-ROM user guides
      - 025.524 Information search and retrieval
  - 025.6 Circulation services (including document delivery and ILL)
  - 025.8 Maintenance and preservation of collections
  - 025.0 Digital libraries and online services
    - 025.0 Digital libraries: general resources
    - 025.0 Digital libraries: national centres
    - 025.0 Digital libraries: journals
    - 025.0 Digital libraries: projects
    - 025.0 Digital libraries in universities and colleges
    - 025.0 Authentication and access management
    - 025.0 Evaluating Internet resources and services
    - 025.0 Image digitisation
    - 025.0 Multilingual access to digital libraries
    - 025.0 Pricing in digital libraries
    - 025.0 Searching and use of digital libraries
    - 025.04 Internet resource directories

Odpowiedniki słowne symboli KDD nie są zgodne we wszystkich przypadkach z ich formą umieszczoną w tablicach. Np. w BUBL wydzielono kategorię tematyczną *025.1 Library administration and management (Administracja i zarządzanie biblioteką)*, której odpowiednik słowny w KDD ma formę *Administration (Administracja)*.

Interesująca jest zasada wyodrębniania kategorii tematycznych na podstawie podziałów wspólnych KDD. W przypadku kategorii tematycznej 020 *Bibliotekoznawstwo i informacja naukowa* na drugim stopniu podziału, wydzielono kilka kategorii podrzędnych, którym przypisano ten sam symbol klasy macierzystej. Tymczasem rzeczywiste symbole tych kategorii zbudowano z wykorzystaniem podziałów wspólnych, czego nie odwzorowano w nazwie kategorii tematycznej (Tab. 5.3).

Tab. 5.3 Różnice w budowie nazwy kategorii BUBL i symbolu KDD.

Fragment rozbudowy kategorii 020 Bibliotekoznawstwo i informacja naukowa w BUBL	Symbol odpowiadającej klasy w KDD
020 Library and information science: journals	020.5
020 Library and information studies departments in the UK	020.711
020 Library and information studies departments worldwide	020.71141

W innych przypadkach wyodrębnia się również kategorie podrzędne odwzorowując to w rozbudowie ich symbolu, np. w tej samej kategorii 020 *Bibliotekoznawstwo i informacja naukowa* wyodrębniono kategorię podrzędną 020.6 *Organizacje bibliotekarskie (Library organisations)* zbudowaną za pomocą podziału wspólnego z KDD o wykładniku „.6”.

Kategorie wyodrębniane na poziomach od czwartego do siódmego charakteryzują się specyficznym sposobem odwzorowania miejsca w strukturze w rozbudowie symbolu stanowiącego nazwę klasy. Chociaż ich zakres jest węższy od kategorii macierzystej, to w planie wyrażania symbol KDD przyjmuje wartość kategorii nadrzędnej, a jego odpowiednik słowny określa rzeczywisty zakres danej kategorii podrzędnej. Sytuacja taka ma miejsce np. w podziale klasy 025.3 *Bibliographic analysis and control (Analiza i kontrola bibliograficzna)*, w której na tym samym stopniu wyodrębniono m.in.:

- 025.3 Analiza i kontrola bibliograficzna – źródła ogólne (ang. Bibliographic analysis and control: general resources),
- 025.3 Katalogowanie źródeł internetowych i multimedialnych (ang. Cataloguing of Internet and multimedia resources),
- 025.3 Indeksowanie (ang. Indexing),
- 025.3 Metadane – standardy i badanie (ang. Metadata standards and research),
- 025.3 OPAC (ang. Online public access catalogues).

Symbole klas współrzędnych są tożsame z symbolem klasy macierzystej, chociaż ich

zakres jest różny, na co wskazują ich nazwy wyrażone w języku naturalnym. Zabieg ten wykonano prawdopodobnie w celu uniknięcia zbyt głębokiej rozbudowy struktury hierarchicznej, która mogłaby przyczynić się do obniżenia funkcjonalności całego narzędzia. ChWD zgromadzone w wymienionych klasach współrzędnych, w polu przeznaczonym do odwzorowania treści za pomocą symboli KDD, zawierają jednak symbole tej klasyfikacji, które są adekwatny do ich zakresu treściowego. Tym samym np. w kategorii *025.3 Katalogowanie źródeł internetowych i multimedialnych* zgrupowano CHWD, którym podczas indeksowania za pomocą KDD nadano m.in. następujące symbole: 025.3, 025.344, 025.3432.

Zasadą wyodrębniania kategorii tematycznych w tym narzędziu jest zasada *literary warrant*\*. Tak więc, w przypadku odpowiedniej liczby ChWD, których treść jest reprezentowana przez symbole KDD o zakresie szczegółowym, w systemie wyodrębnia się odpowiednie kategorie wraz z odpowiadającymi im symbolami klasyfikacyjnymi. Taka sytuacja ma miejsce np. w rozbudowie kategorii tematycznej *700 Sztuka*, gdzie stworzono łańcuch:

*700 Sztuka → 780 Muzyka → 780 Muzyka: kolekcje → 780.9 Muzyka: określone okresy → 780.9031 Muzyka renesansowa: 1450 do 1600.*

Równie częstym sposobem porządkowania punktów dostępu, z wykorzystaniem porządku pionowego ustalonego przez KDD, jest wyodrębnianie na tym samym stopniu kategorii tematycznych pochodzących z różnych stopni podziału danej kategorii macierzystej, np. poprzez podział klasy *551.5 Meteorologia* wyodrębniono klasy współrzędne:

- 551.5 Meteorologia: źródła ogólne,
- 551.5 Meteorologia: centra narodowe,
- 551.552 Huragany,
- 551.578 Śnieg i lawiny.

Takie podejście do wykorzystania strukturalizacji KDD w organizacji informacji w systemie podyktowane jest zachowaniem funkcjonalności przy minimalnych stratach na spójności całej struktury.

Chociaż struktura tego narzędzia w systemie zakłada wykorzystanie siedmiopozomowej struktury hierarchicznej kategorii tematycznych, to przeważającą

---

\* zasada *literary warrant* - uzależnienie tworzenia jednostek leksykalnych reprezentujących określone tematy i nazwy przedmiotów od ich występowania w treści analizowanego zbioru dokumentów. [ANSI/NISO Z39.19-2005, s. 162]

większość ChWD zgromadzono w kategoriach na poziomach wyższych (Tab. 5.4).

Tab. 5.4 Dystrybucja odesłań do ChWD w strukturze hierarchicznej BUBL.

[Źródło: opracowanie własne]

Stopień podziału	Liczba odesłań do ChWD	Udział procentowy
Poziom 1	0	0%
Poziom 2	762	6,72%
Poziom 3	4032	35,57%
Poziom 4	3515	31,01%
Poziom 5	2226	19,64%
Poziom 6	732	6,46%
Poziom 7	67	0,59%
Suma	11334	100,00%

Rozkład odesłań do ChWD na poszczególnych poziomach układu hierarchicznego ma swoje odzwierciedlenie w ilościowej dystrybucji kategorii. W tym przypadku również ponad 80% łącznej liczby odesłań do ChWD zgromadzono na poziomach od trzeciego do piątego. Dystrybucja ChWD przypadających na jedną kategorię została przedstawiona w Tab. 5.5.

Tab. 5.5 Średnia pojemność kategorii tematycznej w BUBL. [Źródło: opracowanie własne]

Stopień podziału	Średnia pojemność klasy (liczba ChWD)
Poziom 1	0
Poziom 2	6
Poziom 3	9
Poziom 4	9
Poziom 5	8
Poziom 6	9
Poziom 7	7

Dane przedstawione w Tab. 5.5 wskazują na stały współczynnik pojemności kategorii na poszczególnych poziomach w strukturze narzędzia dostępu przedmiotowego. Średnio na jedną kategorię w strukturze hierarchicznej przypada około ośmiu ChWD. Świadczy to o efektywnej analizie kolekcji pod kątem dystrybucji tematów, które są reprezentowane przez klasy przejęte z KDD oraz o zachowaniu spójności całego narzędzia.

Katalogi przedmiotowe są kolejnym narzędziem dostępu przedmiotowego, które użyto w systemie. Pierwszy z nich nazwany *menu przedmiotowym* (ang. subject menu) zbudowany jest ze 138 wyrażen języka naturalnego reprezentujących tematy ogólne. [BUBL 2007] Każde z nich jest odpowiednikiem kategorii tematycznej ze struktury hierarchicznej zredukowanym w planie wyrażania do terminu jedno- lub



dwuwyrazowego. Jednocześnie jest węzłem odesłania hipertekstowego, którego aktywacja powoduje wyświetlenie zawartości odpowiadającej mu kategorii ze schematu hierarchicznego. Narzędzie to ma spełniać funkcję pomocniczą w identyfikacji kategorii tematycznych taką, jaką pełni indeks przedmiotowy dla katalogu systematycznego.

Drugim rodzajem katalogu przedmiotowego w systemie BUBL jest alfabetyczny wykaz jednostek leksykalnych stanowiących podzbiór słownika Library of Congress Subject Headings (LCSH). [Dawson, Simpson 1997] Słownik tego JIW jest jednak wyłącznie źródłem pomocniczym w tworzeniu tego narzędzia. Pojawiają się w nim wyrażenia niestosowane w LCSH. Składa się on z 1471 wyrażeń reprezentujących treść oraz formę elementów zbioru informacyjnego. Elementy katalogu mają postać jedno- lub wielowyrazową w formie rozpoczynającej się wyłącznie małą literą (również w przypadku nazw własnych). Warunkiem obecności danej jednostki leksykalnej w katalogu jest jej użycie w odpowiednim polu ChWD podczas indeksowania. Dostęp do kolekcji z jego użyciem polega na identyfikacji grupy stanowiącej wybraną literę alfabetu łańciskiego, będącej jednocześnie węzłem odesłania hipertekstowego. W każdej z nich zgromadzono wyrażenia w układzie alfabetycznym rozpoczynające się od danej litery (Rys. 5.2).



Rys. 5.2 Katalog przedmiotowy w BUBL. [Źródło: <http://bubl.ac.uk/>]

Każde hasło katalogu jest węzłem odesłania hipertekstowego, którego uruchomienie powoduje wyświetlenie przez system listy ChWD, w których je użyto do reprezentacji treści i/lub formy. Omawiany katalog oraz struktura hierarchiczna kategorii tematycznych przejęta z KDD są autonomicznymi narzędziami dostępu.

Katalog przedmiotowy został sprzężony z hierarchicznym schematem kategorii tematycznych. Polega to na wykorzystaniu haseł katalogowych do tworzenia sieci niespecyfikowanych relacji asocjacyjnych. Eksploracja wybranej kategorii tematycznej w strukturze hierarchicznej powoduje, oprócz prezentacji relewantnych ChWD, wyświetlenie skojarzonych wyrażeń pochodzących z katalogu. Wskaźnikiem takiej

relacji jest wyrażenie *zobacz również* (ang. see also) (Rys. 5.3)

### 305.4 Women

See also [Women in history](#), [Women in music](#), [Women in science](#)

<i>Titles</i>	<i>Descriptions</i>
1. <a href="#">American Women's History: A Research Guide</a>	<b><a href="#">American Women's History: A Research Guide</a></b> Citations of print and online reference materials for women's history. <i>Author:</i> Middleton, Ken <i>Dewey Class:</i> 305.4
2. <a href="#">Aviva</a>	
3. <a href="#">Biographies of Women Mathematicians</a>	
4. <a href="#">Diotima: Women and Gender in the Ancient World</a>	<b><a href="#">Aviva</a></b> An international women's magazine which enables women all over the world to make contact with each other. Includes listings of groups, courses, resources, events, and news items. <i>Author:</i> Aviva <i>Dewey Class:</i> 305.4
5. <a href="#">Documents from the Women's Liberation Movement</a>	
6. <a href="#">Emma Goldman Papers</a>	
7. <a href="#">Engendering Development through Attention to Gender</a>	
8. <a href="#">Feminist Theory Website</a>	

Rys. 5.3 Powiązanie katalogu przedmiotowego ze schematem hierarchicznym w BUBL. [Źródło: <http://bubl.ac.uk/>]

Dla przedstawionej na Rys. 5.3 klasy 305.4 *Kobiety* zidentyfikowano również trzy terminy skojarzone: *Kobiety w historii*, *Kobiety w muzyce*, *Kobiety w nauce*. Każdy z nich wyposażony jest w odesłanie hipertekstowe, którego aktywacja powoduje przeniesienie użytkownika do katalogu przedmiotowego i wyświetlenie ChWD, które zostały zaindeksowane z ich wykorzystaniem.

Dodatkowymi narzędziami dostępu są katalogi zbudowane z wyrażień odwzorowujących wyłącznie cechy formalne. Pierwszy z nich, *Państwa* (ang. Countries) zawiera 194 wyrażenia będące nazwami państw. Jednostki te poprzez wyposażenie ich w odesłania hipertekstowe stanowią punkty dostępu do ChWD, w których podczas indeksowania przypisano określoną wartość dla odwzorowania tej cechy. Aktywacja wybranego powiązania powoduje wyświetlenie listy ChWD spełniających wybrane kryterium. Drugi katalog - *Typy* (ang. Types) organizuje zbiór wyszukiwawczy według wymiaru cech formalnych określających typy indeksowanych obiektów informacyjnych. Wyodrębniono tutaj 27 wyrażień reprezentujących tę cechę (Tab. 5.6).

Tab. 5.6 Wartości dla cechy typ źródła w katalogu BUBL.

[Źródło: opracowanie własne na podstawie dokumentacji systemu]

Nazwa kategorii	Odpowiednik w języku polskim
Bibliographies	Bibliografie
Biographies	Biografie
Book and text collections	Książki i kolekcje tekstów
Bookshops	Księgarnie
Dictionaries	Słowniki
Directories	Katalogi internetowe
Discographies	Dyskografie
Ebooks: fiction	Ebooki: beletrystyka
Ebooks: non-fiction	Ebooki: popularnonaukowe

Essays	Eseje
Governments	Instytucje rządowe
Image collections	Zbiory dokumentów graficznych
Journal collections and listings	Czasopisma naukowe
Library catalogues	Katalogi biblioteczne
Magazine listings	Czasopisma nienaukowe
Mailing lists	Listy mailngowe
Map collections	Kolekcje kartograficzne
Moving images	Dokumenty audiowizualne
Museums	Muzea
Poems	Wiersze
Proceedings	Sprawozdania
Satellite images	Obrazy satelitarne
Software archives	Oprogramowanie
Sound collections	Dokumenty dźwiękowe
Thesauri	Tezaurusy
UK higher education colleges	Uczelnie wyższe w Wielkiej Brytanii
UK universities	Uniwersytety w Wielkiej Brytanii

Dane przedstawione w Tab. 5.6 wskazują na zróżnicowany charakter cech odwzorowywanych przez te wrażenia. Znalazły się tam terminy określające zarówno rodzaj dokumentu piśmienniczego (sprawozdanie, biografia, słownik), gatunki literackie (wiersze, eseje), a także sposób kodowania informacji (dokumenty tekstowe, graficzne, audiowizualne). Aktywacja powiązania hipertekstowego dla wybranego wyrażenia powoduje wyodrębnienie i wyświetlenie listy ChWD, którym dla cechy *typ źródła* przypisano określoną wartość.

Organizacja punktów dostępu w narzędziach dostępu przedmiotowego BUBL polega na priorytetowym traktowaniu hierarchicznie uporządkowanych kategorii tematycznych. Pomimo braku pełnej zgodności z porządkiem klas w KDD oraz rozbudową symboli (zwłaszcza na niższych poziomach), schemat hierarchiczny BUBL odznacza się spójnością. Świadczy o tym równomierna dystrybucja odesłań przypadających na jedną klasę w strukturze. Interesującym rozwiązaniem w BUBL jest sprzężenie katalogu przedmiotowego opracowanego na podstawie LCSH ze schematem hierarchicznym i wykorzystanie tych terminów do tworzenia sieci niespecyfikowanych relacji asocjacyjnych.

## 5.2. CISMef

CISMef, akronim francuskiej nazwy Katalog i Indeks Zasobów z Zakresu Zdrowia (fr. Catalogue et Index des Sites Médicaux Francophones), to dziedzinowy system hipertekstowy, który powstał w 1995 roku. [Neveol, Soulamia 2005] Celem systemu było wsparcie rozwoju zawodowego osób związanych z medycyną poprzez

udostępnianie charakterystyk i odesłań do elektronicznych źródeł informacji. System powstał przy współudziale szpitala uniwersyteckiego w Rouen. Zakres systemu obejmuje szeroko pojętą medycynę oraz nauki o zdrowiu. Grupa użytkowników docelowych systemu to:

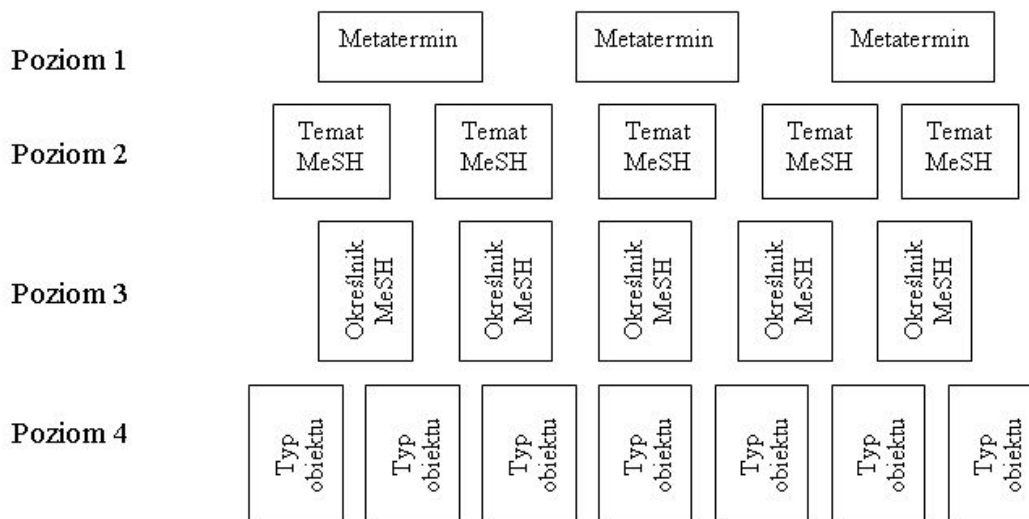
- lekarze (grupa priorytetowa),
- przedstawiciele różnych związanych grup zawodowych: pielęgniarki, położne, weterynarze, dietetycy,
- pacjenci. [Charles 2005]

System gromadzi informacje o internetowych obiektach informacyjnych z podanego zakresu tematycznego, których treść przedstawiono w języku francuskim.

Zbiór wyszukiwawczy systemu tworzą opisy około 13 tysięcy obiektów informacyjnych w World Wide Web (stan na 02-01-2007 r.). Średnie wykorzystanie systemu to około 20 tysięcy odwiedzin dziennie. Przyrost nowych ChWD oscyluje około 55 dziennie. [Neveol, Soulamia, Douyere 2005] Schemat opisu dokumentów sporządzono na podstawie elementów metadanych opartych na kilku standardach m.in. Dublin Core, IEEE LOM. Odwzorowanie treści elementów zbioru informacyjnego odbywa się z wykorzystaniem języka haseł przedmiotowych Medical Subject Headings (MeSH). [Charles 2005] Kryteria selekcji źródeł zostały oparte na wynikach raportu pt. *Net Scoring* wyliczającego 49 kryteriów oceny jakości źródeł elektronicznych. [Darmoni 1997]

Koncepcja organizacji punktów dostępu w narzędziach dostępu przedmiotowego CISMef polega na wykorzystaniu leksyki oraz struktury paradygmatycznej języka Medical Subject Headings, a także pojęcia *metaterminu*, które odgrywa kluczową rolę w tym procesie. Metatermin jest nazwą dla ogólnej specjalności medycznej lub nauki biologicznej. [Thirion, Darmoni 1999] Metaterminy tworzy zbiór 92 (stan na 05-01-2007 r.) terminów przejmowanych najczęściej ze słownika MeSH. Charakteryzują się one wysokim stopniem ogólności. Zbiór tej kategorii leksykalnej tworzą więc np. terminy: anatomia, kardiologia, chirurgia, chirurgia plastyczna, cytologia, etyka, urologia, pediatria, itd. Każdemu metaterminowi przyporządkowano zbiór tematów MeSH o węższym zakresie znaczeniowym, przy czym jeden temat MeSH może należeć do kilku metaterminów. Reprezentacja wiedzy w systemie zakłada wykorzystanie tematów pochodzących ze słownika MeSH oraz odpowiednich określników rzeczowych stosowanych w tym JIW. Tak zbudowane hasła przedmiotowe są wykorzystywane do budowy punktów dostępu w narzędziach dostępu przedmiotowego. Dodatkowo w tym celu, wykorzystuje się zbiór wyrażen reprezentujących typ obiektów informacyjnych,

który pochodzi z listy wartości dla tej cechy stosowanej w bazie Medline (Medline Resource Type). Zależności strukturalne zachodzące pomiędzy wymienionymi kategoriami leksykalnymi prezentuje Rys. 5.4



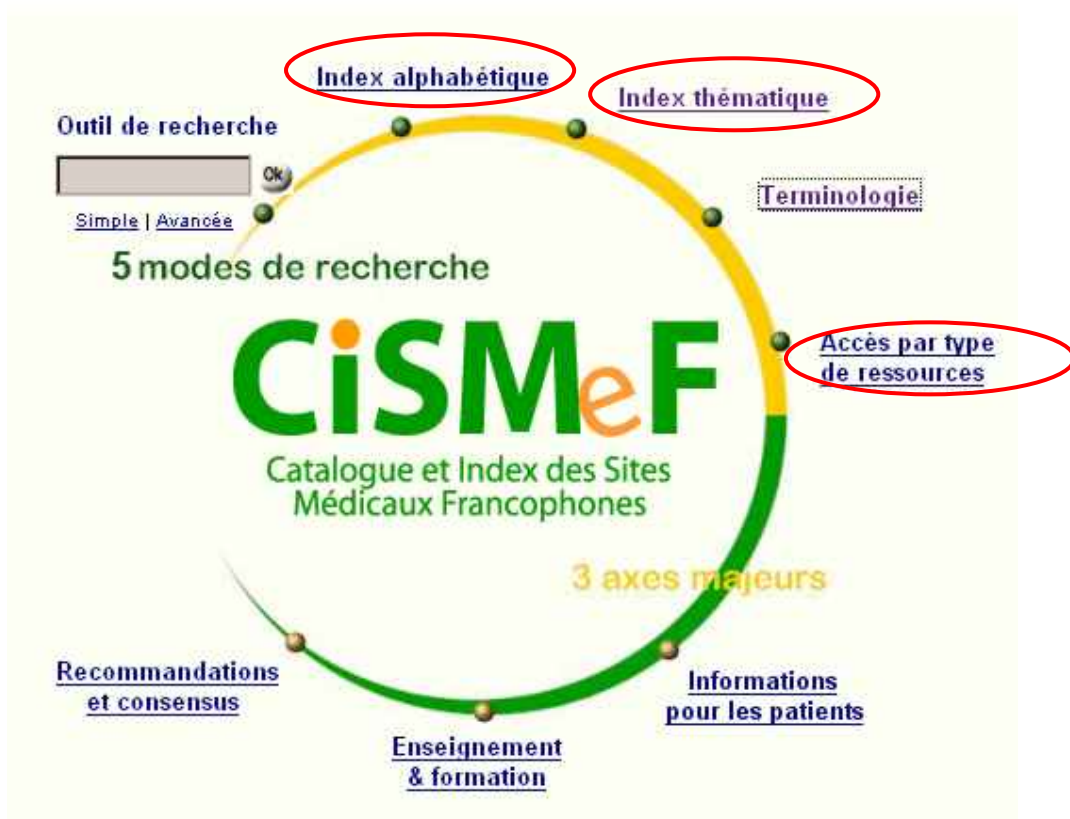
Rys. 5.4 Struktura kategorii punktów dostępu w systemie CISMef. [Źródło: opracowanie własne]

W narzędziu dostępu przedmiotowego CISMef wykorzystuje cztery kategorie punktów dostępu. Są to:

- metaterminy,
- tematy MeSH,
- określniki rzeczowe MeSH (ang. subheadings),
- jednostki leksykalne odwzorowujące cechę *rodzaj źródła* (ang. resource type).

Struktura pola tematycznego w systemie jest dwuwymiarowa [por. Ungurian 1976; Woźniak 2000], co pozwala wyodrębnić wymiar treści oraz formy. Wykładnikami wymiaru treści są metaterminy oraz tematy i określniki rzeczowe MeSH. Wymiar formy reprezentowany jest przez kategorię leksykalną *typ zasobu*.

Dostęp do zasobów systemu umożliwiają trzy katalogi jawne (tj. dostępne na stronie głównej CISMef) oraz dwa niejawne, które dostępne są po aktywowaniu jednego z wcześniej wymienionych. Rys. 5.5 przedstawia punkty dostępu do zbioru informacyjnego w systemie CISMef.



Rys. 5.5 Punkty dostępu do zbioru informacyjnego w CISMef. [Źródło: [www.chu-rouen.fr/cismef/](http://www.chu-rouen.fr/cismef/)]

Pozyskiwanie informacji z zasobu systemu możliwe jest poprzez:

- wyszukiwarę (fr. Outil de recherche),
- katalog alfabetyczny (fr. Index alphabétique),
- katalog tematyczny (fr. Index thématique),
- katalog rodzajów zasobów (fr. Accès par type de ressources).

Katalog alfabetyczny tworzą:

- tematy MeSH,
- określniki rzeczowe MeSH (z etykietą *Qualificatif*),
- wyrażenia odwzorowujące typ obiektów informacyjnych (z etykietą *Type de ressource*).

Wyrażenia tworzące katalog mają postać odpowiedników jednostek leksykalnych MeSH wyrażonych w języku francuskim, do których dołączono ich oryginalną formę w języku angielskim. Każde z haseł wyposażone jest w odpowiednie odesłanie hipertekstowe:

- Aktywacja powiązania dla tematu MeSH powoduje wyświetlenie ChWD zaindeksowanych z jego wykorzystaniem oraz sporządzonego dla niego artykułu hasłowego;

- Aktywacja powiązania dla określnika rzeczowego powoduje wyświetlenie ustrukturyzowanego wykazu określników stosowanych hasłach przedmiotowych w zbiorze wyszukiwawczym systemu. Prezentacja tego typu wyrażeń synsyntaktycznych umożliwia wyświetlenie informacji o zakresie stosowania danego określnika oraz wyodrębnienie podzbioru ChWD, w których wystąpiło jego użycie z dowolnym tematem MeSH;
- Aktywacja powiązania dla hasła odwzorowującego określony typ obiektów informacyjnych powoduje wyświetlenie wykazu ChWD spełniających wskazane kryterium.

Artykuły hasłowe (Rys. 5.6) dla tematów MeSH zbudowane są z:

- tematu MeSH,
- definicji terminu, zasad stosowania (scope note),
- wyrażeń synonimicznych pochodzących z kartoteki MeSH,
- tematów MeSH skojarzonych z wybranym,
- metaterminu (ów), z którym powiązано wybrany temat,
- odsyłacza do miejsca tematu w strukturze hierarchicznej tematów MeSH,

Funkcję wyszukiwawczą pełnią w artykule hasłowym:

- tematy MeSH skojarzone z wybranym wyrazem hasłowym oraz
- metatermin(y), z którym powiązано wybrany temat.

Są one wyposażone w odesłania hipertekstowe, które powodują wyodrębnienie ze zbioru wyszukiwawczego takich ChWD, które spełniają wybrane przez użytkownika kryteria wyszukiwawcze.

**Information :** Un double clic sur un mot permet d'en afficher la définition.

**Définition** [MeSH Scope Note ; traduction CiSMeF] : Spécialité chirurgicale concernant l'étude et le traitement des troubles de l'oreille, du nez et de la gorge.

**Définition** [Fédération Nationale des Centres de Lutte Contre le Cancer] : [http://www.fnclcc.fr/fr/patients/dico/definition.php?id\\_definition=1165](http://www.fnclcc.fr/fr/patients/dico/definition.php?id_definition=1165)

**Synonyme(s) CiSMeF** : *ORL*.

**Synonyme(s) MeSH** : *Oto-rhino-laryngologie ; Otologie ; Otorhinolaryngologie*.

**Voir aussi** : *intervention chirurgicale otologique*.

**Voir aussi le(s) métaterme(s)** : *otorhinolaryngologie*.

**Appartient au(x) Métaterme(s)** : *chirurgie ; otorhinolaryngologie*

**Arborescence(s)** du thesaurus MeSH contenant le mot-clé **otolaryngologie** [otolaryngology] :

► [professions de santé](#)

Position du mot-clé dans l' (les) arborescence(s) :



Rys. 5.6 Artykuł hasłowy w CiSMef. [Źródło: <http://www.chu-rouen.fr/cismef/>]

Pod artykułem hasłowym dla danego terminu wyświetlane są ChWD, w których użyto go do reprezentacji treści. Są one posortowane na grupy charakterystyk według typu obiektu informacyjnego. Dodatkowo użytkownik ma możliwość wyodrębnienia w rezultatach wyszukiwania ChWD według ich przeznaczenia (np. dla pacjentów, o przeznaczeniu edukacyjnym). W ramach prezentacji charakterystyk wyszukiwawczych obiektów informacyjnych opisanych za pomocą danego tematu MeSH, system prezentuje również wykaz określników rzeczowych, które dołączono do eksplorowanego tematu. Każdy z nich posiada odesłanie hipertekstowe, którego aktywacja powoduje wyodrębnienie ChWD, w których pojawił się temat wraz z wybranym określnikiem.

Katalog tematyczny w CiSMef prezentuje w układzie alfabetycznym wyłącznie zbiór metaterminów oraz wyrażen synonimicznych odsyłających do przyjętych form dla tych jednostek leksykalnych. Tutaj również wykorzystanie odesłań hipertekstowych polega na funkcji odsyłania od danego metaterminu do zbioru ChWD zaindeksowanych z jego udziałem.

Katalog rodzaju zasobów jest narzędziem dostępu do zasobów systemu zbudowanym z wyrażen reprezentujących formę obiektów informacyjnych, odwzorowaną podczas indeksowania. Źródło leksyki dla tego zbioru stanowi lista 136 terminów stosowanych w bazie Medline do określania wartości tej cechy. Terminy użyte w ChWD stają się jednocześnie elementami katalogu rodzajów zasobu i tym



samym punktami dostępu do zasobów systemu. Aktywacja odesłania hipertekstowego dla danego terminu powoduje wyodrębnienie w zbiorze informacyjnym podzbioru charakterystyk wyszukiwawczych dokumentów, których forma publikacji odpowiada poszukiwanemu wyrażeniu.

Skorelowanie koncepcji metaterminów z wykorzystaniem języka Medical Subject Headings wydaje się tworzyć efektywne narzędzie dostępu przedmiotowego w tym systemie. Metaterminy pełnią tutaj funkcję tematów ogólnych i jednocześnie klas grupujących punkty dostępu w postaci tematów MeSH. Wyznaczają podstawowe grupy tematyczne, w obrębie których użytkownik może poszukiwać terminów reprezentujących tematy szczegółowe.

### ***5.3. GEM – Gateway to 21st Century Skills***

Misją systemu GEM - Gateway to 21st Century Skills ([www.thegateway.org](http://www.thegateway.org)) jest funkcjonowanie jako internetowego narzędzia rozszerzającego dotychczasowe możliwości opiekunów i nauczycieli w dostępie do cyfrowych materiałów metodycznych i dydaktycznych. Celem GEM jest organizacja dostępu do źródeł informacji opublikowanych w WWW i tworzonych przez instytucje rządowe, akademickie, non-profit, a także komercyjne. W 1996 roku z inicjatywy Wydziału Kształcenia Narodowej Biblioteki Pedagogicznej w USA (National Library of Education) powstało konsorcjum GEM, które opracowało jego zaplecze organizacyjne i programistyczne. Rok później uruchomiono sam system. W skład grupy zarządzającej kolekcją wchodzi reprezentanci instytucji naukowych w USA: Merit Network, Illinois Office of Educational Services, Library of Congress, University of Arizona. [Laundry 2006] Zarząd systemu współpracuje z siecią Achievement Standards Network powołaną przez National Science Foundation, w celu koordynowania standardów nauczania w programach i materiałach dydaktycznych w USA. Dostęp do zasobów systemu jest bezpłatny, jedynie transfer własnych rekordów do bazy wiąże się z ustaleniem kwoty członkostwa instytucjonalnego. System gromadzi około 50000 rekordów zawierających charakterystyki wyszukiwawcze wyselekcjonowanych, cyfrowych obiektów informacyjnych oraz całych kolekcji opublikowanych w Internecie (stan na 01-03-2007 r.). Kryteria włączenia do zbioru wyszukiwawczego GEM są ściśle określone i opublikowane w obszernej dokumentacji systemu.

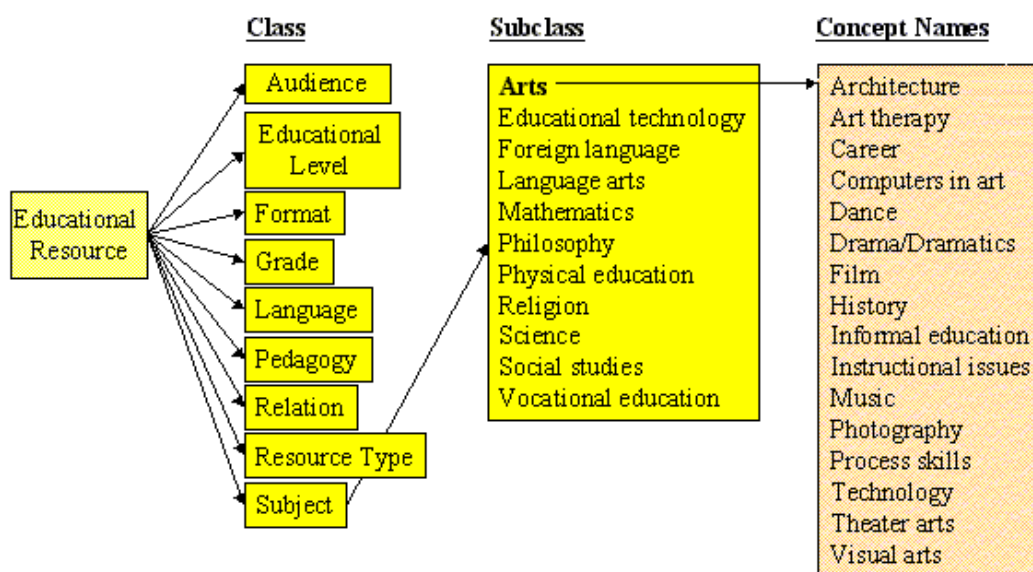
Specyfika indeksowanych w systemie zasobów o charakterze i przeznaczeniu

edukacyjnym ma swoje odzwierciedlenie w stosowanym schemacie metadanych. Format opisu w GEM powstał poprzez zastosowanie zestawu 15 elementów formatu Dublin Core oraz ośmiu elementów dodatkowych. [Sutton 1997, Sutton 1998, Sutton 1999] Służą one do odwzorowania informacji m.in. o grupie użytkowników, do których skierowany jest dokument, przedstawionych metodach kształcenia, zgodności ze standardami nauczania na określonym poziomie. Wyodrębnienie tego rodzaju cech daje dodatkowe możliwości wyszukiwania informacji w systemie.

Budowa narzędzia dostępu przedmiotowego w systemie zakłada sześciowymiarową strukturalizację zbioru wyrażen przyjętych jako punkty dostępu. Są to następujące kategorie:

- typ zasobów,
- poziom kształcenia,
- treść,
- pośrednik,
- beneficjent,
- koszt dostępu.

Dla każdej z nich opracowano w układzie alfabetycznym wykaz wyrażen reprezentujących wartości poszczególnych cechy obiektów informacyjnych. Stanowią one węzły dla powiązań hipertekstowych, których uruchomienie powoduje wyświetlenie listy relewantnych ChWD. Na Rys. 5.7 przedstawiono organizację punktów dostępu w narzędziach dostępu przedmiotowego w GEM.



Rys. 5.7 Organizacja punktów dostępu w GEM. [Źródło: Qin, Palling 2001]

W kategorii *typ zasobów* (ang. Resource Type) wyróżniono 30 wyrażen, które

identyfikują rodzaj internetowego obiektu informacyjnego. Pełną listę wyrażeń prezentuje Tab. 5.7.

Tab. 5.7 Wartość cechy *typ zasobów* w GEM.

[Źródło: <http://www.thegateway.org/help/about/documentation/gem-controlled-vocabularies/vocabulary-resource-type>]

Kategoria	Opis
Aktywność (Activity)	Np. zadanie lub ćwiczenie, jakie ma wykonać uczeń lub instrukcje pomagające w jego ukończeniu
Artefakt (Artifact)	Efekty pracy i umysłu człowieka
Rozwiązania (Best practice)	Proces, praktyka, system, które przyniosły efektywne rezultaty w pracy pedagogicznej lub organizacji jednostki kształcącej
Rekord katalogu (Catalog record)	Rekord opisujący obiekt w inny, bardziej szczegółowy sposób niż GEM
Kolekcja (Collection)	Kolekcja źródeł lub odesłań hipertekstowych do nich
Społeczność (Community)	Np. edukacyjna lista mailingowa, grupy dyskusyjne
Kurs (Course)	Treść kursów wybranych przedmiotów
Program nauczania Curriculum	Program nauczania wraz pełnym wykazem zadań, celów i środków ich realizacji
Program nauczania - materiały pomocnicze (Curriculum support)	Materiały pomocnicze do realizacji programu nauczania
Zbiór danych (Data set)	Źródła zawierające wyłącznie dane numeryczne lub statystyczne
Materiały metodyczne (Educator's guide)	Przewodniki dla nauczycieli
Środowisko (Environment)	Np. multimedia, gry, czat, rzeczywistość wirtualna
Wydarzenie (Event)	
Forma (Form)	Pusty dokument pełniący funkcję szablonu
Zbiór grafik (Image set)	
Plany lekcji (Lesson plan)	Plany pomocne dla uczniów w efektywnym uczestnictwie w lekcjach
Literatura (Literature)	Literatura piękna
Źródło pierwotne (Primary source)	Osobowe, instytucjonalne lub dokumentalne
Projekt (Project)	Np. projekt badawczy
Realia	Reprezentacje obiektów uniwersum, w których można wyróżnić dodatkowe realia (zjawiska, obiekty) niebędące dziełem człowieka, np. zdjęcie drzewa

Informator (Reference)	Informatory bezpośrednie
Wyniki badań (Research study)	Krytyczna analiza wyników badań
Źródła pochodne (Secondary source)	Abstrakty, recenzje, itp.
Wydawnictwo ciągłe (Serial)	
Serwis, usługa (Service)	Usługa dostępna online
Opowiadanie (Story)	
Poradniki dydaktyczne (Study guide)	
Test (Test)	
Narzędzie (Tool)	Urządzenie, narzędzie dostępne poprzez Internet
Jednostka instruktażowa (Unit of instruction)	Sekwencja planu zajęć, stworzona w celu opanowania umiejętności, przyswojenia informacji

Kategoria *Poziom kształcenia* (ang. Level) zawiera wyrażenia przyjęte do określania poziomu edukacyjnego obiektu informacyjnego. Wyróżniono w niej dwie podkategorie:

- Stopień kształcenia (Grade), zawierającą wartości tej cechy w postaci: *przedszkole, klasa 1, 2, 3, 4, ... 12*;
- Poziom edukacyjny, gdzie wyodrębniono:
  - o Dla wszystkich (All),
  - o Niesprecyzowany (Unspecified),
  - o Kształcenie dorosłych (Adult/continuing education),
  - o Szkoła pomaturalna (Community College),
  - o Szkolnictwo wyższe (Higher education),
  - o Edukacja przedszkolna (Preschool education),
  - o Szkoły zawodowe (Vocational education).

Kolejnym wymiarem, zwanym w dokumentacji systemu również *fasetą* [Laundry 2006], jest wymiar cech treści obiektów informacyjnych. Tworzą ją dwa indeksy punktów dostępu - przedmiotowy (ang. subject) oraz słów kluczowych (ang. keywords). Indeks przedmiotowy tworzy 12 alfabetycznie uporządkowanych kategorii tematycznych, których nazwy odpowiadają wyrażen przedmiotom nauczania lub grupom zagadnieniowym. Są to: sztuka, technologia edukacyjna, języki obce, zdrowie, językoznawstwo i literatura, matematyka, filozofia, kultura fizyczna, religia, nauki ścisłe, nauki społeczne, przedmioty zawodowe. Jest to zbiór deskryptorów tezauryasa

Thesaurus of ERIC Descriptors uzupełniony o niekontrolowane słowa kluczowe\*. W indeksie niekontrolowanych słów kluczowych występują wyrażenia bezpośrednio odwzorowujące temat i przedmiot obiektu informacyjnego.

Faseta *Pośrednik* (ang. Mediator), grupuje wyrażenia odwzorowujące nazwy grup lub stanowisk pośredniczących w procesie edukacji pomiędzy użytkownikiem a edukacyjnym obiektem informacyjnym. Wyodrębniono tutaj 18 wyrażeń:

1. Administratorzy (ang. Administrators),
2. Instruktorzy w Collegach/ Uniwersytetach (ang. College/ University Instructors),
3. Zarządzający wdrażaniem programów zajęć (ang. Curriculum Supervisors),
4. Nauczyciele szkół podstawowych (ang. Elementary School Teachers),
5. Społeczeństwo – ogólnie (ang. General Public),
6. Bibliotekarze (ang. Librarians),
7. Nauczyciele szkół średnich (ang. Middle School Teachers),
8. Inne zawody związane z edukacją (ang. Other Educational Professionals),
9. Rodzice (ang. Parents),
10. Osoby odpowiedzialne za politykę kształcenia (ang. Policy-Makers),
11. Nauczyciele szkół drugiego stopnia (ang. Secondary School Teachers),
12. Nauczyciele akademicy (ang. Student Teachers),
13. Uczniowie (ang. Students),
14. Inspektorzy szkolni (ang. Supervisors),
15. Metodocy (ang. Teacher Educators),
16. Nauczyciele (ang. Teachers),
17. Specjaliści prowadzący szkolenia (ang. Teaching Professionals),
18. Koordynatorzy wdrażania technologii (ang. Technology Coordinators).

Faseta *Beneficjent* (ang. Beneficiary) grupuje wyrażenia pochodzące z listy słownictwa kontrolowanego GEM stosowane do określania kategorii osób lub grup użytkowników dla których przeznaczone jest źródło internetowe. W tym celu wyodrębniono 47 jednostek leksykalnych odwzorowujących tę cechę.

Ostatnią z faset organizujących dostęp do zasobów systemu jest *Koszt dostępu* (ang. Price Code), która zawiera trzy wyrażenia: *Darmowy* (ang. Free), *Częściowo darmowy* (ang. Partially free), *Odpłatny* (ang. Not free).

---

\* <http://www.thegateway.org/help/about/documentation/gem-controlled-vocabularies/vocabulary-subject>

## Gateway to 21<sup>st</sup> Century Skills

start learning & teaching leading & managing partnering asn about help

jesteś w: start

nawigacja

- Start
- Learning & Teaching
- Leading & Managing
- Partnering
- ASN
- About
- Help

portal\_seamarksearch

**Browse by subject** (glossary) show ▼

[all](#) (362) | [arts](#) (4977) | [educational technology](#) (3198) | [foreign languages](#) (870) | [health](#) (3707) | [language arts](#) (8287) | [mathematics](#) (5856) | [philosophy](#) (343) | [physical education](#) (1810) | [religion](#) (526) | [science](#) (12161) | [social studies](#) (13362) | [vocational education](#) (990) |

**Browse by type** (glossary) show ▼

**Browse by level** (glossary) hide ▲


[1](#) (9493) | [10](#) (16532) | [11](#) (16584) | [12](#) (16493) | [2](#) (9982) | [3](#) (11997) | [4](#) (13547) | [5](#) (14702) | [6](#) (15920) | [7](#) (16239) | [8](#) (16476) | [9](#) (16956) | [adult/continuing education](#) (2152) | [all](#) (1851) | [community college](#) (1630) | [higher education](#) (4563) | [kindergarten](#) (6930) | [preschool education](#) (1504) | [unspecified](#) (3182) | [vocational education](#) (424) |

**Browse by keywords** (glossary) show ▼

**Browse by mediator** (glossary) show ▼

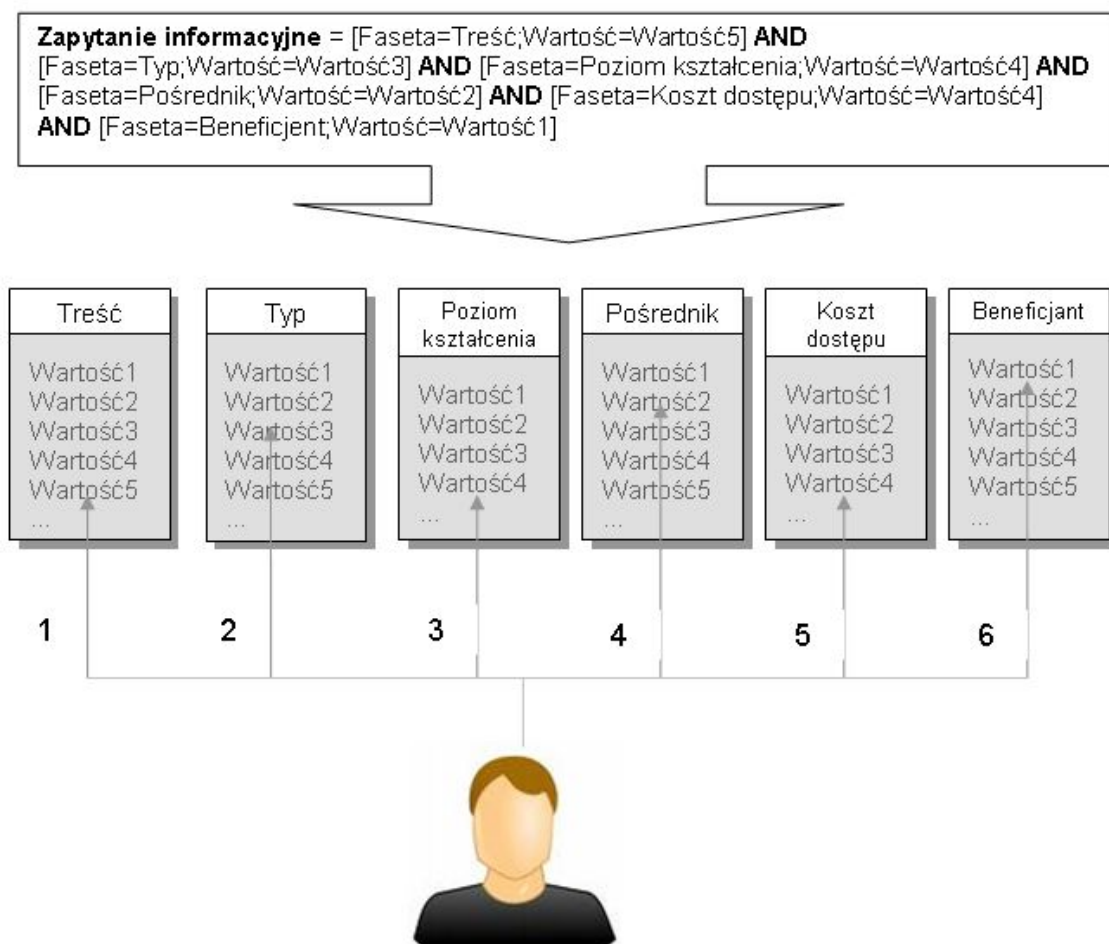
**Browse by beneficiary** (glossary) show ▼

**Browse by priceCode** (glossary) show ▼



Rys. 5.8. Narzędzie dostępu przedmiotowego w systemie GEM. [Źródło: <http://thegateway.org>]

Dostęp do zasobów systemu za pomocą tak skonstruowanego narzędzia polega na wyborze z określonej faset wyrażenia jako punktu dostępu. Jego rodzaj oraz wartość reprezentuje jedną z wymienionych cech obiektów informacyjnych. Każde wyrażenie jest węzłem dla odesłania hipertekstowego, którego uruchomienie powoduje wyodrębnienie i wyświetlenie ChWD, w których w opisie go zastosowano. Wybór wyrażenia z określonej fasety powoduje wyświetlenie relewantnych ChWD, ale również wyrażen z pozostałych siedmiu faset zaproponowanych przez system jako kryteria zawężenia zakresu zapytania informacyjnego. System wyświetla jednak tylko takie terminy w pozostałych fasetach, które zastosowano do indeksowania w ChWD wraz z wyrażeniem wybranym jako pierwsze. Tym samym wybór kolejnych wyrażen z faset prowadzi do dynamicznego zawężania zakresu zapytania poprzez dołączanie do kryteriów wyszukiwawczych wyrażen z konkretnych faset. Przebieg tego procesu przedstawia rysunek 5.9.



Rys. 5.9. Hybrydowe narzędzie dostępu przedmiotowego w systemie GEM.

[Źródło: opracowanie własne]

Narzędzie dostępu przedmiotowego w systemie GEM reprezentuje hybrydowy wariant organizacji punktów dostępu. Siedem różnych alfabetycznych indeksów gromadzących wyrażenia reprezentujące sześć kategorii atrybutów zostało połączonych w jedno narzędzie dostępu do zasobów. Jego koncepcja została opracowana na podstawie zaprojektowanej dla systemu ontologii internetowej. [Qin, Palling 2004] Opierając się na słownictwie kontrolowanym zidentyfikowano edukacyjny obiekt informacyjny poprzez wyodrębnienie w nim cech uznanych za istotne z punktu widzenia wyszukiwania informacji. To pozwoliło na odpowiednią konstrukcję formatu opisu danych oraz na wykorzystanie hipertekstu do dynamicznego przetwarzania zapytania informacyjnego i reorganizacji zbioru rezultatów wyszukiwania.

#### 5.4. Intute

System Intute ([www.intute.ac.uk](http://www.intute.ac.uk)) uruchomiono 13 sierpnia 2006 roku. (<http://www.intute.ac.uk/news/71.html>) Jest on jednym z największych dziedzinowych

systemów hipertekstowych. Jego geneza związana jest pośrednio z projektem eLib, którego efekty prac przyczyniły się do powstania pierwszej sieci DSH. Był to brytyjski Research Discovery Network (RDN). [Hiom 2006c] Celem jego twórców była koordynacja prac nad wdrażaniem nowych systemów oraz opieka nad już funkcjonującymi. RDN zrzeszało osiem dziedzinowych systemów hipertekstowych. Były to:

- Artifact – sztuka oraz tzw. kreatywne gałęzie gospodarki (sztuka, architektura, media, muzyka, moda, itp.),
- EEVL – inżynieria, matematyka, informatyka,
- GeSource – geografia,
- BIOME – medycyna i nauka o zdrowiu,
- ALTIS – sport, turystyka, spędzanie wolnego czasu,
- Humbul – nauki humanistyczne,
- PSIGate – astronomia, chemia, nauka o Ziemi, materiałoznawstwo,
- SOSIG – nauki społeczne.

Zmiany w kierunkach projektowania sieciowych systemów informacyjnych funkcjonujących w środowisku WWW objęły również dziedzinowe systemy hipertekstowe, w tym RDN. Przeformułowano więc zasady funkcjonowania sieci i skierowano jej rozwój w stronę modelu zintegrowanej platformy zaspokajającej potrzeby informacyjne wybranej grupy użytkowników. Uruchomienie systemu Intute było odpowiedzią na potrzeby informacyjne środowiska akademickiego Wielkiej Brytanii. System Intute gromadzi informacje o ponad 100000 obiektów informacyjnych w zasobach WWW (stan na 05-11-2006 r.). Jest on zbudowany z czterech podsystemów, których zakresy wyznaczają cztery grupy tematyczne:

- nauka i technologia – Intute: Science, Engineering & Technology (<http://www.intute.ac.uk/sciences/>),
- sztuka i nauki humanistyczne – Intute: Arts & Humanities (<http://www.intute.ac.uk/artsandhumanities/>),
- nauki społeczne – Intute: Social Sciences (<http://www.intute.ac.uk/socialsciences/>),
- medycyna i nauka o zdrowiu – Intute: Medicine & Health Sciences (<http://www.intute.ac.uk/healthandlifesciences/>).

Pracę systemu koordynuje konsorcjum, w skład którego wchodzi przedstawiciele siedmiu uniwersytetów z Wielkiej Brytanii. Finansowaniem systemu zajmuje się rada zrzeszająca przedstawicieli Joint Information Systems Committee (JISC), Arts and



Humanities Research Council (AHRC) oraz Economic and Social Research Council (ESRC).

System Intute skierowany jest do użytkowników wywodzących się z sektora szkolnictwa wyższego w Wielkiej Brytanii (HE – ang. higher education) oraz kształcenia na poziomie poniżej szkolnictwa wyższego dla osób, które przekroczyły wiek obowiązku szkolnego (FE – ang. further education). Istota tego systemu, polega na pełnieniu przez niego funkcji repozytorium metadanych dla wirtualnych środowisk edukacyjnych (VLE – ang. virtual learning environments).

Istotną cechą systemu jest wykorzystanie standaryzowanych schematów metadanych do odwzorowania treści i formy obiektów informacyjnych. Struktura rekordów bazy danych zakłada funkcjonowanie elementów różnych schematów metadanych. Ich celem jest efektywna charakterystyka źródła internetowego. Format opisu zastosowany w systemie Intute opracowano na podstawie:

- Dublin Core,
- RLLOMAP – profil aplikacyjny RDN/LTSN Learning Object Metadata,
- AACR2 – Anglo American Cataloging Guidelines, wydanie drugie,
- RDA – Resource Description and Access – nowe zasady katalogowania, które w założeniu mają zastąpić AACR2, i które uwzględniają specyfikę cyfrowych źródeł informacji. [Champan 2006; Knowlton 2006]

Charakterystyka elementów zbioru informacyjnego jako potencjalnych obiektów edukacyjnych (LOM – ang. learning object metadata) [Najjar, Duval 2006] zakłada wyodrębnianie cech wyszukiwawczych istotnych z dydaktycznego punktu widzenia. W przypadku systemu Intute określono kilka dodatkowych cech dokumentu, którym podczas indeksowania przypisuje się wartości pobierane z list kontrolowanych. Są to m.in.:

*Audience* - określenie grupy użytkowników docelowych,

*Educational Level* - określenie poziomu dydaktycznego obiektu informacyjnego.

Kryteria selekcji stosowane w systemie Intute oparte są na zaleceniach projektu DESIRE [Hofman, Worsfold 1997] dotyczących jakościowych wskaźników oceny źródeł internetowych włączanych do zbiorów informacyjnych DSH. [Hiom 2006a] Szczegółowy przebieg oceny źródła dostępny jest w dokumentacji systemu. [Hiom 2006a; Hiom 2006b]

Dostęp do zasobów systemu realizowany jest poprzez mechanizm wyszukiwawczy w formie wyszukiwarki oraz narzędzia dostępu przedmiotowego. W pierwszym przypadku, użytkownik może przeszukiwać zasoby całego systemu lub

konkretnej grupy tematycznej. W drugim, dla każdej grupy tematycznej przewidziano osobne, autonomiczne narzędzia dostępu przedmiotowego. Dalsza charakterystyka narzędzi dostępu przedmiotowego w Intute zostanie przedstawiona na wybranych przykładach podsystemów działających w sieci.

#### **5.4.1. Intute: nauka, inżynieria i technologia (*Intute: Science, Engineering & Technology*)**

Zakres tej grupy tematycznej obejmuje nauki przyrodnicze, inżynierię i technologię. Grupę tematyczną stworzono poprzez połączenie zbiorów wyszukiwawczych trzech dziedzinowych systemów hipertekstowych:

- EEVL – inżynieria, matematyka, technika komputerowa,
- GEsorce – geografia, nauka o środowisku,
- PSI-Gate – fizyka, astronomia, chemia, nauka o Ziemi, materiałoznawstwo.

W omawianej grupie tematycznej zgromadzono 33452 rekordów (stan na 30-08-2006 r.). Narzędzie dostępu przedmiotowego zbudowano z zagnieżdżonych w strukturze hipertekstowej kategorii tematycznych w układzie hierarchicznym. Na podstawie podziału dyscyplinarnego na pierwszym stopniu podziału wyodrębniono dziesięć kategorii tematycznych. Są to:

1. Astronomia,
2. Chemia,
3. Technika komputerowa,
4. Nauki o Ziemi,
5. Inżynieria,
6. Nauka o środowisku,
7. Nauka - ogólnie,
8. Geografia,
9. Matematyka,
10. Fizyka.

Każda z wymienionych kategorii tematycznych podlega dalszemu podziałowi. Organizację kategorii *Matematyka*, *Inżynieria* oraz *Technika komputerowa* przejęto z narzędzia dostępu przedmiotowego systemu EEVL. Jako nazwy klas zastosowano tutaj:

- odpowiedniki słowne klas Mathematics Subject Classification (MSC 2000) - *Matematyka*,
- deskryptory tezaurusu Ei Thesaurus – *Inżynieria* oraz
- odpowiedniki słowne symboli klasyfikacyjnych ACM Computing Classification System (1998) – *Technika komputerowa*.

Tylko w przypadku kategorii *Technika komputerowa* zastosowano oryginalny porządek pionowy klas, jaki ma miejsce w klasyfikacji ACM Computing Classification System. W zakresie kategorii *Technika komputerowa* wyodrębniono kategorie podrzędne, do trzeciego stopnia podziału, zgodnie z podziałem uniwersum, jaki ma miejsce ACM Computing Classification System. W pozostałych przypadkach zastosowano stworzone na potrzeby systemu schematy hierarchiczne wykorzystujące jedynie leksykę wymienionych JIW jako nazwy punktów dostępu grupujących ChWD.

Nazwy współrzędnych kategorii tematycznych (otrzymanych z podziału tej samej klasy węzłowej) są uporządkowane w kolejności alfabetycznej. Wyjątek stanowi tutaj organizacja kategorii tematycznej *Technika komputerowa*, gdzie zastosowano porządek klas ACM Computing Classification System.

Kategorie tematyczne *Astronomia*, *Chemia*, *Nauka o Ziemi*, *Geografia*, *Materialoznawstwo* oraz *Fizyka* poddano dalszemu podziałowi wyodrębniając kategorie podrzędne, których nazwy są m.in. odpowiednikami słownymi określonych symboli klasyfikacyjnych Klasyfikacji Dziesiątej Deweya. W Tab. 5.8 przedstawiono fragment organizacji kategorii tematycznej *Chemia* oraz odpowiadające wyodrębnionym podkategoriom symbole klasyfikacji KDD. Obok nazwy każdej kategorii tematycznej w nawiasie kwadratowym przedstawiono liczbę zgrupowanych w nich ChWD. Jest to charakterystyczne dla narzędzi dostępu przedmiotowego w całym systemie Intute.

Tab. 5.8 Organizacja kategorii tematycznej *Chemia* w systemie Intute. [Źródło: opracowanie własne]

Nazwa klasy w narzędziu dostępu	Symbol KDD skojarzony z kategorią tematyczną w systemie
<b>Chemistry [4187]</b>	
- Analytical chemistry [493]	- 543
o Chromatography [58]	- 543.089
o Mass Spectrometry [56]	- 543.0873
o Other techniques [87]	- 543.08
o Spectroscopy [197]	- 543.0858
- Chemicals and materials safety [67]	- 363.17
- Chemistry of materials and environments [306]	- 540
o Atmospheric chemistry [56]	- 551.511
o Chemical oceanography [27]	- 551.4601
o Food chemistry [38]	- 664
o Geochemistry [122]	- 551.9
▪ Soil chemistry [19]	- 631.41
- Computational chemistry [153]	- 540.72
- Crystallography [164]	- 548
- General Chemistry [539]	- 540
o Chemistry dictionaries and encyclopedias [69]	- 540.03
o Chemistry news [10]	- 540.05
o General chemistry animations and multimedia [16]	- 540.0208
o General chemistry databases [25]	- 540.021
o General chemistry journals [23]	- 540.05
o General chemistry software [10]	- 540.0285
o General chemistry Web directories and portals [32]	- 540.025
o Laboratory chemistry	- 542

Analiza danych przedstawionych w Tab. 5.8 wskazuje na brak pełnej zgodności schematów kategorii tematycznych w systemie z porządkiem klas w KDD. W narzędziu dostępu przedmiotowego na tym samym stopniu podziału występują kategorie tematyczne, których odpowiedniki w postaci symboli KDD reprezentują klasy niewspółrzędne. Prawdopodobnie w celu uniknięcia zbyt głębokiego podziału w schemacie kategorii tematycznych zdecydowano się na taki zabieg.

Określenie, jaki symbol KDD odpowiada kategorii tematycznej w analizowanym narzędziu było możliwe w tym przypadku dzięki specyficznej budowie odesłania hipertekstowego dla każdego wyrażenia. Wspomniane kategorie tematyczne, w których zastosowano podział na podstawie KDD tworzyły przed utworzeniem Intute narzędzie dostępu przedmiotowego w systemach GEsources i PSI-

Gate\*. Każda kategoria tematyczna wyposażona była wówczas w odesłanie hipertekstowe (generujące zbiór rezultatów wyszukiwania), które zawierało odpowiadający jej symbol z klasyfikacji KKD. W Tab. 5.9 przedstawiono przykładowe zagnieżdżenie symbolu KDD w odesłaniu hipertekstowym.

Tab. 5.9 Zagnieżdżanie symbolu KDD w strukturze odesłania hipertekstowego.

[Źródło: opracowanie własne]

Nazwa klasy	Forma odesłania hipertekstowego	Ekstrahowany symbol KDD
<a href="#">Mass Spectrometry</a>	http://www.psigate.ac.uk/roads/cgi-bin/psibrowse.pl?limit=0&tolevel=chemistry&subject= <b>543.0873</b>	543.0873
<a href="#">Chemical oceanography</a>	http://www.psigate.ac.uk/roads/cgi-bin/psibrowse.pl?limit=0&tolevel=chemistry&subject= <b>551.4601</b>	551.4601

Dzięki zagnieżdżeniu symbolu KDD w odesłaniu hipertekstowym możliwe było odwzorowanie struktury kategorii tematycznych i odpowiadających im symboli KKD oraz odkrycie, czy są one porównywalne.

Narzędzia dostępu przedmiotowego w charakteryzowanej grupie tematycznej systemu Intute zostały zbudowane z wykorzystaniem czterech języków informacyjno-wyszukiwawczych (Tab. 5.10).

Tab. 5.10 JIW jako narzędzie organizacji informacji w grupie tematycznej Intute: nauka, inżynieria i technologia. [Źródło: opracowanie własne]

Klasa	JIW który wykorzystano do organizacji informacji
Matematyka	Mathematical Subject Classification
Inżynieria	Ei Thesaurus
Technika komputerowa	ACM Computing Classification System
Astronomia Chemia Nauki o Ziemi Nauki ogólne Geografia Materiałoznawstwo Matematyka	Klasyfikacja Dziesiątka Deweya

Wskazane JIW stanowią przede wszystkim źródło słownictwa wykorzystywanego jako nazwy wyodrębnianych kategorii tematycznych. Porządkowanie zbioru kategorii tematycznych przyjmuje formę zgodną do określonego stopnia z odpowiadającymi im układami klas w KDD i ACM Computing Classification System. W pozostałych przypadkach są to tworzone na potrzeby narzędzia struktury hierarchiczne wykorzystujące jedynie leksykę JIW jako nazwy dla kategorii tematycznych.

Każda jednostka leksykalna przyjęta za nazwę kategorii tematycznej

\* Do Intute przejęto oryginalny układ tych kategorii tematycznych.

w strukturze hierarchicznej jest jednocześnie węzłem dla powiązania hipertekstowego. Jego aktywacja powoduje wyświetlenie przez system:

- listy ChWD, w których użyto ją do odwzorowania treści i/lub formy jej użycia,
- listy wyrażen pozostajacych z wyodrębnionym w hierarchicznej relacji bezpośredniej podrzędności zakresowej (Rys. 5.10).

The screenshot shows a search interface on the Intute website. At the top, there is a breadcrumb trail: [Astronomy](#) > **Observational Astronomy**. Below this, the section is titled "Sub-categories" and lists several links with their respective record counts: [Gamma Ray Astronomy](#) [15], [Infrared Astronomy](#) [16], [Optical Astronomy](#) [29], [Radio Astronomy](#) [14], [Sky Surveys](#) [40], and [X-ray Astronomy](#) [31]. A search bar is present with the text "Search" and a dropdown menu set to "Observational Astronomy". Below the search bar, there are links for "Advanced search", "Subject A-Z", "New resources", and "Help". A filter bar shows "Filter by All resource types". Below the filter bar, there are buttons for "Details", "Full record", "Go to website", and "Save Record". The results section shows "Now showing: 1 - 25 of 110 records (Filter: All resource types)" and "Page: 1 2 3 4 5". The first result is titled "20th century exploration : journey through the galaxy" and includes a brief description of the content, which discusses space exploration in the 20th century, mentioning figures like Hubble, Hawking, and Einstein, and various missions. The URL for the resource is <http://filer.case.edu/~sjr16/advanced/20th.html>.

Rys. 5.10 Artykuł hasłowy w grupie tematycznej Intute: nauka, inżynieria i technologia. [Źródło: <http://intute.ac.uk/>]

Aktywacja powiązania hipertekstowego powoduje tym samym wyświetlenie artykułu hasłowego przewidzianego dla określonego wyrażenia. Jego budowa zakłada:

- wyodrębnienie i wyróżnienie hasła w postaci wyrażenia przyjętego za nazwę kategorii tematycznej: w analizowanym przypadku jest to *Astronomia obserwacyjna* (ang. *Observational astronomy*),
- wskazanie wyrażenia lub wyrażen nadrzędnych (stanowiących ogniwa łańcucha, którego klasą końcową jest aktualnie wyróżniona) za pomocą wskaźnika miejsca: *Astronomia* > *Astronomia obserwacyjna* (ang. *Astronomy* > *Observational Astronomy*),
- identyfikację wyrażen pozostajacych z wyróżnionym w hierarchicznej relacji podrzędności zakresowej: wyrażenia następujące po wykładniku relacji *Podkategorie* (ang. *Sub-categories*).

Każde z wyrażen podrzędnych wskazanych w artykule hasłowym tworzy również węzeł dla powiązania hipertekstowego, który umożliwia dalszą eksplorację zbioru wyszukiwawczego systemu.

W badaniach ilościowych, narzędzie dostępu przedmiotowego w Intute: nauki przyrodnicze, inżynieria i technologia zostało potraktowane jako jeden spójny układ kategorii tematycznych. Pomimo różnych źródeł przejmowania nazw dla kategorii oraz odmiennych sposobów ich porządkowania, badaniom statystycznym poddano cały zbiór wyrażen pełniących te funkcje.

Dostęp do zbioru wszystkich ChWD został tutaj zorganizowany z wykorzystaniem 1176 kategorii tematycznych w sześciopoziomowej strukturze hierarchicznej. Tab. 5.11 przedstawia rozkład kategorii tematycznych na poszczególnych stopniach podziału.

Tab. 5.11 Rozkład kategorii tematycznych w grupie Intute: nauka, inżynieria i technika.

[Źródło: opracowanie własne]

Poziom	Liczba kategorii tematycznych	Procentowy rozkład kategorii tematycznych
Poziom 1	10	0,85%
Poziom 2	120	10,20%
Poziom 3	485	41,24%
Poziom 4	417	35,46%
Poziom 5	120	10,20%
Poziom 6	24	2,05%
Suma	1176	100%

Analiza danych w tabeli 5.11 wskazuje, że ponad 76% łącznej liczby wyodrębnionych kategorii zostało skupionych na trzecim i czwartym stopniu podziału. Biorąc pod uwagę liczbę kategorii tematycznych wyodrębnionych na poziomach od drugiego do czwartego, stanowi to ponad 86% łącznej liczby zidentyfikowanych kategorii.

Analiza budowy narzędzi dostępu przedmiotowego pod kątem dystrybucji odesłań na poszczególnych stopniach podziału oraz średniej pojemności kategorii tematycznej została przedstawiona w tabelach 5.12 oraz 5.13.

Tab. 5.12. Ilościowy rozkład odesłań na poszczególnych poziomach w grupie *Intute: nauka, inżynieria i technologia*. [Źródło: opracowanie własne]

Poziom	Liczba odesłań	Udział procentowy
Poziom 1	0	0%
Poziom 2	12429	19,05%
Poziom 3	27714	42,46%
Poziom 4	17025	26,08%
Poziom 5	6835	10,47%

Poziom 6	1266	1,94%
Suma	65269	100%

Przedstawione w Tab. 5.12 i 5.13 dane wskazują na bezpośrednią zależność pomiędzy liczbą wyodrębnionych kategorii tematycznych na poszczególnych stopniach podziału, a liczbą ChWD przyporządkowanych do nich. W tym przypadku około 68% elementów zbioru wyszukiwawczego została sklasyfikowana z wykorzystaniem kategorii tematycznych umiejscowionych na trzecim i czwartym stopniu podziału. W kategoriach wyodrębnionych na poziomie od drugiego do czwartego zgromadzono łącznie około 87% wszystkich odesłań do ChWD.

Tab. 5.13. Średnia pojemność klas w Intute: nauka, inżynieria i technologia.

[Źródło: opracowanie własne]

Poziom	Średnia pojemność kategorii tematycznej (średnia liczba odesłań do ChWD)
Poziom 1	0
Poziom 2	103
Poziom 3	57
Poziom 4	41
Poziom 5	57
Poziom 6	53

Kategorie wyodrębnione na drugim stopniu podziału charakteryzują się relatywnie największą pojemnością. Na jedną, na tym stopniu podziału przypada średnio 103 ChWD. Na niższych stopniach wskaźnik ten utrzymuje się na poziomie od 41 do 57 ChWD przyporządkowanych do jednej kategorii tematycznej. Zbiór wyszukiwawczy omawianej grupy tematycznej liczył podczas badania 33126 rekordów (stan na 19-12-2006 r.). Do jego organizacji użyto 65269 odesłań do ChWD. Oznacza to, że dany internetowy obiekt informacyjny został przyporządkowany średnio do dwóch kategorii tematycznych.

#### 5.4.2. Intute: nauki społeczne (*Intute: Social Sciences*)

Grupa tematyczna Intute: nauki społeczne powstała poprzez połączenie zbiorów informacyjnych systemów SOSIG oraz Altis. System SOSIG był przed utworzeniem Intute jednym z najważniejszych i największych (pod względem wielkości zbioru wyszukiwawczego) dziedzinowych systemów hipertekstowych. SOSIG powstał w 1994 roku i był rozwijany w ramach projektu eLib Access to Network Resources. [Hiom 2000] Był on postrzegany jako modelowe zastosowanie DSH. Zakres tematyczny



systemu obejmował szeroko rozumiane nauki społeczne, w tym m.in.: socjologię, prawo, edukację, ekonomię oraz psychologię. Zakres systemu Altis, który stanowi drugi człon omawianej grupy tematycznej obejmował sport, turystykę oraz zagadnienia spędzania czasu wolnego.

Narzędzie dostępu przedmiotowego tworzy tutaj również hierarchiczny układ kategorii tematycznych. Do 2000 roku system SOSIG wykorzystywał Uniwersalną Klasyfikację Dziesiątą jako źródło porządkowania kategorii tematycznych. [Koch 1997] Wówczas nazwy kategorii tematycznych tworzyły odpowiedniki słowne 161 symboli UKD. Ich zakres obejmował zarówno tematy ogólne, np. *1 – Filozofia* oraz bardzo szczegółowe, np. *658.562 – Zarządzanie jakością*. [Huxley, Joyce 2004; Slavic 2006] Wraz z rozwojem systemu, a w końcu powstaniem grupy tematycznej Intute: nauki społeczne zreorganizowano dotychczasową strukturę narzędzia. Podczas projektowania układu wykorzystano słowniki wybranych JIW, stosując zawarte tam wyrażenia jako nazwy kategorii grupujących ChWD. Są to:

- tezaurus HASET (Humanities And Social Sciences Electronic Thesaurus),
- słownik języka słów kluczowych International Bibliography of the Social Sciences (IBSS),
- tezaurus Social Care Institute for Excellence Thesaurus (SCIE),
- tezaurus CAB,
- tezaurus Social Research Methodology (SRM-Thesaurus).

Na pierwszym stopniu podziału w strukturze hierarchicznej wyodrębniono 19 kategorii tematycznych:

1. Antropologia (ang. Anthropology),
2. Biznes i zarządzanie (ang. Business and Management),
3. Ekonomia (ang. Economics),
4. Edukacja (ang. Education),
5. Nauka o środowisku (ang. Environmental Sciences),
6. Studia europejskie (ang. European Studies),
7. Ustrój polityczny (ang. Government Policy),
8. Gastronomia i hotelarstwo (ang. Hospitality and Catering),
9. Antropogeografia (ang. Human Geography),
10. Prawo (ang. Law),
11. Polityka (ang. Politics),
12. Psychologia (ang. Psychology),
13. Narzędzia i metody badawcze (ang. Research Tools and Methods),

14. Opieka społeczna (ang. Social Welfare),
15. Socjologia (ang. Sociology),
16. Sport i spędzanie czasu wolnego (ang. Sport and Leisure Practice),
17. Statystyka (ang. Statistics and Data),
18. Podróże i turystyka (ang. Travel and Tourism),
19. Studia nad problematyką kobiet (ang. Women's Studies).

Kategorie te, chociaż zostały wyodrębnione na najwyższym stopniu podziału, cechuje różny stopień szczegółowości. Dokonano tutaj kategoryzacji uniwersum z wykorzystaniem podziału dyscyplinarnego. Wyjątkiem jest tutaj kategoria *Narzędzia i metody badawcze*, której zakres obejmuje obiekty informacyjne, których treść dotyczy metod i narzędzi badawczych wykorzystywanych w naukach społecznych, a więc zagadnień, które są wspólne dla obszarów wiedzy reprezentowanych przez nazwy pozostałych kategorii.

Ilościowy rozkład kategorii tematycznych na poszczególnych poziomach podziału został przedstawiony w Tab. 5.14.

Tab. 5.14. Rozkład kategorii w grupie *Intute: nauki społeczne*. [Źródło: opracowanie własne]

Poziom	Liczba kategorii	Procentowy rozkład kategorii
Poziom 1	19	1,05%
Poziom 2	307	17,12%
Poziom 3	723	40,32%
Poziom 4	672	37,47%
Poziom 5	59	3,29%
Poziom 6	13	0,75%
Suma	1793	100,00%

Narzędzie dostępu przedmiotowego w tej grupie tematycznej tworzy szczęściopoziomowy układ hierarchiczny kategorii tematycznych. Ponad 94% z 1793 kategorii tworzących tę strukturę, zostało wyodrębnionych na poziomach od drugiego do czwartego. Na poziomach trzecim i czwartym zidentyfikowano ponad 78% łącznej liczby kategorii.

Podstawową relacją semantyczną wyznaczającą porządek kategorii tematycznych jest relacja hierarchii tematycznej. Dużą rolę odgrywa tutaj również sieć relacji asocjacyjnych. Podczas ekstrakcji struktury hierarchicznej, w większości analizowanych przypadków, każdej wyodrębnionej kategorii tematycznej przyporządkowano przynajmniej jedną kategorię powiązaną z nią za pomocą niespecyfikowanej relacji asocjacyjnej. Tym samym aktywacja powiązania hipertekstowego, przewidzianego dla wyrażenia stanowiącego nazwę dla danej

kategorii prowadzi do wyświetlenia użytkownikowi (Rys. 5.11):

- wyrażen stanowiących ogniwa łańcucha, którego kategorią końcową jest aktualnie wyróżniona (za pomocą wskaźnika miejsca),
- listy przyporządkowanych ChWD,
- listy kategorii tematycznych powiązanych z aktualnie eksplorowaną relacją bezpośredniej podrzędności zakresowej,
- listy kategorii tematycznych powiązanych z aktualnie eksplorowaną relacją asocjacyjną.

Psychology > Forensic Psychology and Legal Issues

**Sub-categories**

- [Civil Rights and Civil Law](#) [2]
- [Crime Prevention](#) [1]
- [Criminal Law and Criminal Adjudication](#) [5]
- [Mediation and Conflict Resolution](#) [0]
- [Police and Legal Personnel](#) [0]

**Related sections**

- [Expert Witnesses](#)
- [Criminology](#)

Search  in

[Advanced search](#) [Thesaurus](#) [Subject A-Z](#) [New resources](#) [Help](#)

Filter by

Full record   Save Record

Now showing: 1 - 25 of 30 records (Filter: All resource types)  
Page: 1  Order by [Date added](#) | [Title](#)

**Division of Forensic Psychology, British Psychological Society** Editor's Choice

The UK's professional group for forensic psychologists - a Division of The British Psychological Society (BPS). The Division aims to represent the interests of psychologists whose work involves them in the criminal and civil Justice . Only those who have completed an approved training may join as a full member. The website describes the aims and work of the Division.

Rys. 5.11. Specyfikacja relacji semantycznych w grupie tematycznej *Intute: nauki społeczne*. [Źródło: <http://intute.ac.uk/>]

Kategorie powiązane z przedstawioną na rysunku 5.10 kategorią *Psychologia sądowa i zagadnienia prawne* zorganizowano w dwóch grupach:

- sub-categories: wskaźnik hierarchicznej relacji podrzędności zakresowej,
- related sections: wskaźnik niespecyfikowanej relacji asocjacyjnej.

Przeprowadzona przez autora analiza wyrażen pod kątem przynależności do określonego typu kategorii wykazała, że wyrażenia stanowiące nazwę kategorii tematycznych w strukturze hierarchicznej odwzorowują wyłącznie wymiar treści obiektów informacyjnych. Nie odnotowano wyrażen odwzorowujących cechy formalne, które stanowiłyby punkty dostępu do zbioru wyszukiwawczego w omawianym

narzędziu.

Porządek pionowy kategorii tematycznych, tak jak na pierwszym, tak i na kolejnych stopniach podziału jest wyznaczony przez układ alfabetyczny wyrażeń przyjętych za ich nazwy.

Wyodrębnianie kategorii tematycznych w narzędziach dostępu przedmiotowego we wszystkich podsystemach Intute jest zgodne z zasadą *literary warrant*. Na podstawie analizy treściowej kolekcji wyodrębnia się wyłącznie tematy posiadające co najmniej jednego reprezentanta w postaci obiektu informacyjnego, którego opis przechowywany jest w bazie systemu. W tym przypadku utworzono również kategorie puste (tj. bez egzemplarza w postaci ChWD), np. wśród kategorii wyodrębnionych poprzez podział kategorii drugiego stopnia *Prawo międzynarodowe* znalazły się m.in.: *Prawo konsularne* i *Prawo dyplomatyczne*, do których aktualnie nie przyporządkowano żadnej ChWD.

Hierarchiczny układ kategorii tematycznych umożliwia dostęp do 31008 charakterystyk wyszukiwawczych (stan na 19-12-2006) za pomocą 88753 punktów dostępu. Oznacza to, że jedną ChWD przyporządkowano średnio do trzech kategorii tematycznych w strukturze. Ilościową i procentową dystrybucję ChWD na poszczególnych poziomach struktury hierarchicznej kategorii przedstawia Tab. 5.15.

Tab. 5.15. Dystrybucja ChWD na poszczególnych poziomach w układzie hierarchicznym w grupie Intute: nauki społeczne. [Źródło: opracowanie własne]

Poziom	Liczba ChWD	Udział procentowy
Poziom 1	0	0%
Poziom 2	24957	28,12%
Poziom 3	32305	36,39%
Poziom 4	27049	30,48%
Poziom 5	3281	3,69%
Poziom 6	1161	1,31%
Suma	88753	99,99%

Kategorie tematyczne wyodrębnione na stopniach podziału od drugiego do czwartego, skupiają średnio po 30% zasobów systemu. Średnia pojemność kategorii na poziomach drugim i szóstym jest zbliżona i wynosi około 85 ChWD, zaś na poziomach od trzeciego do piątego wynosi średnio 45 ChWD (Tab. 5.16).

Tab. 5.16. Średnia pojemność kategorii na poszczególnych stopniach podziału.

[Źródło: opracowanie własne]

Poziom	Średnia pojemność kategorii
Poziom 1	0

Poziom 2	81
Poziom 3	44
Poziom 4	40
Poziom 5	55
Poziom 6	89

Dystrybucja ChWD na poszczególnych poziomach struktury hierarchicznej kategorii tematycznych jest odzwierciedleniem ich rozkładu ilościowego przedstawionego w Tab. 5.14, przy czym w kategoriach na pierwszym stopniu podziału nie grupuje się ChWD. Ponadto 94% łącznej liczby charakterystyk wyszukiwawczych zgromadzono na poziomach od drugiego do czwartego.

### 5.4.3. Intute: zdrowie i nauki o życiu (*Intute: Health and Life Sciences*)

Grupa tematyczna Intute: zdrowie i nauki o życiu powstała, tak jak w przypadku poprzednio analizowanych grup, poprzez połączenie zbiorów informacyjnych kilku dziedzinowych systemów hipertekstowych. Były to systemy działające w podsieci RDN z zakresu medycyny i nauki o zdrowiu zwanej BIOME:

- OMNI: Health and Medicine – medycyna i zdrowie,
- NMAP: Nursing, Midwifery and Allied Health – pielęgniarstwo, położnictwo, opieka zdrowotna,
- BioResearch: Biological Sciences - biologia,
- AgriFor: Agriculture, Food and Forestry - rolnictwo, leśnictwo, nauka o żywieniu,
- VetGate: Animal Health – weterynaria,
- Natural Selection - botanika, zoologia, paleontologia, ekologia, nauka o środowisku.

Dodatkowo zakres grupy tematycznej został poszerzony o zbiory trzech DSH:


- BioEthicsWeb – etyka w biomedycynie,
- MedHist – historia medycyny,
- psci-com – społeczny ruch naukowy, komunikacja w nauce.

Pomimo wspólnej nazwy dla grupy tematycznej, poszczególne kolekcje zachowały swoją autonomię. Tym samym dla każdej z nich przewidziano odrębne narzędzia dostępu przedmiotowego. Dostęp do ich zasobów w tym trybie realizowany jest poprzez eksplorację kategorii tematycznych zgromadzonych w dziewięciu klasach

głównych grupujących ChWD i tworzących zbiór wyszukiwawczy tej grupy tematycznej (Rys. 5.11). Są to:

1. Medycyna (ang. Medicine),
2. Pielęgniarstwo, położnictwo, opieka zdrowotna (ang. Nursing, Midwifery and Allied Health),
3. Weterynaria (ang. Veterinary),
4. Biologia (ang. Bioresearch),
5. Historia naturalna (ang. Natural History),
6. Rolnictwo, leśnictwo, nauka o żywieniu (ang. Agriculture, Food and Forestry),
7. Etyka w biomedycynie (ang. BioethicsWeb),
8. MedHist - historia medycyny (ang. MedHist - history of medicine resources),
9. Psci-com - społeczny ruch naukowy, komunikacja w nauce (ang. Psci-com - public engagement with science and technology).

[Intute](#) > Health and Life Sciences



<b>Subject links</b>	<p>Welcome to the Health and Life Sciences pages of Intute. We are a free online service providing you with access to the very best web resources for education and research, evaluated and selected by a network of subject specialists. There are over 31,000 resource descriptions listed here that are freely accessible for keyword searching or <a href="#">browsing</a>.</p> <p>This service was formerly known as BIOME</p> <p><b>Search our Health and Life Sciences database</b></p> <p>Search <input type="text"/> <a href="#">Go</a></p> <p><a href="#">Advanced search</a> <a href="#">New resources</a> <a href="#">Help</a></p> <p>Quick search: <a href="#">Avian influenza</a></p> <p>You can also search or browse by the subjects below:</p> <p><a href="#">Medicine</a>  <a href="#">Nursing, Midwifery and Allied Health</a>  <a href="#">Veterinary</a>  <a href="#">Bioresearch</a>  <a href="#">Natural History</a>  <a href="#">Agriculture, Food and Forestry</a>  <a href="#">BioethicsWeb</a> - biomedical ethics  <a href="#">MedHist</a> - history of medicine resources  <a href="#">Psci-com</a> - public engagement with science and technology</p>	<b>Login to MyIntute</b>
<a href="#">About us</a>		<b>Group News</b>
<a href="#">A-Z of services</a>		<ul style="list-style-type: none"> <li>• <a href="#">Free Agriculture booklet</a></li> <li>• <a href="#">Animal diseases - Hot Topic</a></li> <li>• <a href="#">New Internet tutorials for Health and Life Sciences</a></li> <li>• <a href="#">August Hot Topic - Depressive Disorders</a></li> <li>• <a href="#">Ethics of stem cell research - Hot topic for June</a></li> </ul>
<b>Internet catalogue</b>		
<a href="#">Advanced search</a>		
<a href="#">Browse</a>		
<a href="#">New resources</a>		
<a href="#">Suggest a site</a>		
<b>Internet training</b>		
<a href="#">Virtual Training Suite</a>		
<a href="#">Support materials</a>		
<a href="#">Intute events</a>		
<b>Additional services</b>		
<a href="#">Blog</a>		
<a href="#">Hot topics</a>		
<a href="#">Newsround</a>		
<a href="#">Timelines</a>		
<a href="#">MedHist events</a>		
<a href="#">Psci-com events</a>		
<a href="#">Psci-com books and articles</a>		

>> [More news](#)

Rys. 5.12. Podkolekcje w grupie tematycznej Intute: zdrowie i nauki o życiu

[Źródło: <http://intute.ac.uk/healthandlifesciences/>]

W zdecydowanej większości narzędzi dominuje alfabetyczna (ahierarchiczna) organizacja kategorii tematycznych. Ich szczegółowa charakterystyka w tej grupie zostanie przedstawiona na konkretnych przykładach kolekcji z zakresu: medycyny, pielęgniarstwa i położnictwa, historii naturalnej oraz historii medycyny.

#### **5.4.3.1. Medycyna**

Zakres klasy głównej obejmuje medycynę oraz naukę o zdrowiu. Kolekcja ta stanowiła przed powstaniem Intute zbiór informacyjny systemu OMNI. Narzędzie dostępu przedmiotowego zbudowane było wówczas z kategorii tematycznych, których nazwy były odpowiednikami słownymi symboli UKD. W 1998 roku zbiór kategorii tematycznych zreorganizowano do postaci alfabetycznego wykazu wyrażań, których formę stanowiły odpowiedniki słowne symboli klasyfikacji National Library of Medicine Classification. Wybór ten podyktowany był specjalistycznym zakresem tego JIW, który był zbieżny z zakresem tematycznym systemu. [Slavic 2006] Obecnie narzędzie dostępu przedmiotowego tworzy alfabetycznie uporządkowany zbiór kategorii tematycznych, których forma przejmowana jest ze słownika MeSH. Ten JIW jest wykorzystywany do indeksowania zasobów kolekcji, a narzędzie dostępu przedmiotowego tworzy się generując wykaz autosyntaktycznych jednostek leksykalnych (tematów) zastosowanych do reprezentacji treści elementów zbioru informacyjnego.

Dostęp do zbioru wyszukiwawczego realizuje 4190 (stan na 30-08-2006 r.) haseł przedmiotowych MeSH pełniących funkcję punktów dostępu. Wielkość zbioru to 10130 ChWD (stan na 30-08-2006 r.).

W strukturze tego narzędzia wyodrębniono 26 kategorii odpowiadających literom alfabetu łacińskiego. Zgrupowano w nich tematy MeSH wraz z kwalifikatorami rozpoczynające się od konkretnej litery alfabetu. W charakteryzowanym zbiorze leksyki można wyodrębnić dwie kategorie wyrażań. Pierwsze to tematy MeSH (występujące samodzielnie lub z odpowiednimi kwalifikatorami rzeczowymi), reprezentujące cechy treści obiektów informacyjnych. Druga kategoria wyrażań pełniących funkcję punktów dostępu odwzorowuje formę obiektów informacyjnych. Każde z nich pełniąc funkcję węzła dla odesłania hipertekstowego odsyła do zbioru odpowiednich ChWD.

Wśród wyrażań reprezentujących treść obiektów informacyjnych można wyróżnić jedno- lub wielowyrazowe tematy MeSH oraz hasła przedmiotowe MeSH

wraz z kwalifikatorami rzeczowymi (ang. subheadings). Forma hasła przedmiotowego MeSH wraz z kwalifikatorem rzeczowym w narzędziu dostępu polega na formalizacji zapisu do postaci:

*temat MeSH / kwalifikator,*

np.

[Alcohol Drinking / epidemiology](#)

Forma tematu MeSH oraz kwalifikatora jest równoznaczna z ich zapisem w słowniku tego JIW. Druga kategoria wyrażen służy do odwzorowywania cech formalnych, a przede wszystkim do identyfikacji rodzaju obiektu informacyjnego. Przynależność jednostki do tej kategorii leksykalnej jest zaznaczana za pomocą kwalifikatora *typ publikacji* (ang. publication type):

*wartość cechy [Publication Type],*

np.

[Database \[Publication Type\]](#)

#### **5.4.3.2. Pielęgniarstwo, położnictwo, opieka zdrowotna**

Zakres klasy głównej obejmuje zagadnienia pielęgniarstwa, położnictwa oraz opieki zdrowotnej. Kolekcja ta stanowiła przed uruchomieniem systemu Intute zbiór informacyjny systemu NMAP (Nursing, Midwifery and Allied health Professions). Dostęp do zasobów tej grupy tematycznej jest realizowany przez dwa narzędzia dostępu przedmiotowego. Jest to alfabetyczny wykaz kategorii tematycznych, których nazwy pochodzą ze słownika MeSH. Drugie, komplementarne narzędzie zbudowane jest z kategorii tematycznych, których nazwy mają formę deskryptorów teaurusu The Royal College of Nursing Thesaurus. Zbiór wyszukiwawczy tej klasy głównej liczy 3901 ChWD.

W tym przypadku mamy do czynienia z polireprezentacją wiedzy w formie ukrytej\*. Polega to na stosowaniu kilku JIW do indeksowania obiektów informacyjnych, bez identyfikacji źródła pochodzenia użytych jednostek leksykalnych w strukturze rekordu. Tutaj są to MeSH oraz wspomniany teaurus The Royal College of Nursing Thesaurus (RCN). W rekordzie dla obiektu informacyjnego wyświetlanego użytkownikowi nie uwzględniono osobnych pól lub podpól dla wyrażen pochodzących

---

\* Zjawisko polireprezentacji wiedzy w DSH przedstawiono w rozdziale 4.2.6.



ze wspomnianych słowników JIW. Wykorzystano jednak osobne narzędzia dostępu przedmiotowego zbudowane na ich podstawie. Wygenerowano dwa indeksy alfabetyczne, wyrażeń MeSH oraz deskryptorów tezauryśa RCN.

W pierwszym przypadku zastosowano 4191 kategorii tematycznych, reprezentujących zarówno cechy treści oraz formy obiektów informacyjnych. W drugim, użyto 1528 jednostek jedno- lub wielowyrazowych równokształtnych z deskryptorami wspomnianego tezauryśa, które odwzorowują wyłącznie treść obiektów informacyjnych.

Aktywacja powiązania hipertekstowego dla wybranej kategorii tematycznej w narzędziu dostępu przedmiotowego zbudowanego z wykorzystaniem słownika MeSH powoduje wyświetlenie oprócz listy relewantnych ChWD, jednej lub dwóch kategorii tematycznych pozostających z eksplorowaną w hierarchicznej relacji podrzędności lub nadrzędności zakresowej, np. aktywacja powiązania hipertekstowego dla wyodrębnionej kategorii tematycznej *Zaburzenia językowe* (ang. *Language Disorders*) powoduje odpowiedź systemu w postaci przedstawionej na Rys. 5.13.

The screenshot shows a search interface for 'Language Disorders'. At the top, the title 'Language Disorders' is displayed in green. Below it, two subcategories are listed: 'Language Disorders / therapy' and 'Language Disorders / complications'. A search bar is present with the text 'Search' and a dropdown menu set to 'Nursing, Midwifery and Allied Health'. Below the search bar, there are links for 'Advanced search', 'Subject A-Z', 'New resources', and 'Help'. A filter section shows 'Filter by' with a dropdown menu set to 'All resource types'. Below this, there are buttons for 'Details', 'Full record', and 'Go to website'. The interface indicates 'Now showing: 1 - 23 of 23 records (Filter: All resource types)' and 'Page: 1'. The search results are ordered by 'Date | Title'. The first result is for the 'American Speech-Language-Hearing Association (ASHA)', with a brief description and a link to 'http://www.asha.org/default.htm'. The second result is for 'Aphasia', with a brief description and a link to 'http://www.asha.org/public/speech/disorders/Aphasia\_info.htm'. The type of the results is identified as 'Other organisations'.

Rys. 5.13. Relacje paradygmatyczne w narzędziu dostępu przedmiotowego.

[Źródło: <http://intute.ac.uk/>]

Dla wyrażenia *Zaburzenia językowe* przewidziano w systemie dwie powiązane kategorie tematyczne *Zaburzenia językowe / terapia* oraz *Zaburzenia językowe / powikłania* pozostających z wybranym w hierarchicznej relacji podrzędności zakresowej, specyfikowanej za pomocą określenia *Subcategory* (podkategoria). Taka opcja pozwala na modyfikację przez użytkownika zapytania poprzez wybór ChWD zgromadzonych w sugerowanych przez system kategoriach tematycznych. Odpowiednie relacje paradygmatyczne pomiędzy wyrażeniami w MeSH stanowią tutaj podstawę do budowy takich powiązań w tym narzędziu dostępu. Przy czym nie wszystkie punkty dostępu, w postaci tematów MeSH, w narzędziu są wyposażone w dodatkowe powiązania hierarchiczne z innymi tematami.

#### **5.4.3.3. Historia naturalna**

Zakres klasy głównej obejmuje zagadnienia historii naturalnej, czyli ewolucji i zmian zachodzących w przyrodzie. Kolekcja ta stanowiła przed włączeniem do Intute zbiór wyszukiwawczy systemu Natural Selection działający w ramach sieci RDN/BIOME. Liczy ona 5551 ChWD (stan na 30-08-2006 r.).

Narzędzie dostępu przedmiotowego ma tutaj formę hierarchicznego układu kategorii tematycznych. Jego struktura została zaprojektowana na potrzeby systemu. Dostępna dokumentacja systemu nie wskazuje na jawne wykorzystanie do tego celu konkretnego JIW. Na pierwszym stopniu podziału wyodrębniono tutaj 16 następujących kategorii tematycznych:

1. Różnorodność biologiczna (ang. Biodiversity),
2. Biogeografia (ang. Biogeography),
3. Biologia (ang. Biology),
4. Botanika – rośliny (ang. Botany – Plants),
5. Przechowywanie i ochrona (ang. Conservation & protection),
6. Ekologia (ang. Ecology),
7. Nauka o środowisku (ang. Environmental sciences),
8. Ewolucja (ang. Evolution),
9. Mikologia (ang. Mycology),
10. Oceanografia (ang. Oceanography & aquatic sciences),
11. Paleobotanika (ang. Palaeobotany),
12. Paleozoologia (ang. Palaeozoology),
13. Protisologia lub protozoologia (ang. Protistology),

14. Paleontologia stratygraficzna (ang. Stratigraphic palaeontology),
15. Systematyka i taksonomia (ang. Systematics & Taxonomy),
16. Zoologia – zwierzęta (ang. Zoology – Animals).

Wszystkie wyodrębnione kategorie tematyczne na pierwszym stopniu podziału, reprezentują cechy treściowe obiektów informacyjnych. Kategorie tematyczne wyodrębniane na kolejnych stopniach podziału uporządkowano alfabetycznie. Przy czym rozbudowa poszczególnych kategorii odbywa się często z wykorzystaniem kilku zasad podziału, np. w kategorii pierwszego poziomu *Ekologia* wyodrębniono podkategorie o nazwach:

- zwierzęta (ang. animals),
- teorie ewolucji (ang. evolutionary theories),
- wymarłe gatunki (ang. extinct species),
- tematy ogólne (ang. general topics),
- historia / ludzie (ang. history / people),
- mikroorganizmy / grzyby (ang. microorganisms / fungi),
- organizacje (ang. organisations),
- rośliny (ang. plants),
- oprogramowanie – modelowanie komputerowe (ang. software - computer modelling),
- oprogramowanie – przetwarzanie danych (ang. software - data processing),
- terminologia (ang. terminology).

Tym samym w jednym szeregu umieszczono kategorie, których nazwy należą do kategorii leksykalnych odwzorowujących zarówno cechy treściowe, jak i formalne. Ich ilościowy rozkład na poszczególnych stopniach podziału przedstawia Tab. 5.17.

Tab. 5.17. Ilościowy rozkład kategorii tematycznych w klasie głównej Historia Naturalna  
 . [Źródło: opracowanie własne]

Poziom	Liczba kategorii	Procentowy rozkład kategorii
Poziom 1	16	4,53%
Poziom 2	148	41,82%
Poziom 3	155	43,76%
Poziom 4	35	9,89%
Suma	354	100,00%

Przedstawione w Tab. 5.17 dane wskazują na powtarzającą się zależność gromadzenia znaczącego odsetku kategorii tematycznych na poziomach drugim i trzecim. W tym przypadku ponad 85% łącznej liczby wyodrębnionych kategorii zgromadzono na tych

poziomach. Ma to bezpośrednie przełożenie na dystrybucję odesłań do ChWD na poszczególnych poziomach w strukturze narzędzia (Tab. 5.18).

Tab. 5.18. Ilościowy rozkład odesłań do ChWD w klasie głównej Historia Naturalna.

[Źródło: opracowanie własne]

Poziom	Liczba odesłań do ChWD	Udział procentowy
Poziom 1	3	0,03%
Poziom 2	4927	48,59%
Poziom 3	4581	45,18%
Poziom 4	629	6,2%
Suma	10140	100,00%

W tym przypadku zachodzi ścisła korelacja pomiędzy wspomnianymi zmiennymi. Na poziomach drugim i trzecim zgromadzono ponad 93% odesłań do ChWD.

#### 5.4.3.4. Historia medycyny

Zakres tej klasy głównej obejmuje zagadnienia związane z historią medycyny i nauk pokrewnych uwzględniające problematykę rozwoju wiedzy medycznej. Kolekcja przyporządkowana do tej klasy stanowiła przez włączeniem do Intute zbiór wyszukiwawczy systemu MedHist. Wielkość zbioru wyszukiwawczego wynosi 1127 ChWD (stan na 30-08-2006). Zastosowano tutaj dwa narzędzia dostępu przedmiotowego. Są to alfabetyczny wykaz kategorii tematycznych utworzonych na podstawie tematów MeSH oraz czteropoziomowa struktura hierarchiczna kategorii tematycznych. Zasady wykorzystania MeSH w postaci alfabetycznego wykazu tematów jako punktów dostępu są identyczne jak w przypadku wcześniej omawianych klas głównych stosujących to narzędzie. W tym przypadku wyodrębniono 478 tematów MeSH.

Drugim narzędziem dostępu przedmiotowego w tej klasie głównej jest czteropoziomowa struktura hierarchiczna kategorii tematycznych. Na pierwszym stopniu podziału wyodrębniono 11 kategorii w układzie alfabetycznym:

1. Choroby (ang. Diseases),
2. Edukacja (ang. Education),
3. Publikacje elektroniczne (ang. Electronic Publications),
4. Kalendarium wydarzeń (ang. Event listings),
5. Okresy historyczne (ang. Historical Period),
6. Regiony (ang. Locality),

7. Specjalności, techniki i materiały medyczne (ang. Medical specialties, techniques & materials),
8. Rozmaitości związane z medycyną (ang. Miscellany relating to medicine),
9. Ludzie (ang. People),
10. Źródła (ang. Reference),
11. Nauka i technologia (ang. Science & Technology).

Schemat ten pokazuje, że dokonano tutaj podziału uniwersum rzeczywistości dokumentacyjnej w z zastosowaniem różnych zasad podziału:

- Kategorie *Publikacje elektroniczne* oraz *Źródła* ewidentnie wskazują na formę obiektów informacyjnych.
- Kategorie tematyczne wyodrębnione w kategorii *Regiony* reprezentują odniesienie przedmiotu obiektu informacyjnego do jego lokalizacji w przestrzeni geograficznej.
- Kategorie *Kalendarium wydarzeń* oraz *Okresy historyczne* wskazują na cechy należące do wymiaru czasu.

W kategorii *Kalendarium wydarzeń* zaprezentowano interesujący sposób organizacji i prezentacji informacji. System Intute jako jeden z czołowych reprezentantów dziedzinowych systemów hipertekstowych należy do dokumentacyjnych systemów informacyjno-wyszukiwawczych. System na wyjściu udostępnia charakterystyki wyszukiwawcze zaindeksowanych internetowych obiektów informacyjnych. W omawianej klasie głównej mamy do czynienia z hybrydyzacją tego modelu, gdyż na wyjściu systemu udostępniane są informacje faktograficzne, dotyczące wybranego fragmentu rzeczywistości pozadokumentacyjnej. Są to informacje o bieżących i planowanych sympoziach, konferencjach czy też ważnych wydarzeniach dotyczących medycyny. W kategorii *Kalendarium wydarzeń* wyodrębniono na drugim stopniu podziału kategorie *Wydarzenia – dla specjalistów* oraz *Wydarzenia – dla wszystkich*. W pierwszej, zgrupowano informacje przeznaczone dla specjalistów, czyli wyodrębniono szczegółowe wydarzenia mające znaczenie dla omawianego zagadnienia. W drugiej, dokonano selekcji wyodrębniając tylko najważniejsze, przeznaczone dla użytkowników niebędących specjalistami w tym zakresie. W obydwu kategoriach informacje o wydarzeniach zorganizowano w układzie chronologicznym tj. na podstawie daty dziennej, rozpoczynając od wydarzenia najbardziej aktualnego (Rys. 5.14).

## MedHist Events

[Events aimed at a professional audience](#)

[Events aimed at a public audience](#)

Some events are suitable for both audiences and will appear in both sections

### Search the events database

Enter keywords	<input type="text" value="All terms"/>
Dates	From: <input type="text" value="02"/> <input type="text" value="2007"/> to: <input type="text" value="12"/> <input type="text" value="2010"/>
Audience	<input type="radio"/> Any <input type="radio"/> Professional <input checked="" type="radio"/> Public

[Help on searching events](#)

[Submit an event](#)

There are 41 events from 02/2007 to 12/2010

Audience: public audience

Keyword(s): All terms

Start	Event
11 Jun 2007	<a href="#">Medicinal plants study week</a>
04 Jun 2007	<a href="#">'From creation to inheritance: how reproduction changed its meaning'</a>
19 May 2007	<a href="#">The poetry show</a>
16 May 2007	<a href="#">What happened to the polymaths?</a>
26 Apr 2007	<a href="#">War of the mind : battle and its psychological effects, 1914-1945</a>
25 Apr 2007	<a href="#">Peter Ritchie Calder and the public culture of 20th century science</a>

Rys. 5.14. Dostęp do informacji w kategorii *Kalendarium wydarzeń* [Źródło: <http://intute.ac.uk/>]

Przedstawiony na rys. 5.14 dostęp do tego rodzaju zasobów możliwy jest za pomocą wykazu wydarzeń w układzie chronologicznym lub wyszukiwarki. Nazwa wydarzenia lub imprezy stanowi tutaj węzeł dla powiązania hipertekstowego. Jego aktywacja powoduje wyświetlenie przez system:

- nazwy wydarzenia,
- jego charakterystyki,
- adresu strony internetowej instytucji organizującej,
- daty rozpoczęcia i zakończenia, ewentualnie godzin odbywania się imprezy,
- opłat związanych z uczestnictwem,
- informacji umożliwiających kontakt z osobą odpowiedzialną (e-mail, nr telefonu, faksu),
- lokalizacji wydarzenia.

Dostęp do ChWD w klasie głównej Historia medycyny jest realizowany również poprzez czteropoziomą strukturę hierarchiczną zbudowaną z 436 kategorii tematycznych. Ilościowy rozkład kategorii na poszczególnych poziomach przedstawia Tab. 5.19.

Tab. 5.19. Rozkład kategorii w klasie głównej Historia medycyny. [Źródło: opracowanie własne]

Poziom	Liczba kategorii	Procentowy rozkład kategorii
Poziom 1	11	2,51%
Poziom 2	162	37,14%
Poziom 3	220	50,46%
Poziom 4	43	9,87%
Suma	436	100,00%

Tak, jak w przypadku pozostałych hierarchicznych narzędzi dostępu przedmiotowego, widoczna jest tutaj tendencja do wyodrębniania największej liczby kategorii na poziomach drugim i trzecim. W tym przypadku stanowi to około 87% łącznej liczby wyodrębnionych kategorii.

Tab. 5.20. Rozkład odesłań w klasie głównej *Historia medycyny*. [Źródło: opracowanie własne]

Poziom	Liczba odesłań do ChWD	Udział procentowy
Poziom 1	138	2,27%
Poziom 2	2952	48,72%
Poziom 3	2729	45,04%
Poziom 4	240	3,96%
Suma	6059	100,00%

Przedstawiony w Tab. 5.20 rozkład odesłań na poszczególnych poziomach układu hierarchicznego również potwierdza tendencję do grupowania ChWD w kategoriach tematycznych na drugim i trzecim stopniu podziału (ponad 93% ChWD). Zbiór kolekcji liczy 1127 ChWD, do których skierowuje 6059 kategorii pełniąc funkcję punktów dostępu. Oznacza to, że jedną ChWD przyporządkowano średnio do pięciu kategorii tematycznych.

W klasie głównej Historia medycyny zastosowano dodatkowe narzędzie dostępu w postaci Kartoteki Wzorcowej Nazw Osobowych Biblioteki Kongresu (Library of Congress Name Authority File). Podczas indeksowania dokumentów, formę dla tematów osobowych przejmuje się ze wspomnianego zbioru leksyki. Na jego podstawie opracowano odrębne narzędzie dostępu do kolekcji w oparciu o wartości dla tej cechy. Formy nazw osobowych, przejęte ze zbioru słownictwa kontrolowanego, są tutaj uporządkowane alfabetycznie. Każda wyodrębniona nazwa osobowa stanowi węzeł dla powiązania hipertekstowego, którego aktywacja powoduje wyświetlenie listy ChWD, w których je użyto do odwzorowania cech treści.

Intute jest bardzo rozbudowanym systemem, który z oferowanymi funkcjami pretenduje do miana zintegrowanej platformy informacyjnej. Grupy tematyczne stanowiące główny trzon systemu, udostępniają odmienne narzędzia dostępu

przedmiotowego reprezentujące wariant ahierarchiczny oraz hierarchiczny porządkowania kategorii tematycznych jako punktów dostępu. Wnioski z badań przeprowadzonych przez autora nad systemem Intute dotyczyły przede wszystkim tendencji lub prawidłowości, jakie zaobserwowano analizując budowę narzędzi dostępu przedmiotowego i udziału w niej JIW.

W narzędziach dostępu przedmiotowego Intute, języki informacyjne są wykorzystywane przede wszystkim jako źródło leksyki do tworzenia nazw dla kategorii tematycznych. W większości analizowanych przypadków, głównym narzędziem dostępu przedmiotowego jest struktura hierarchiczna kategorii tematycznych. Tylko w niewielu przypadkach porządek pionowy kategorii opracowywano na podstawie języków informacyjnych typu klasyfikacji. Zazwyczaj schematy te miały postać hierarchicznych układów kategorii zaprojektowanych na potrzeby narzędzi dostępu do systemu.

Ekstrahowane struktury hierarchiczne kategorii tematycznych w narzędziach dostępu przedmiotowego Intute poddano badaniom statystycznym. W przedstawionych w tym miejscu rozprawy wynikach badań autora nad systemem Intute, uwzględniono również narzędzia dostępu przedmiotowego dodatkowych podkolekcji, tworzących zasób informacyjny tego systemu. Na pierwszym poziomie wyodrębniono w nich od sześciu (w humanistyce) do dziewiętnastu kategorii (w naukach społecznych) (Tab. 5.21).

Tab. 5.21. Liczba klas wyodrębnionych na pierwszym stopniu podziału. [Źródło: opracowanie własne]

Nazwa	Liczba kategorii na pierwszym poziomie
Intute Nauka, Inżynieria i Technologia	10
Intute Humanistyka i Sztuka	6
Intute Nauki Społeczne	19
Intute Historia Naturalna	16
Intute Bioetyka	10
Intute Historia Medycyny	11
Intute Społeczny Ruch Naukowy	9

Głębokość struktur dla poszczególnych hierarchicznych przedstawia Tab. 5.22.

Tab. 5.22 Głębokość struktur hierarchicznych w Intute. [Źródło: opracowanie własne]

Nazwa	Liczba poziomów	Liczba ChWD
Intute Nauka, Inżynieria i Technologia	6	33452
Intute Nauki społeczne	6	31008



Intute Humanistyka i Sztuka	7	18000
Intute Historia Naturalna	4	5551
Intute Historia Medycyny	4	1127
Intute Społeczny Ruch Naukowy	5	960
Intute Bioetyka	4	655

Trudno tutaj orzekać o bezpośredniej zależności pomiędzy wielkością zbioru wyszukiwawczego, a głębokością struktur hierarchicznych opierając się wyłącznie na kilku przypadkach. Niemniej ostrożne wnioski wskazują, że w kolekcjach powyżej 15000 ChWD głębokość schematów wynosi od sześciu do siedmiu poziomów, a w kolekcjach mniejszych od czterech do pięciu stopni podziału.

Tab. 5.23. Liczba wyodrębnionych kategorii w poszczególnych grupach Intute.

[Źródło: opracowanie własne]

Nazwa	Liczba kategorii	Liczba ChWD
Intute Nauka, Inżynieria i Technologia	1176	33452
Intute Nauki Społeczne	1793	31008
Intute Humanistyka i Sztuka	2199	18000
Intute Historia Naturalna	354	5551
Intute Historia Medycyny	436	1127
Intute Społeczny ruch naukowy	85	960
Intute Bioetyka	175	655

Podobna sytuacja ma miejsce przy zestawieniu całkowitej liczby wyodrębnionych kategorii tematycznych z wielkością tych kolekcji (Tab. 5.23). W przypadku Intute, wielkość kolekcji przekłada się na liczbę wyodrębnionych kategorii w narzędziach dostępu przedmiotowego. W organizacji zbiorów powyżej 15000 ChWD wykorzystuje się średnio 1700 kategorii, zaś w mniejszych kolekcjach około 260.

We wszystkich analizowanych przypadkach hierarchicznych narzędzi dostępu przedmiotowego w kolekcjach Intute, gromadzi się ponad 80% odesłań do ChWD w kategoriach tematycznych na drugim i trzecim stopniu podziału.

W większości przypadków kategorie wyodrębniane na poszczególnych stopniach podziału uporządkowane są w kolejności alfabetycznej. Powszechny jest również jednoczesny podział kategorii tematycznych na tym samym stopniu według kilku zasad podziału.

W narzędziach dostępu przedmiotowego w postaci alfabetycznego wykazu kategorii tematycznych, JIW pełnił jedynie funkcję pomocniczą jako źródło tworzenia ich nazw. W przypadku wyrażen odwzorowujących formę obiektów informacyjnych

stosuje się wskaźnik *rodzaj zasobów* określający przynależności kategorialną wyrażenia. Nie jest to jednak stosowane konsekwentnie i w analizowanym systemie pojawiły się narzędzia dostępu przedmiotowego, w których nie wyodrębniono takiego wskaźnika (np. Historia medycyny).

## 5.5. *MathGuide*

MathGuide jest dziedzinowym systemem hipertekstowym działającym w ramach niemieckiej sieci Special Subject Guides / SSG-Fachinformation (SSG-FI). Sieć powstała w 1996 roku jako efekt projektu badawczego realizowanego przez Bibliotekę Uniwersytecką w Göttingen (Niedersächsische Staats- und Universitätsbibliothek Göttingen) oraz Deutsche Forschungsgemeinschaft - centralną niemiecką fundację do spraw promocji badań naukowych. [Fischer, Neuroth 2000] Zakres gromadzonych dokumentów w ramach sieci obejmuje cztery dziedziny: matematykę teoretyczną, nauki o Ziemi, leśnictwo oraz kulturę angloamerykańską. Grupę użytkowników docelowych stanowią studenci i pracownicy naukowci. [Fischer, Neuroth 2000]

Dostęp do zasobów systemu jest realizowany poprzez mechanizm wyszukiwawczy oraz dwa narzędzia dostępu przedmiotowego. Pierwsze nosi nazwę *katalogu przedmiotowego* (ang. *subject catalog*), drugie - *katalog typów zasobów* (ang. *source type catalog*). W budowie pierwszego narzędzia wykorzystano porządek klas w klasyfikacji Mathematical Subject Classification (MSC). Kategorie tematyczne grupujące ChWD zostały uporządkowane do postaci struktury hierarchicznej, gdzie na pierwszym stopniu podziału wyodrębniono kategorie:

1. Matematyka ogólna (ang. General mathematics),
2. Historia i podstawy (ang. History and Foundations),
3. Algebra (ang. Algebra),
4. Analiza (ang. Analysis),
5. Geometria (ang. Geometry),
6. Statystyka (ang. Statistics),
7. Matematyka stosowana (ang. Applied Mathematics).

Wyrażenia te pełnią funkcję „superklas” grupujących kategorie niższego rzędu przejęte z klasyfikacji MSC. W ramach każdej z nich wyodrębniono kategorie tematyczne, których nazwy zbudowane są z symbolu klasyfikacji MSC oraz jego odpowiednika słownego. W ten sposób wyodrębniono 58 kategorii tematycznych pełniących funkcję

punktów dostępu do zasobów systemu (Tab. 5.24).

Tab. 5.24. Organizacja kategorii tematycznych w MathGuide  
z wykorzystaniem Mathematics Subject Classification. [Źródło: opracowanie własne]

Nazwa kategorii tematycznej pierwszego stopnia	Nazwa kategorii tematycznej drugiego stopnia w MathGuide	Forma symbolu i odpowiednika słownego w MSC
General mathematics	00 General 00A35 Methodology of Mathematics, Didactics	<a href="#">00-xx</a> General 00A35 Methodology of mathematics, didactics
History and Foundations	01 History 03 Mathematical logic and foundations 04 Set theory  05 Combinatorics	<a href="#">01-xx</a> History and biography <a href="#">03-xx</a> Mathematical logic and foundations 04-xx This section has been deleted {For set theory see <a href="#">03Exx</a> } – klasa usunięta  <a href="#">05-xx</a> Combinatorics
Algebra	06 Order, lattices, ordered algebraic structures 11 Number theory 12 Field theory and polynomials 13 Commutative rings and algebras 14 Algebraic Geometry 15 Linear and multilinear algebra, matrix theory 16 Associative rings and algebras  17 Nonassociative rings and algebras 18 Category theory, homological algebra 19 K-theory 20 Group theory and generalizations 22 Topological groups, Lie algebras	<a href="#">06-xx</a> Order, lattices, ordered algebraic structures <a href="#">11-xx</a> Number theory <a href="#">12-xx</a> Field theory and polynomials <a href="#">13-xx</a> Commutative rings and algebras <a href="#">14-xx</a> Algebraic geometry <a href="#">15-xx</a> Linear and multilinear algebra; matrix theory <a href="#">16-xx</a> Associative rings and algebras <a href="#">17-xx</a> Nonassociative rings and algebras <a href="#">18-xx</a> Category theory; homological algebra <a href="#">19-xx</a> K-theory <a href="#">20-xx</a> Group theory and generalizations <a href="#">22-xx</a> Topological groups,
Analysis	26 Real functions 28 Measure and integration 30 Functions of a complex variable  32 Several complex variables  34 Ordinary differential equations 35 Partial differential equations 39 Finite differences and functional equations 40 Sequences, series, summability  41 Approximation and expansion  42 Fourier analysis 43 Abstract harmonic analysis 45 Integral equations 46 Functional analysis 47 Operator theory 49 Calculus of variations and optimal control, optimization	<a href="#">26-xx</a> Real functions <a href="#">28-xx</a> Measure and integration <a href="#">30-xx</a> Functions of a complex variable  <a href="#">32-xx</a> Several complex variables and analytic spaces <a href="#">34-xx</a> Ordinary differential equations <a href="#">35-xx</a> Partial differential equations <a href="#">39-xx</a> Difference and functional equations  <a href="#">40-xx</a> Sequences, series, summability  <a href="#">41-xx</a> Approximations and expansions  <a href="#">42-xx</a> Fourier analysis <a href="#">43-xx</a> Abstract harmonic analysis <a href="#">45-xx</a> Integral equations <a href="#">46-xx</a> Functional analysis <a href="#">47-xx</a> Operator theory <a href="#">49-xx</a> Calculus of variations and optimal control; optimization
Geometry	51 Geometry 52 Convex sets and related geometric topics 53 Differential geometry 54 General Topology 55 Algebraic topology 57 Manifolds and cell complexes	<a href="#">51-xx</a> Geometry <a href="#">52-xx</a> Convex and discrete geometry  <a href="#">53-xx</a> Differential geometry <a href="#">54-xx</a> General <a href="#">55-xx</a> Algebraic topology <a href="#">57-xx</a> Manifolds and cell complexes

	58 Global analysis, analysis on manifolds	<a href="#">58-xx</a> Global analysis, analysis on manifolds
Statistics	60 Probability theory and stochastic processes 62 Statistics 65 Numerical analysis	<a href="#">60-xx</a> Probability theory and stochastic processes <a href="#">62-xx</a> Statistics <a href="#">65-xx</a> Numerical analysis
Applied Mathematics	68 Computer science 70 Mechanics of particles and systems 73 Mechanics of solids  76 Fluid mechanics 78 Optics, electromagnetic theory  80 Classical thermodynamics, heat transfer 81 Quantum theory 82 Statistical mechanics, structure of matter  83 Relativity and gravitational theory 85 Astronomy and astrophysics 86 Geophysics 90 Economics, operations research, programming, games 92 Biology and behavioral sciences 93 Systems theory, control 94 Information and communication, circuit	<a href="#">68-xx</a> Computer science <a href="#">70-xx</a> Mechanics of particles and systems 73-xx This section has been deleted {For mechanics of solids, see <a href="#">74-xx</a> }- klasa usunięta <a href="#">74-xx</a> Mechanics of deformable solids  <a href="#">76-xx</a> Fluid mechanics <a href="#">78-xx</a> Optics, electromagnetic theory  <a href="#">80-xx</a> Classical thermodynamics, heat transfer <a href="#">81-xx</a> Quantum theory <a href="#">82-xx</a> Statistical mechanics, structure of matter  <a href="#">83-xx</a> Relativity and gravitational theory <a href="#">85-xx</a> Astronomy and astrophysics <a href="#">86-xx</a> Geophysics <a href="#">90-xx</a> Operations research, mathematical programming <a href="#">92-xx</a> Biology and other natural sciences <a href="#">93-xx</a> Systems theory; control <a href="#">94-xx</a> Information and communication, circuits

Dane przedstawione w Tab. 5.24 wskazują, poza kilkoma wyjątkami, na pełną zgodność odpowiedników symboli w tablicach MSC i form kategorii tematycznych. Kategorie tematyczne tworzące narzędzie dostępu w systemie przejęto z wykazu symboli klasyfikacyjnych w tablicach MSC na pierwszym stopniu podziału.

W klasyfikacji tej na pierwszym stopniu wyodrębniono 66 klas. [Eilbeck 1996] Tymczasem w MathGuide wykorzystano tylko 58. Do klas pominiętych należą m.in. *08-xx Systemy algebraiczne - ogólnie* (ang. *General algebraic systems*), czy *97-xx Edukacja matematyczna* (ang. *Mathematics education*). Zgodnie z zasadą *literary warrant* [Vizine-Goetz, Beall 2004], kategorie tematyczne są wyodrębniane na podstawie wystąpienia określonego tematu w treści obiektów informacyjnych tworzących kolekcję systemu. Tymczasem wysłanie zapytania, za pomocą wyszukiwarki w systemie, o treści *mathematics education* (ang. *edukacja matematyczna*) spowodowało wyświetlenie przez system dwóch obiektów informacyjnych, które zaindeksowano tym terminem. Tym samym pominięcie kategorii *Edukacja matematyczna*, w strukturze hierarchicznej kategorii tematycznych podyktowane było zapewne innymi względami lub stanowi dowód niekonsekwencji projektantów systemu. O braku konsekwencji w tworzeniu schematu narzędzia dostępu

przedmiotowego świadczy też występowanie kategorii pustych, w których nie zgrupowano żadnej ChWD (są to np. 80 *Termodynamika klasyczna, transfer ciepła* (ang. *Classical thermodynamics, heat transfer*) czy też 82 *Mechanika statystyczna, struktura materii* (ang. *Statistical mechanics, structure of matter*).

Każde wyrażenie przyjęte za nazwę kategorii tematycznej w schemacie stanowi węzeł odesłania hipertekstowego, którego aktywacja powoduje wyświetlenie użytkownikowi listy relewantnych ChWD. Charakterystyki te, wyświetlane są w tzw. formacie uproszczonym. Wizualizacja pełnych danych możliwa jest poprzez aktywację węzła hipertekstowego o nazwie *Metadane*, dołączonego do opisu uproszczonego. Struktura rekordu zawiera oprócz pól przeznaczonych do odwzorowywania cech formalnych obiektu informacyjnego, również sekcję pól w których opisuje się treść dokumentu (Rys. 5.15).

#### Metadata: Singular

Data Source: SUB

BIBLIOGRAPHIC DATA	
<b>Title</b>	Singular
<b>Publisher</b>	Singular Research Group - Department of Mathematics - University of Kaiserslautern: Kaiserslautern, DE
<b>Distributor</b>	University of Kaiserslautern - Department of Mathematics: Kaiserslautern, DE
<b>Language</b>	English
<b>Country (State)</b>	Germany
<b>Format of data</b>	text/html; ps; dvi; TeXinfo; LaTeX; tar; hqx; zip
<b>Keywords</b>	Kaiserslautern; ZCA; Singular; software; grÅ bner base; syzygies; computer algebra
<b>Description</b>	This is the Singular home page. Singular is a software package which has been developed at the centre for computer algebra at the University of Kaiserslautern. It is free and available for almost all platforms. The current version claims to beat commonly used Programs like Macaulay 2 in efficiency. There is an extensive collection of documentation in various formats available. The group cooperates with MuPad and plans to fulfill the MP Standard. Characteristic zero computation is possible. Unfortunately there is no GUI available.
<b>URL</b>	<a href="http://www.singular.uni-kl.de/">http://www.singular.uni-kl.de/</a>
CLASSIFICATION	
<b>Source Type</b>	Software and Expert Systems
<b>MSCverbal</b>	Commutative rings and algebras; Algebraic Geometry; Polynomials and matrices; Explicit machine computation and programs
<b>MSC</b>	13-04; 14-04; 11Cxx
ADDITIONAL INFORMATION	
<b>Access</b>	free; no restriction
<b>Contents</b>	<b>Level</b> graduate; professional
<b>★★★</b>	
STATISTICS	
<b>Backlinks</b>	1240
<b>© SSG-FI</b>	1998-06-26

Rys. 5.15. Struktura rekordu w systemie MathGuide. [Źródło: <http://mathguide.de>]

Na rysunku 5.15 przedstawiono strukturę rekordu w systemie MathGuide. Pola przeznaczone do odwzorowania treści dokumentu to:

- słowa kluczowe (ang. keywords) – niekontrolowane słowa kluczowe,

- klasyfikacja (ang. classification) – sekcja pól, gdzie określa się rodzaj zasobu (ang. source type) oraz identyfikuje wyrażenia reprezentujące treść dokumentu z wykorzystaniem symbolu MSC oraz jego odpowiednika wyrażonego w języku naturalnym.

W wyświetlanej użytkownikowi strukturze rekordu, wyrażenia użyte do odwzorowania treści obiektu informacyjnego nie są węzłami hipertekstowymi, jak to miało miejsce np. w systemie Intute. Pomimo wystąpienia kilku symboli MSC oraz ich odpowiedników słownych, nie istnieje możliwość użycia tych wyrażen jako odesłania hipertekstowego do listy innych dokumentów zaindeksowanych z ich użyciem.

W narzędziu dostępu przedmiotowego, stworzonego na podstawie systematyki klas MSC, na pierwszym stopniu podziału występują kategorie tematyczne o względnie szerokim zakresie znaczeniowym. Tym samym ChWD, w których do indeksowania zastosowano symbole szczegółowe MSC są włączane do kategorii tematycznych w schemacie, które są odpowiednikami jednej z klas MSC na pierwszym stopniu podziału. Tego rodzaju „uogólnianie” zapobiega nadmiernej rozbudowie schematu kategorii tematycznych.

Komplementarnym wobec narzędzia dostępu przedmiotowego w MathGuide jest wykaz punktów dostępu odwzorowujących cechy formalne obiektów informacyjnych. Jest to zbiór wyrażen stosowanych podczas indeksowania do określenia cechy *typ źródła* (ang. resource type). Jest to kontrolowany zbiór wyrażen stosowanych do indeksowania, na podstawie którego zbudowano narzędzie dostępu do zasobów systemu. W czterech kategoriach:

- Dostawcy informacji / Serwey (Information Providers / Servers),
- Informatory bezpośrednie (Factual Reference Works),
- Źródła bibliograficzne (Bibliographic Sources),
- Źródła pierwotne (Primary Information Sources),

zgrupowano 48 wyrażen pełniących funkcję węzłów hipertekstowych, które odsyłają do ChWD spełniających określone kryteria formalnego utrwalenia informacji.

Charakterystyka dwóch narzędzi dostępu do zasobów MathGuide wskazuje na zastosowanie dwóch kategorii wyrażen pełniących funkcje punktów dostępu. Pierwszy zbiór reprezentuje wymiar treści obiektów informacyjnych, drugi zaś wyłącznie ich formę.

## **5.6. Podsumowanie**

Przedstawiona w rozdziale charakterystyka narzędzi dostępu przedmiotowego w wybranych dziedzinowych systemach hipertekstowych miała na celu wskazanie funkcji, jaką pełni w nich język informacyjno-wyszukiwawczy. Dodatkowym celem badań było wskazanie prawidłowości, jakie zachodzą w strukturalizacji punktów dostępu za pomocą zbiorów słownictwa w tego rodzaju narzędziach.

Narzędzia dostępu przedmiotowego w wybranych DSH prezentują z jednej strony typowe rozwiązania dla tej klasy systemów informacyjno-wyszukiwawczych, z drugiej zaś wskazują na ciekawe sposoby adaptacji JIW. Do typowych cech strukturalnych tego rodzaju narzędzi należą alfabetyczne lub hierarchiczne zbiory kategorii tematycznych oraz autonomiczne lub włączone w narzędzie dostępu przedmiotowego alfabetyczne indeksy haseł reprezentujących formę obiektów informacyjnych. Rola JIW w tych narzędziach to przede wszystkim pełnienie funkcji źródła słownictwa do tworzenia nazw dla kategorii tematycznych. Tam, gdzie wykorzystuje się język informacyjno-wyszukiwawczy do projektowania schematów hierarchicznych punktów dostępu, mamy do czynienia z rozbieżnościami w wykorzystaniu jego struktury paradygmatycznej do organizacji pola tematycznego kolekcji.

## **6. Wykorzystanie języka informacyjno-wyszukiwawczego w budowie narzędzi dostępu przedmiotowego**

Przedmiotem rozdziału jest budowa narzędzi dostępu przedmiotowego dziedzinowych systemów hipertekstowych oraz udział w niej elementów strukturalnych języków informacyjno-wyszukiwawczych. Prezentowane w tej części rozprawy treści są wnioskami z badań przeprowadzonych przez autora nad tego rodzaju narzędziami dostępu w DSH. Głównym celem badawczym była ekstrakcja rozwiązań modelowych, jakie mają miejsce w narzędziach dostępu przedmiotowego analizowanej grupy reprezentatywnej oraz wskazanie funkcji i roli, jaką pełni w tego rodzaju narzędziach język informacyjno-wyszukiwawczy.

Analiza materiału badawczego w postaci szkieletów narzędzi dostępu przedmiotowego oddzielonych od ich interfejsów graficznych w grupie reprezentatywnej DSH, pozwoliła na stworzenie trzech wariantów strukturalnych tych narzędzi. Zostały one opracowane na podstawie sposobów uporządkowania punktów dostępu do zasobów, reprezentowanych przez kategorie tematyczne oraz kategorie odwzorowujące formę elementów zbioru informacyjnego. Są to warianty: ahierarchiczny, hierarchiczny oraz hybrydowy. Przedstawione w poprzednim rozdziale przykłady narzędzi dostępu przedmiotowego w DSH uwzględniały już taką kategoryzację. W tym miejscu warianty te zostaną poddane szczegółowej analizie poprzez wskazanie na ich charakterystyczne cechy budowy oraz na udział w ich tworzeniu i funkcjonowaniu języka informacyjno-wyszukiwawczego.

Wyniki badań przeprowadzonych przez autora pozwalają na stwierdzenie, że narzędzia dostępu przedmiotowego w dziedzinowych systemach hipertekstowych pełnią funkcję zbliżoną do katalogów rzeczowych. Ich kompozycja oraz realizacja podstawowych funkcji są uzależnione od środowiska informacyjnego hipertekstu. W zależności od sposobu porządkowania punktów dostępu w postaci haseł katalogowych można je porównywać do katalogów przedmiotowych lub systematycznych (lub działowych). Jednocześnie możliwa jest interpretacja tego typu narzędzi jako indeksów, odpowiednio przedmiotowych lub systematycznych (lub działowych). Różnicą jest tutaj techniczny aspekt realizacji funkcji wyszukiwawczej. W narzędziach dostępu przedmiotowego nie można mówić, ani o bezpośrednim grupowaniu ChWD przy hasłach katalogowych, ani też o umieszczaniu adresów ChWD obok terminów indeksowych. Specyfika hipertekstu umożliwia automatyczne przetwarzanie zbioru wyszukiwawczego i generowanie rezultatów wyszukiwania, czyli



wykazu ChWD bez konieczności fizycznego powiązania haseł katalogowych z określoną grupą opisów.

### **6.1. *Wariant ahierarchiczny***

Budowa narzędzi dostępu przedmiotowego w DSH z wykorzystaniem ahierarchicznej (płaskiej) organizacji punktów dostępu polega na generowaniu wykazów tych wyrażeń z odpowiednich pól rekordów bazy danych systemu. Są to zbiory jednostek leksykalnych tworzących indeks alfabetyczny, którego elementy są wejściami do zbioru wyszukiwawczego. Każde z nich jest węzłem dla odesłania hipertekstowego, którego aktywacja powoduje wyświetlenie listy charakterystyk wyszukiwawczych obiektów informacyjnych, w których użyto je do zaindeksowania.

Ahierarchiczna budowa tego wariantu narzędzi cechuje organizację „[...] wyrażeń, w której nie wykorzystuje się żadnej relacji hierarchicznej, a pojęcia reprezentowane przez jednostki leksykalne są traktowane jako wzajemnie współrzędne. Strukturami ahierarchicznymi są zatem wszelkie równoważnościowe struktury monorelacyjne”. [Sosińska-Kalata 1999 s. 166]

Wykorzystanie JIW do generowania struktur ahierarchicznych polega na traktowaniu słownika wybranego JIW jako źródła leksyki bezpośrednio do indeksowania obiektów informacyjnych, a pośrednio do tworzenia nazw dla punktów dostępu. Zastosowanie określonego typu JIW do odwzorowania treści i formy obiektów informacyjnych w systemie, określa formę jego wyrażeń w omawianym wariantcie narzędzia dostępu przedmiotowego. Najczęściej w tym celu wykorzystuje się JIW o notacji paranaturalnej, czyli język słów kluczowych, język haseł przedmiotowych oraz tezaurus jako słownik języka deskryptorowego. Jest to podyktowane względami praktycznymi, tj. możliwością wykorzystania jednostek leksykalnych w formie równokształtnej z wyrażeniami języka naturalnego do bezpośredniego oznaczania zakresu nazwy punktu dostępu. Zastosowanie zasobu leksykalnego JIW o notacji sztucznej do generowania tego rodzaju indeksów wiązałoby się z koniecznością opracowania haseł indeksowych w języku naturalnym w celu ułatwienia interpretacji zakresu znaczeniowego punktów dostępu dla użytkownika.

W tego rodzaju narzędziach występują wyrażenia reprezentujące zazwyczaj dwie kategorie cech obiektów informacyjnych. Są to treść oraz forma elementów zbioru informacyjnego. Budowa tych narzędzi może uwzględniać jednoczesne występowanie dwóch wspomnianych kategorii punktów dostępu lub zakładać funkcjonowanie

osobnych indeksów stanowiących narzędzie dostępu do zasobów systemu z uwzględnieniem kryterium treściowego lub formalnego. Tym samym biorąc pod uwagę rodzaj informacji o zbiorze informacyjnym reprezentowaną przez hasła indeksowe, w wariacie ahierarchicznym można wyróżnić indeksy rzeczowe, formalne oraz mieszane.

### **6.1.1. Indeksy rzeczowe**

Indeksy rzeczowe to umowna nazwa dla typu narzędzi dostępu przedmiotowego w omawianym wariacie, których hasła odsyłają do opisów obiektów informacyjnych według kryterium treściowego. Oznacza to, że skierowują do ChWD poprzez wskazanie na ich temat lub przedmiot.

Ahierarchiczny wariant dostępu przedmiotowego z użyciem haseł odwzorowujących treść obiektów informacyjnych cechuje zastosowanie zasobu leksykalnego języka niekontrolowanych słów kluczowych lub zbiorów słownictwa kontrolowanego o notacji paranaturalnej. W praktyce nie stosuje się indeksów niekontrolowanych słów kluczowych jako jedyne narzędzie dostępu do informacji w DSH. Indeksy rzeczowe generowane na podstawie zbioru niekontrolowanych słów kluczowych są zazwyczaj wykorzystywane jako dodatkowe narzędzie dostępu do zasobów.\*

Najbardziej rozpowszechnionym sposobem budowania narzędzi dostępu przedmiotowego w omawianym wariacie jest wykorzystanie kontrolowanego słownika zasobu leksykalnego wybranego JIW o notacji paranaturalnej. Zasada tworzenia punktów dostępu pozostaje niezmienna.\*\* Jego wybór przez użytkownika, z wykorzystaniem techniki przeglądania, powoduje realizację jego funkcji wyszukiwawczej poprzez wyodrębnienie w zbiorze wyszukiwawczym takich ChWD, w których pełni wskazaną funkcję.

Budowa tego rodzaju indeksów, w zależności od ich wielkości, zakłada wyświetlenie użytkownikowi pełnej listy wyrażen (Rys. 6.1.) lub jednostek zgrupowanych w sekcjach, których nazwy stanowią litery alfabetu łacińskiego (Rys. 6.2.).

---

\* Taka sytuacja ma miejsce np. w systemie Infomine (<http://infomine.ucr.edu/>), w którym dostęp do zasobów realizowany jest za pomocą indeksu opartego na Library of Congress Subject Headings, a jako narzędzie pomocnicze wykorzystuje się indeks niekontrolowanych słów kluczowych.

\*\* Każde wyrażenie pochodzące ze słownika stosowanego w systemie JIW, które użyto w ChWD do odwzorowania jego treści staje się punktem dostępu w indeksie.

Please take our 2006 HealthWeb user survey

<a href="#">AIDS &amp; HIV</a>	<a href="#">Health Informatics</a>	<a href="#">Pediatrics</a>
<a href="#">Allied Health</a>	<a href="#">Hematology</a>	<a href="#">Pharmacy &amp; Pharmacology</a>
<a href="#">Alternative Medicine</a>	<a href="#">Immunology</a>	<a href="#">Physical Medicine &amp; Rehabilitation</a>
<a href="#">Anatomy</a>	<a href="#">Laboratory Medicine &amp; Pathology</a>	<a href="#">Physiology</a>
<a href="#">Anesthesiology</a>	<a href="#">Men's Health</a>	<a href="#">Preventive Medicine</a>
<a href="#">Biochemistry</a>	<a href="#">Microbiology &amp; Infectious Diseases</a>	<a href="#">Psychiatry/Psychology</a>
<a href="#">Bioethics</a>	<a href="#">Minority Health</a>	<a href="#">Public Health</a>
<a href="#">Cardiology</a>	<a href="#">Molecular Biology</a>	<a href="#">Pulmonary Medicine</a>
<a href="#">Cell Biology</a>	<a href="#">Mortuary Science</a>	<a href="#">Radiology</a>
<a href="#">Chiropractic</a>	<a href="#">Neuroanatomy</a>	<a href="#">Reference Resources</a>
<a href="#">Consumer Health</a>	<a href="#">Neurology</a>	<a href="#">Rheumatology</a>
<a href="#">Dentistry</a>	<a href="#">Nuclear Medicine</a>	<a href="#">Rural Health</a>
<a href="#">Dermatology</a>	<a href="#">Nursing</a>	<a href="#">Speech-Language Pathology &amp; Audiology</a>
<a href="#">Developmental Disabilities</a>	<a href="#">Nutrition</a>	<a href="#">Sports Medicine</a>
<a href="#">Diabetes</a>	<a href="#">Obstetrics &amp; Gynecology</a>	<a href="#">Substance Abuse</a>
<a href="#">Emergency Medicine &amp; Critical Care</a>	<a href="#">Occupational Therapy</a>	<a href="#">Surgery</a>
<a href="#">Endocrinology</a>	<a href="#">Oncology</a>	<a href="#">Telemedicine</a>
<a href="#">Evidence Based Health Care</a>	<a href="#">Ophthalmology</a>	<a href="#">Toxicology</a>
<a href="#">Family Practice/Primary Care</a>	<a href="#">Optometry</a>	<a href="#">Transplantation</a>
<a href="#">Gastroenterology</a>	<a href="#">Orthopedics</a>	<a href="#">Urban Health</a>
<a href="#">Genetics</a>	<a href="#">Osteopathic Medicine</a>	<a href="#">Urology</a>
<a href="#">Geriatrics &amp; Gerontology</a>	<a href="#">Otolaryngology</a>	<a href="#">Veterinary Medicine</a>
<a href="#">Health Administration</a>	<a href="#">Palliative and Hospice Care</a>	<a href="#">Women's Health</a>

Rys.6.1. Dostęp do informacji z wykorzystaniem struktury płaskiej w HealthWeb.

[Źródło: <http://healthweb.org/>]

<b>Subject links</b>	<b>A B C D E F G H I J K L M N O P Q R S T U V W X Y Z</b>
<b>About us</b> <input type="button" value="▶"/>	<a href="#">abdomen</a> [1]
<b>A-Z of services</b> <input type="button" value="▶"/>	<a href="#">ablation</a> [1]
<b>Internet catalogue</b>	<a href="#">abnormal behaviour</a> [3]
<b>Advanced search</b> <input type="button" value="▶"/>	<a href="#">abnormalities</a> [4]
<b>Browse</b> <input type="button" value="▶"/>	<a href="#">abomasum</a> [2]
<b>New resources</b> <input type="button" value="▶"/>	<a href="#">aborigines</a> [1]
<b>Suggest a site</b> <input type="button" value="▶"/>	<a href="#">abortion</a> [5]
<b>Internet training</b>	<a href="#">abuse</a> [3]
<b>Virtual Training Suite</b> <input type="button" value="▶"/>	<a href="#">academic standards</a> [1]
<b>Support materials</b> <input type="button" value="▶"/>	<a href="#">Acanthocephala</a> [1]
<b>Intute events</b> <input type="button" value="▶"/>	<a href="#">accident prevention</a> [1]
<b>Additional services</b>	<a href="#">accidents</a> [1]
	<a href="#">acne</a> [1]
	<a href="#">Actinobacillus</a> [1]
	<a href="#">actinomycosis</a> [1]
	<a href="#">acupuncture</a> [4]
	<a href="#">acute diseases</a> [1]
	<a href="#">acute phase proteins</a> [1]

Rys. 6.2. Dostęp do informacji z wykorzystaniem struktury płaskiej w systemie Intute: weterynaria.

[Źródło: <http://intute.ac.uk/>]

Często stosowanym rozwiązaniem jest umieszczanie obok wyrażenia pełniącego

funkcję wyszukiwawczą wskaźnika informującego o liczbie charakterystyk wyszukiwawczych obiektów informacyjnych, do których odsyła (Rys. 6.2).

W zależności od rodzaju języka informacyjnego stosowanego w systemie, indeksy rzeczowe w DSH zbudowane są z alfabetycznego wykazu jednego rodzaju wyrażeń. Są to:

- słowa kluczowe,
- hasła przedmiotowe,
- deskryptory,
- odpowiedniki słowne symboli klasyfikacyjnych.

Charakterystyczne zastosowanie języka haseł przedmiotowych do indeksowania elementów zbioru informacyjnego przekłada się również na specyficzną budowę indeksów rzeczowych, wygenerowanych na ich podstawie. Zastosowanie języka haseł przedmiotowych do budowy tego typu narzędzi przyjmuje dwie formy. W pierwszej, funkcję organizującą i wyszukiwawczą pełnią wyłącznie tematy jako jednostki autosyntaktyczne (Rys.6.3).

**intute : health & life sciences** [Intute Home](#) [Contacts](#) [Help desk](#) [Site map](#)

[Quick links](#)

[Medicine](#) | [Nursing, Midwifery and Allied Health](#) | [Veterinary](#) | [Bioresearch](#) | [Natural History](#) | [Agriculture, Food and Forest](#)  
[BioethicsWeb](#) | [MedHist](#) | [Psci-com](#)

●● Intute > [Health and Life Sciences](#) > [BioethicsWeb](#) > Browse by MeSH headings

Subject links	
About us	▶
A-Z of services	▶
Internet catalogue	
Advanced search	▶
Browse	▶
New resources	▶
Suggest a site	▶

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

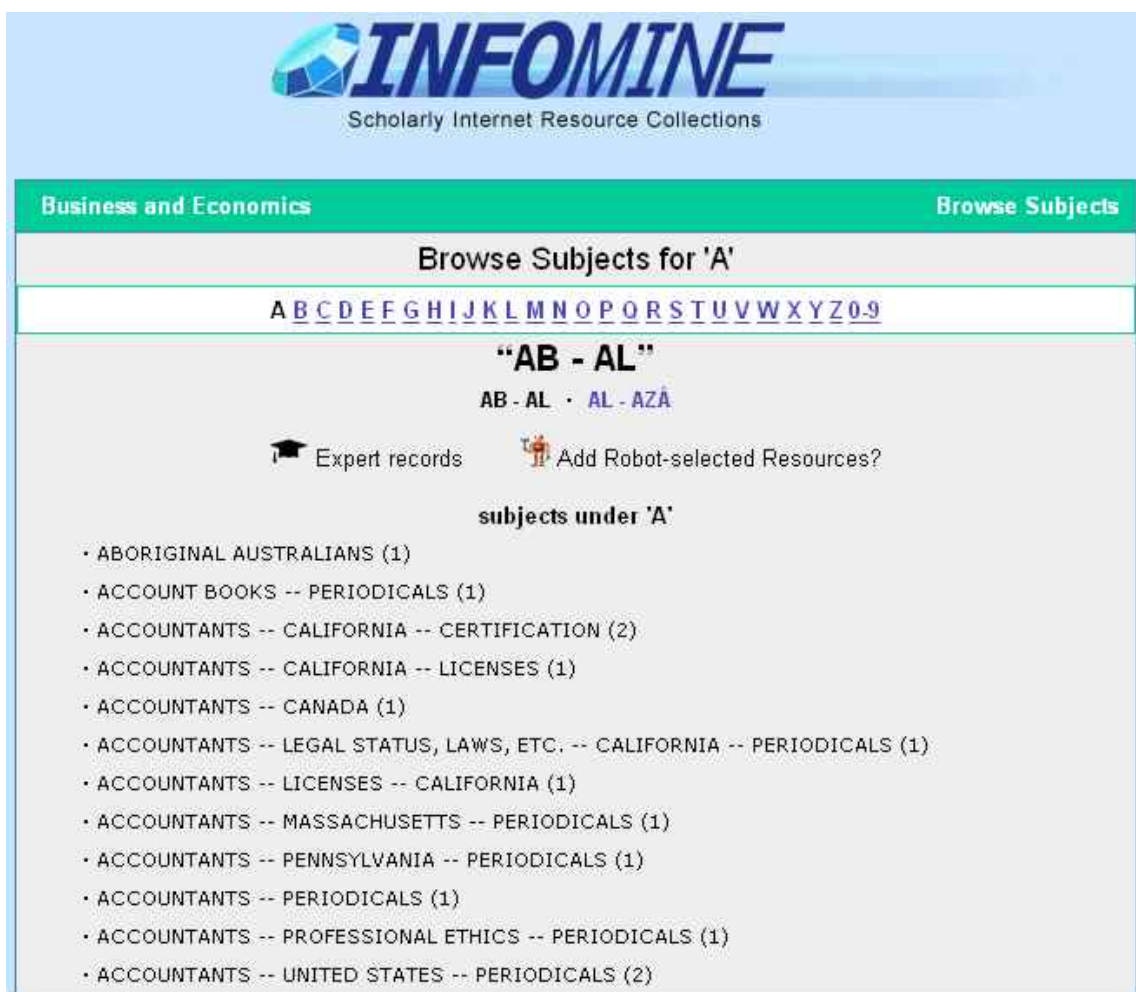
Teaching Materials [2]  
Technology [1]  
Tissue Banks [27]  
Transplantation [12]  
Transplantation, Heterologous [26]

Rys. 6.3. Tematy JHP jako autosyntaktyczne jednostki organizujące w systemie Intute: zdrowie i nauki o życiu – Bioetyka. [Źródło: <http://intute.ac.uk/>]

Na rysunku 6.3 przedstawiono fragment indeksu rzeczowego w ahierarchicznym wariacie narzędzia dostępu przedmiotowego w systemie BiotethicsWeb. Wchodzi on w skład sieci Intute: zdrowie i nauki o życiu. Narzędzie powstało poprzez wygenerowanie zbioru wyrażeń MeSH zastosowanych tutaj do indeksowania kolekcji. Treść obiektów informacyjnych reprezentują tutaj wyłącznie jednostki autosyntaktyczne tego języka informacyjnego, czyli tematy oraz tematy z dopowiedzeniem

(np. identyfikującym, objaśniającym, wskazującym). W tym przypadku funkcję wyszukiwawczą pełni temat wraz z kwalifikatorem tworząc węzeł dla powiązania hipertekstowego.

Drugim sposobem wykorzystania języka haseł przedmiotowych do budowy narzędzi dostępu przedmiotowego reprezentujących ten wariant to zastosowanie pełnych haseł przedmiotowych, czyli tematu wraz z grupą określników (Rys. 6.4).



The screenshot shows the Infomine website interface. At the top, there is a logo for 'INFOMINE Scholarly Internet Resource Collections'. Below the logo, there are two navigation tabs: 'Business and Economics' and 'Browse Subjects'. The main content area is titled 'Browse Subjects for 'A'' and features a navigation bar with letters A through Z and 0-9. Below this, there is a section for 'AB - AL' with sub-links for 'AB - AL' and 'AL - AZA'. There are also two icons: a graduation cap for 'Expert records' and a robot for 'Add Robot-selected Resources?'. The main list of subjects under 'A' includes:

- ABORIGINAL AUSTRALIANS (1)
- ACCOUNT BOOKS -- PERIODICALS (1)
- ACCOUNTANTS -- CALIFORNIA -- CERTIFICATION (2)
- ACCOUNTANTS -- CALIFORNIA -- LICENSES (1)
- ACCOUNTANTS -- CANADA (1)
- ACCOUNTANTS -- LEGAL STATUS, LAWS, ETC. -- CALIFORNIA -- PERIODICALS (1)
- ACCOUNTANTS -- LICENSES -- CALIFORNIA (1)
- ACCOUNTANTS -- MASSACHUSETTS -- PERIODICALS (1)
- ACCOUNTANTS -- PENNSYLVANIA -- PERIODICALS (1)
- ACCOUNTANTS -- PERIODICALS (1)
- ACCOUNTANTS -- PROFESSIONAL ETHICS -- PERIODICALS (1)
- ACCOUNTANTS -- UNITED STATES -- PERIODICALS (2)

Rys. 6.4. Organizacja informacji w systemie Infomine z wykorzystaniem języka haseł przedmiotowych.

[Źródło: <http://infomine.ucr.edu/cgi-bin/search?category=bioag>]

Na rysunku 6.4. przedstawiono fragment indeksu rzeczowego w ahierarchicznym wariacie narzędzia dostępu przedmiotowego w systemie Infomine. Tutaj również sposób wykorzystania JIW do indeksowania wyznacza możliwości adaptacji jego wyrażeń jako punktów dostępu. Przedstawione na Rys. 6.4 narzędzie dostępu przedmiotowego wykorzystuje zasób leksykalny oraz reguły gramatyczne języka haseł przedmiotowych Biblioteki Kongresu - Library Of Congress Subject Headings (LCSH). Tutaj funkcję porządkującą i wyszukiwawczą pełni całe wyrażenie, czyli temat wraz

z grupą określników. Cechą tego typu rozwiązań jest widoczna na rysunku 6.4. prekoordynacja, czyli w tym przypadku wykorzystanie rozbudowanych haseł przedmiotowych do grupowania odesłań do ChWD. Takie rozwiązanie powoduje, że jest to narzędzie statyczne, w porównaniu chociażby ze schematami hierarchicznymi.\*

U podstaw ahierarchicznych schematów organizacji punktów dostępu wykorzystujących hasła przedmiotowe leżą reguły gramatyczne porządkowania grup określników w tego typu językach informacyjnych. Podstawowym elementem porządkującym jest tutaj forma tematu, zaś kolejność w grupie określników w dalszej rozbudowie hasła wyznacza miejsce odesłania w porządku alfabetycznym.

Wykorzystanie języka deskryptorowego do generowania narzędzi dostępu indeksów rzeczowych w wariacie ahierarchicznym polega na wykorzystaniu tezauryasa jako źródła leksyki. Elementy słownika tego typu języka informacyjnego są tak, jak we wcześniej analizowanych przypadkach, traktowane jako punkty dostępu do zbioru wyszukiwawczego. Warunkiem otrzymania statusu punktu dostępu jest użycie wyrażenia ze słownika do indeksowania w ChWD (Rys. 6.5).

---

\* Środowisko hipertekstu umożliwia wprowadzenie elementu „dynamicznego” konstruowania punktu dostępu poprzez np. dołączanie do zakresu tematu wyrażeń reprezentujących dodatkowe fasy, czy kategorie leksykalne. Przykładem takiego rozwiązania jest narzędzie dostępu przedmiotowego w systemie GEM (Rozdz. 5.3)

Subject links	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
<b>About us</b>	<a href="#">Ear</a> [10]
<b>A-Z of services</b>	<a href="#">Ear Diseases</a> [6]
<b>Internet catalogue</b>	<a href="#">Ear Nose and Throat</a> [9]
<b>Advanced search</b>	<a href="#">Ear Nose and Throat Diseases</a> [1]
<b>Browse</b>	<a href="#">Eating Disorders</a> [16]
<b>New resources</b>	<a href="#">Ectopic Pregnancy</a> [10]
<b>Suggest a site</b>	<a href="#">Education</a> [23]
<b>Internet training</b>	<a href="#">Educational Assessment</a> [1]
<b>Virtual Training Suite</b>	<a href="#">Educational Economics</a> [1]
<b>Support materials</b>	<a href="#">Educational Grants</a> [2]
<b>Intute events</b>	<a href="#">Educational Guidance</a> [1]
<b>Additional services</b>	<a href="#">Educational Institutions</a> [10]
<b>Hot topics</b>	<a href="#">Educational Management</a> [2]
<b>Timelines</b>	<a href="#">Educational Performance</a> [1]
<b>MedHist events</b>	<a href="#">Educational Policy</a> [4]
<b>Psci-com events</b>	<a href="#">Educational Research</a> [3]
<b>Psci-com books and articles</b>	<a href="#">Educational Standards</a> [5]
	<a href="#">Educational Technology</a> [1]
	<a href="#">Efficiency</a> [1]
	<a href="#">Elbow</a> [2]
	<a href="#">Electrotherapy</a> [6]
	<a href="#">Electrothermy</a> [1]
	<a href="#">Embryo Transfer</a> [1]
	<a href="#">Embryology</a> [2]
	<a href="#">Emergency Health Services</a> [4]
	<a href="#">Emergency Medical Units</a> [1]

Rys. 6.5. Wykorzystanie tezausa Royal College of Nursing Thesaurus w systemie Intute: zdrowie i nauki o życiu – Pielęgniarstwo. [Źródło: <http://intute.ac.uk/>]

Zastosowanie tezausa w konstrukcji ahierarchicznych wariantów organizacji narzędzia dostępu zakłada wykorzystanie wyłącznie deskryptorów pozbawionych modyfikatorów oraz wskaźników semantycznych.

W przedstawionym na rysunku 6.5 narzędziu dostępu do zasobów systemu Pielęgniarstwo, położnictwo, opieka zdrowotna (ang. NMAP: Nursing, Midwifery and Allied Health) w Intute: zdrowie i nauki o życiu wygenerowano alfabetyczny wykaz wyrażen pochodzących z tezausa Royal College of Nursing, które użyto do indeksowania zbioru systemu. Funkcję organizującą oraz wyszukiwawczą pełnią tutaj wyłącznie deskryptory. Nie stosuje się odsyłaczy od wyrażen posiadających status askryptora. W celu poprawienia efektywności wyszukiwania mogłyby one zostać wykorzystane i pełnić funkcję pustych punktów dostępu skierowujących do wyrażen posiadających pełny status wyszukiwawczy.

Rzadko spotykanym rodzajem indeksów rzeczowych w wariacie płaskim jest wykorzystanie do tego celu języków informacyjnych typu klasyfikacji. Narzędzie dostępu przedmiotowego przyjmuje wówczas formę wykazu odpowiedników słownych

symboli klasyfikacyjnych, którymi scharakteryzowano treść dokumentów w procesie indeksowania. Taka sytuacja miała miejsce we wspomnianym systemie NMAP przed włączeniem go w struktury systemu Intute. System ten stosował wówczas jako punkty dostępu odpowiedniki słowne symboli klasyfikacji National Library of Medicine w układzie alfabetycznym. Wyrażenia te w przeważającej większości były równokształtne z odpowiednikami słownymi symboli w języku naturalnym zamieszczonymi w tablicach tej klasyfikacji. Obecnie narzędzia dostępu przedmiotowego w NMAP zbudowane są z dwóch indeksów reprezentujących wariant ahierarchiczny. Pierwsze zostało wygenerowane na podstawie wykorzystanych do indeksowania deskryptorów teaurusu Royal College of Nursing Thesaurus, drugie - to alfabetyczny wykaz tematów Medical Subject Headings. Jest to klasyczny dla DSH przykład polireprezentacji wiedzy, która przekłada się na budowę narzędzi dostępu przedmiotowego.

Do dodatkowych cech treściowych odwzorowywanych przez punkty dostępu w omawianym wariantcie należą tematy jednostkowe w postaci nazw własnych. Należą do nich m.in. tematy osobowe, korporatywne oraz geograficzne. Dla wymienionych trzech kategorii tematów stosuje się odrębną metodę kontroli słownictwa. Polega to na wykorzystaniu dostępnych kartotek wzorcowych normalizujących formy językowe wartości wskazanych cech, np. w systemie ADAM wchodzącym w skład Intute (z zakresu architektury, mediów i sztuki) do odwzorowania tematu osobowego (twórca, artysta) oraz korporatywnego wykorzystuje się kartotekę wzorcową nazw osobowych Union List of Artist Names. System MedHist również wchodzący w skład Intute (z zakresu historii medycyny) stosuje osobny indeks nazw osobowych, gdzie źródłem słownictwa jest kartoteka nazw osobowych Library of Congress Name Authority File (Rys. 6.6). Do najczęściej wykorzystywanych kartotek wzorcowych stanowiących źródło form językowych dla tematów geograficznych należy The Getty Thesaurus of Geographic Names (TGN), dostępny w formacie XML oraz hipertekstowej autonomicznej bazy danych dostępnej poprzez WWW:

[http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)



<b>Subject links</b>	<a href="#">Albertus Magnus, Saint 1193?-1280</a> [1]
<b>About us</b>	<a href="#">Alessandri, Francesco Degli 1529-1587</a> [1]
<b>A-Z of services</b>	<a href="#">Alibert, J. L. (Jean Louis Marie) 1768-1837</a> [1]
	<a href="#">Archimedes</a> [1]
<b>Internet catalogue</b>	<a href="#">Aretaeus, of Cappadocia</a> [1]
<b>Advanced search</b>	<a href="#">Aristotle 384-322 B.C.</a> [7]
<b>Browse</b>	<a href="#">Avery, Oswald T. (Oswald Theodore) 1877-1955</a> [1]
<b>New resources</b>	<a href="#">Avicenna 980-1037</a> [5]
<b>Suggest a site</b>	<a href="#">Axelrod, Julius 1912-</a> [1]
	<a href="#">Bacon, Francis 1561-1626</a> [1]
	<a href="#">Baldwin, James Mark 1861-1934</a> [1]
	<a href="#">Ballard, Martha 1735-1812</a> [1]
<b>Internet training</b>	<a href="#">Banting, Frederick Grant, Sir 1891-1941</a> [5]
<b>Virtual Training Suite</b>	<a href="#">Barnes, David S.</a> [1]
<b>Support materials</b>	<a href="#">Berkeley, George 1685-1753</a> [1]
<b>Intute events</b>	<a href="#">Best, Charles Herbert 1899-1978</a> [2]
	<a href="#">Bevan, Aneurin 1897-1960</a> [1]
	<a href="#">Bichat, Xavier 1771-1802</a> [2]
<b>Additional services</b>	<a href="#">Biruni, Aburalkhan Mukhamed ben-Akhmed 973?-1048</a> [1]
<b>Hot topics</b>	<a href="#">Bisseau, Michel 1676-1743</a> [1]
<b>Timelines</b>	<a href="#">Blackwell, Elizabeth 1821-1910</a> [3]
<b>MedHist events</b>	<a href="#">Blackwell, Elizabeth fl.1737</a> [2]
<b>Psci-com events</b>	<a href="#">Boccaccio, Giovanni 1313-1375</a> [1]
<b>Psci-com books and articles</b>	<a href="#">Boerhaave, Hermann 1668-1738</a> [1]
	<a href="#">Booth, Charles 1840-1916</a> [1]
	<a href="#">Boyer, Paul S.</a> [1]
	<a href="#">Brahe, Tycho 1546-1601</a> [6]

Rys. 6.6. Library of Congress Name Authority File jako źródło słownictwa w systemie MedHist.

[Źródło: <http://intute.ac.uk/>]

Tym samym zastosowanie kartoteki haseł wzorcowych do indeksowania umożliwia zbudowanie, na podstawie zgromadzonych jednostek leksykalnych, indeksu alfabetycznego pełniącego funkcję dodatkowego narzędzia dostępu do zasobów systemu.

### 6.1.2. Indeksy formalne

Nazwa *indeksy formalne*, tak jak *indeksy rzeczowe* ma również charakter umowny, a jej użycie ma na celu wskazanie na kategorię cech reprezentowaną przez hasła indeksowe. Do cech pozat treściowych odwzorowywanych w narzędziach wariantu ahierarchicznego DSH należą:

- autor obiektu informacyjnego,
- tytuł obiektu informacyjnego,
- rodzaj obiektu informacyjnego (rodzaj zasobów).

Forma indeksów wyznaczana jest poprzez porządek alfabetyczny wyrażen przyjętych za

nazwy punktów dostępu. Narzędzia te, tak jak wszystkie w wariacie ahierarchicznym są generowane ze zbioru wartości określonego pola w ChWD. Tym samym kontrola semantyczno-strukturalna wyrażen stosowanych podczas odwzorowywania wskazanych atrybutów określa sposoby adaptacji zbiorów słownictwa, w tym słowników JIW do budowy tego rodzaju indeksów.

Kategoria cech pozatreściowych reprezentowanych przez zbiory słownictwa kontrolowanego w ahierarchicznym wariacie organizacji punktów dostępu przyjmuje formę osobnych wykazów jednostek leksykalnych odwzorowujących wartości tych cech lub ich włączenie do wykazu wyrażen reprezentujących wymiar treści dokumentów internetowych. Najczęściej stosowanym rozwiązaniem w dziedzinowych systemach hipertekstowych, realizujących omawiany wariant organizacji punktów dostępu, jest połączenie jednostek leksykalnych stanowiących punkty dostępu do zbioru wyszukiwawczego w jednym narzędziu dostępu przedmiotowego.

Wartości cech autor/autorzy oraz tytuł obiektu informacyjnego stanowią stosunkowo rzadko podstawę do generowania narzędzi dostępu przedmiotowego. Wykorzystanie tych atrybutów wiąże się z koniecznością kontroli formy wyrażen odwzorowujących te cechy. W przypadku narzędzi dostępu wykorzystujących te atrybuty, ich forma jest kontrolowana odpowiednimi kartotekami wzorcowymi. Taka sytuacja ma miejsce szczególnie w dużych DSH, np. w Infomine oraz CISMef. Stosuje się tam rzadko wykorzystywane narzędzie dostępu do zasobów w postaci indeksu tytułów (Rys. 6.7).



Rys. 6.7. Indeks tytułów w systemie Infomine. [Źródło: <http://infomine.ucr.edu/cgi-bin/search?category=bioag>]

Wyrażenia stanowiące nazwy tytułów obiektów informacyjnych są uporządkowane w układzie alfabetycznym w podziale na sekcje, których nazwom odpowiadają litery alfabetu łacińskiego lub ich odpowiednie zakresy (np. Ca-Co) - Rysunek 6.7.

Tak, jak w przypadku tematów osobowych, które wystąpiły jako hasła opisu rzeczowego, również wartość cechy dla atrybutu autor/autorzy dokumentów może stanowić punkt dostępu w tego rodzaju indeksach. W tym celu najczęściej wykorzystywanym źródłem słownictwa w DSH jest kartoteka wzorcowa nazw osobowych Biblioteki Kongresu (Library of Congress Name Authority File), jednak w narzędziach dostępu przedmiotowego w DSH w zasadzie nie stosuje się indeksów autorskich. Do nielicznych wyjątków należy system Infomine, w którym wygenerowano również indeks autorski jako narzędzie dostępu do zasobów.

Wymiar cech pozatrześciowych reprezentowanych przez kategorię *rodzaj zasobów* jest najczęściej stosowanym kryterium dostępu poprzez indeksy formalne w wariacie ahierarchicznym. Tutaj również wartości tego atrybutu, przechowywane w zbiorze ChWD, stanowią podstawę do generowania tego typu indeksów. Forma

językowa wyrażen przyjętych w danym systemie jako wartości omawianej cechy jest kontrolowana za pomocą zamkniętego zbioru wyrażen języka naturalnego. W większości przypadków jest to lista wybranych wyrażen języka naturalnego służąca do identyfikacji wskazanej wartości atrybutu podczas indeksowania i wykorzystana do budowy tego typu narzędzia. W systemie MathGuide z zakresu matematyki składa się ona z 48 wyrażen wyposażonych w odesłania hipertekstowe (Rys. 6.8).



Rys. 6.8 Dostęp do zbioru w systemie MathGuide z wykorzystaniem indeksu rodzaj zasobu. [Źródło: <http://mathguide.de>]

Przedstawione na rys. 6.8 narzędzie dostępu do zbioru wyszukiwawczego uwzględniające wartość atrybutu rodzaj zasobu organizuje wspomniane 48 wyrażen za pomocą czterech kategorii nadrzędnych:

- Dostawcy informacji / Serwery (Information Providers / Servers),
- Informatory bezpośrednie (Factual Reference Works),

- Źródła bibliograficzne (Bibliographic Sources),
- Źródła pierwotne (Primary Information Sources).

Należy tutaj zauważyć stosowanie jako wartości omawianej cechy wyrażenia *Other* (*Inny rodzaj zasobu*). Jest ono nazwą kategorii grupującej dokumenty internetowe reprezentujące rodzaj zasobu nieodwzorowany przez którekolwiek z pozostałych wyrażeń z listy. Przy czym jej efektywność w narzędziu jest zdaniem autora dyskusyjna.

Indeks rodzajów zasobu tworzy alfabetyczny wykaz wyrażeń będących węzłami dla powiązań hipertekstowych (Rys. 6.9).

**Liste des types de ressource présents dans Doc'CISMeF**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

[voir aussi la [liste annotée et hiérarchisée](#) des types de ressources]

Terme	Nombre de résultats
<a href="#">abdomen sans préparation</a>	8
<a href="#">absorptiométrie photonique</a>	0
<a href="#">actes de congrès</a>	142
<a href="#">âge dentaire</a>	0
<a href="#">âge osseux</a>	0
<a href="#">agence gouvernementale</a>	23
<a href="#">agence régionale</a>	74
<a href="#">algorithme</a>	353
<a href="#">amélioration image radiographique</a>	47
<a href="#">angiocardiographie</a>	0
<a href="#">angiographie</a>	35
<a href="#">angiographie cérébrale</a>	0
<a href="#">angiographie isotopique</a>	0
<a href="#">angiographie numérisée</a>	0
<a href="#">angiographie résonance magnétique</a>	1
<a href="#">annuaire</a>	26
<a href="#">aortographie</a>	0
<a href="#">apprentissage du raisonnement clinique</a>	8
<a href="#">apprentissage par problème</a>	18
<a href="#">arthrographie</a>	0
<a href="#">article de périodique</a>	1510
<a href="#">assistance par téléphone</a>	102
<a href="#">association</a>	2067
<a href="#">association patients</a>	951
<a href="#">association professionnels santé</a>	738
<a href="#">atlas</a>	21

Rys. 6.9. Indeks alfabetyczny rodzaju zasobu w systemie CISMeF.

[Źródło: <http://www.chu-rouen.fr/cismef/>]

W zależności od szczegółowości indeksowania lista wyrażeń stanowiących punkty dostępu według omawianego kryterium formalnego, może mieć różną wielkość. W jednym z największych dziedzinowych systemów hipertekstowych – Intute, zbiór jednostek leksykalnych odwzorowujących cechę rodzaju zasobów liczy 71 pozycji.

[Hiom 2007] W narzędziu dostępu przedmiotowego w systemie Intute: zdrowie i nauki o życiu – Medycyna, źródłem słownictwa do tworzenia nazw dla tej kategorii punktów dostępu są odpowiednie wyrażenia ze słownika MeSH. W systemie CISMef wartość atrybutu *rodzaj zasobów* przyjmuje formę listy słownictwa kontrolowanego stosowanego do tego celu w bazie Medline (Medline Resource Type). [Charles 2005]

### 6.1.3. Indeksy mieszane

Indeksy mieszane w ahierarchicznym wariacie narzędzi dostępu przedmiotowego pełnią funkcje zbliżone do indeksów krzyżowych stosowanych np. w bibliografiach. Tego typu narzędzia łączą w jednym wykazie alfabetycznym hasła z indeksów rzeczowych oraz punkty dostępu wskazujące na typ zasobów. Jest to bardzo często praktykowany sposób budowy tego rodzaju narzędzi. Charakterystyczną cechą jest w tym przypadku stosowanie wykładnika dla cechy *rodzaj zasobu*. Polega to na umieszczaniu obok punktu dostępu reprezentującego tę kategorię wyrażen odpowiedniego kwalifikatora. Jego forma jest zależna od języka interfejsu użytkownika w systemie i sprowadza się do użycia w odpowiedniej formie wyrażenia *typ publikacji* (np. ang. *publication type*).

Taki typ narzędzia dostępu przedmiotowego zastosowano m.in. w klasie głównej Medycyna w systemie Intute: zdrowie i nauki o życiu. Ma ono formę indeksu rzeczowego, w którym obok tematów MeSH (występujących także z określnikami rzeczowymi) zastosowano punkty dostępu z kwalifikatorem [*Publication Type*] wskazującym na reprezentację formy obiektów informacyjnych przez wybrane wyrażenie. (Rys. 6.10)

Subject links	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
<b>About us</b> 	<a href="#">Abbreviations</a> [2]
<b>A-Z of services</b> 	<a href="#">Abbreviations [Publication Type]</a> [2]
<b>Internet catalogue</b>	<a href="#">Abdomen</a> [9]
<b>Advanced search</b> 	<a href="#">Abdominal Injuries</a> [2]
<b>Browse</b> 	<a href="#">Abdominal Muscles / anatomy &amp; histology</a> [1]
<b>New resources</b> 	<a href="#">Abdominal Muscles / physiology</a> [1]
<b>Suggest a site</b> 	<a href="#">Abdominal Neoplasms / case studies</a> [1]
<b>Internet training</b>	<a href="#">Abdominal Pain</a> [2]
<b>Virtual Training Suite</b> 	<a href="#">Abdominal Pain / case studies</a> [1]
<b>Support materials</b> 	<a href="#">Abnormalities</a> [11]
<b>Intute events</b> 	<a href="#">Abnormalities, Multiple</a> [1]
<b>Additional services</b>	<a href="#">ABO Blood-Group System</a> [1]
<b>Hot topics</b> 	<a href="#">Abortion, Habitual</a> [1]
<b>Timelines</b> 	<a href="#">Abortion, Habitual / diagnosis</a> [1]
<b>MedHist events</b> 	<a href="#">Abortion, Habitual / immunology</a> [1]
<b>Psci-com events</b> 	<a href="#">Abortion, Induced</a> [6]
<b>Psci-com books and articles</b> 	<a href="#">Abortion, Induced / prevention and control</a> [1]
	<a href="#">Abortion, Legal</a> [4]
	<a href="#">Abortion, Spontaneous</a> [3]
	<a href="#">Abortion, Threatened</a> [1]
	<a href="#">Abruptio Placentae</a> [1]
	<a href="#">Abscess / radiology</a> [1]
	<a href="#">Academies and Institutes</a> [222]
	<a href="#">Access to Information</a> [2]
	<a href="#">Accident Prevention</a> [19]

Rys. 6.10. Wskaźnik kategorii leksykalnej *rodzaj zasobu* w systemie Intute: zdrowie i nauki o życiu – Medycyna. [Źródło: <http://intute.ac.uk/>]

Budowa narzędzi dostępu przedmiotowego reprezentujących ahierarchiczny wariant porządkowania punktów dostępu polega przede wszystkim na generowaniu tego rodzaju struktur. Udział JIW jest tutaj uzależniony od jego wykorzystania w procesie indeksowania. Aktywizacja jego funkcji wyszukiwawczej polega na pełnieniu przez jego wyrażenia funkcji punktów dostępu. Każde z nich jest bowiem węzłem dla powiązania hipertekstowego, którego uruchomienie powoduje skierowanie użytkownika do zbioru relewantnych ChWD. Warunkiem uzyskania przez jednostkę leksykalną JIW statusu punktu dostępu jest jej użycie w procesie indeksowania do odwzorowania informacji o obiekcie.

Specyfika ahierarchicznego wariantu narzędzia dostępu zakłada wykorzystanie struktur równoważnościowych punktów dostępu. Tym samym porządek wyrażen pełniących funkcje wyszukiwawcze ma charakter formalny i wyznacza go ich układ alfabetyczny. W związku z tym, wykorzystanie języka informacyjnego w dostępie do kolekcji polega na traktowaniu jego słownika przede wszystkim jako źródła słownictwa w procesie automatycznego wyboru nazw dla punktów dostępu.

## 6.2. *Wariant hierarchiczny*

Hierarchiczny wariant budowy narzędzi dostępu przedmiotowego w DSH polega na uporządkowaniu punktów dostępu do postaci struktury drzewiastej, którą tworzą kategorie tematyczne zagnieżdżone w środowisku hipertekstowym. Dostęp do zasobów systemu z wykorzystaniem tego typu narzędzi polega na eksploracji układu kategorii tematycznych, aż do momentu odnalezienia takiej, której zakres odpowiada zakresowi potrzeby informacyjnej użytkownika. Zagnieżdżenie układu w środowisku hipertekstowym umożliwia poruszanie się użytkownika w strukturze, wyznaczając tym same ścieżki dostępu do kategorii grupujących relewantne ChWD. Proces aktywacji powiązań hipertekstowych, w które wyposażone są nazwy kategorii (posiadające tym samym status węzła hipertekstowego) powoduje każdorazowe wyodrębnianie ze zbioru wyszukiwawczego zestawu ChWD o zakresach wyznaczanych przez zakres kategorii aktualnie eksplorowanej. W takim układzie każda z kategorii posiada status węzła hipertekstowego wyposażonego w odpowiednie odesłanie. Jego aktywacja powoduje:

- odpowiedź systemu w postaci wyświetlenia listy relewantnych ChWD zgrupowanych w danej kategorii,
- aktualizację relacji semantycznych wiążących kategorię eksplorowaną z pozostałymi w układzie, poprzez wyświetlenie ich nazw użytkownikowi.

W ten sposób użytkownik ma możliwość analizy rezultatów wyszukiwania i jednocześnie modyfikacji zapytania (budowanego w ten nieformalny sposób) poprzez eksplorowanie kategorii powiązanych z aktualnie przeglądaną. Poruszanie się w głąb struktury pomiędzy węzłami, których nazwy stanowią nazwy kategorii wykazuje podobieństwo do wyodrębniania łańcucha klasyfikacyjnego. Jego ostatnim ogniwem jest tutaj kategoria, której zakres jest relewantny dla użytkownika

Wariant hierarchiczny jest typem narzędzia dostępu przedmiotowego w DSH, w którym projektując jego strukturę wykorzystuje się organizację pola semantycznego języków informacyjno-wyszukiwawczych. Są to języki typu klasyfikacji, ze szczególnym uwzględnieniem znanych klasyfikacji uniwersalnych (np. UKD, KDD) oraz klasyfikacji specjalistycznych.

Charakterystyczną cechą tego rodzaju narzędzi jest ich świadoma konstrukcja, w przeciwieństwie do generowania, co jest charakterystyczne dla narzędzi w wariacie ahierarchicznym. Ten aspekt, stanowi o wartości dodanej narzędzia, gdyż pełni ono funkcję mapy konceptualnej dając użytkownikowi możliwość wglądu w dystrybucję tematów w obiektach informacyjnych z kolekcji. Taką rolę pełnią kategorie tematyczne,



które jednocześnie realizując funkcję wyszukiwawczą skierowują użytkownika do zbioru ChWD do nich przyporządkowanych. W przeciwieństwie do narzędzi wariantu ahierarchicznego, tutaj charakterystyka obiektu informacyjnego jest przyporządkowywana do narzędzia dostępu w postaci określonego zbioru kategorii tematycznych.

### 6.2.1. Metodyka ekstrakcji

Analiza budowy narzędzi dostępu w wariacie hierarchicznym była poprzedzona ich ekstrakcją z interfejsu użytkownika w grupie reprezentatywnej DSH (Zał. 2). Polegało to na przeniesieniu struktury kategorii tematycznych do postaci arkuszy kalkulacyjnych Microsoft Excel. Pozwoliło to na przeprowadzenie badań nad strukturalizacją pola tematycznego reprezentowanego przez układ kategorii tematycznych. Zasadniczą kwestią było poprawne odwzorowanie hierarchicznych zależności pomiędzy punktami dostępu. W przypadku narzędzi dostępu przedmiotowego, skonstruowanego z wykorzystaniem podziału klas w funkcjonujących klasyfikacjach, odwzorowywano rzeczywiste położenie punktu dostępu w strukturze narzędzia. Nie kierowano się miejscem symbolu klasyfikacyjnego (który był podstawą do utworzenia kategorii tematycznej) w drzewie klasyfikacyjnym. Tabela 6.1 przedstawia sposób identyfikacji rozmieszczenia kategorii dla narzędzia dostępu w przykładowym systemie BUBL, który stosuje symbole Klasyfikacji Dziesiątej Deweya i ich odpowiedniki słowne.

Tab. 6.1. Metodyka identyfikacji miejsca kategorii w schemacie. [Źródło: opracowanie własne]

Symbol i nazwa kategorii			Poziom
-	600 Technology (applied sciences)		1
-	-	<u>600 Technology</u>	2
-	-	<u>610 Medical sciences, medicine</u>	2
-	-	620 Engineering and allied operations	2
-	-	- 620 Engineering: general resources	3
-	-	- 620 Engineering: companies	3
-	-	- 620 Engineering: departments	3
-	-	- 620 Engineering: education and research	3
-	-	- 620 Engineering: journals and magazines	3
-	-	- 620 Engineering: societies	3
-	-	- <b>620.11 Materials science</b>	3
-	-	- <b>620.5 Nanotechnology</b>	3
-	-	- 621 Electronics and electrical engineering	3

-	-	-	621 Mechanical engineering	3
-	-	-	621 Optical engineering	3
-	-	-	621 Telecommunications and communications engineering	3
-	-	-	622 Mining and related operations	3
-	-	-	623 Military and nautical engineering	3
-	-	-	624 Civil engineering	3
-	-	-	625 Engineering of roads and railways	3
-	-	-	627 Hydraulic engineering	3
-	-	-	628 Sanitary and municipal engineering, environmental protection engineering	3
-	-	-	<b>629.1 Aerospace engineering</b>	3
-	-	-	<b>629.4 Astronautics</b>	3
-	-	-	<b>629.892 Robots</b>	3
-	-	-	<u>630 Agriculture and related technologies</u>	2
-	-	-	<u>640 Home economics and family living</u>	2
-	-	-	<u>650 Management and auxiliary services</u>	2
-	-	-	<u>660 Chemical engineering</u>	2
-	-	-	<u>670 Manufacturing</u>	2
-	-	-	<u>680 Manufacture for specific uses</u>	2
-	-	-	<u>690 Buildings</u>	2

W Tab. 6.1. pogrubioną czcionką zaznaczono ewidentne elementy braku zgodności schematu z budową drzewa klasyfikacyjnego KDD. Wyróżnione kategorie tematyczne reprezentują klasy w KDD na niższych stopniach podziału niż te, na których umieszczono je w narzędziu dostępu przedmiotowego w systemie BUBL.

Przedmiotem badań o charakterze ilościowym było określenie wartości dla zmiennych:

- głębokość podziału – liczba stopni podziału,
- łączna liczba kategorii w schemacie,
- liczba kategorii na pierwszym stopniu podziału,
- rozkład kategorii na poszczególnych poziomach w strukturze,
- pojemność kategorii na poszczególnych poziomach w strukturze (liczba odesłań do ChWD),
- średnia pojemność kategorii na poszczególnych poziomach w strukturze,
- dystrybucja ChWD na poszczególnych poziomach w strukturze,
- średnia liczba kategorii, do których przyporządkowano jedną ChWD.

Celem badań o charakterze ilościowym było opracowanie modelu typowego narzędzia dostępu przedmiotowego reprezentującego wariant hierarchiczny. Uogólnienia i obiektywny charakter zmiennych nie uwzględniają zróżnicowania w budowie tego

narzędzia, ale są konieczne do stworzenia obrazu jego struktury, który spełniać ma funkcje poznawcze.

Badania o charakterze jakościowym miały na celu analizę strukturalną uwzględniającą:

- porządek kategorii na pierwszym stopniu podziału,
- porządek kategorii w całym schemacie,
- przynależność kategoryjną działów na pierwszym stopniu podziału,
- zasady budowy struktur hierarchicznych,
- identyfikację typów wykorzystanych relacji semantycznych,
- sposoby wykorzystania języka informacyjno-wyszukiwawczego w strukturalizacji pola tematycznego.

Dalsza charakterystyka tego narzędzia będzie przeprowadzona według sformułowanych założeń badawczych.

### 6.2.2. Analiza ilościowa

Głębokość struktur hierarchicznych mierzono liczbą stopni podziału, jaką zastosowano do organizacji kategorii tematycznych. Analiza grupy reprezentatywnej pod tym kątem wykazała, że głębokość struktur waha się pomiędzy dwoma i siedmioma stopniami podziału. Szczegółowe dane przedstawia Tab. 6.2.

Tab. 6.2. Głębokość struktur hierarchicznych. [Źródło: opracowanie własne]

Głębokość podziału	2	3	4	5	6	7
Liczba systemów o określonej głębokości struktur hierarchicznych w narzędziach dostępu przedmiotowego	2	6	8	7	4	4
Udział procentowy	6,45%	19,35%	25,81%	22,58%	12,90%	12,90%

Dane przedstawione w Tab. 6.2 wskazują na częste zastosowanie struktur hierarchicznych do czwartego lub piątego stopnia podziału (25,81% oraz 22,58%). Są one wykorzystywane w prawie połowie zanalizowanych systemów, gdzie stosuje się tego rodzaju struktury drzewiaste. Około 19% łącznej liczby zbadanych systemów stosuje podział do trzeciego stopnia. Zdecydowanie głębokie struktury hierarchiczne (do szóstego lub siódmego stopnia podziału) stosuje się łącznie w jednej czwartej całkowitej liczby zbadanych systemów.

Głębokość struktur hierarchicznych wiąże się z wyodrębnianiem i nazywaniem pojęć o wąskich zakresach znaczeniowych. Im struktura wykazuje większy stopień podziału, tym ogniwa końcowe reprezentują pojęcia o węższym zakresie znaczeniowym. Tym samym zwiększa się długość łańcucha wolnego\*, którego ogniwem początkowym jest klasa główna, a ostatnim ogniwo końcowe. Powoduje to dłuższą ścieżkę eksploracji narzędzia przez użytkownika w celu odnalezienia kategorii tematycznej o wąskim zakresie.

Kolejny element analizowanych struktur hierarchicznych dotyczył łącznej liczby wyodrębnionych kategorii. Przyjmowała ona wartości od około 60 do ponad 2000. Narzędzia dostępu przedmiotowego w wariacie hierarchicznym, w grupie reprezentatywnej DSH były zbudowane średnio z 633 kategorii. (Tab. 6.3).

Tab. 6.3. Liczba kategorii w strukturze. [Źródło: opracowanie własne]

Nazwa	Liczba kategorii
Aerade	227
Agrifor	533
Altis	417
AnglistikGuide	101
Arkade	320
Artifact	953
BioethicsWeb	175
Biogate	66
BioResearch	375
BUBL	1362
Chemdex	582
EdNa	1702
EELS	834
EEVL Informatyka	259
EEVL Inżynieria	205
EEVL Matematyka	165
ForestryGuide	69
GEES	205
GEM	238
GeoGuide	278
Humbul	1356
Infomine Bio, Ag Med. Sciences	2148
Intute Arts&Humanities	2199
Intute Science, Engineering & Technology	1176
Intute Social Sciences	1405
MathForum	374
MathGuide	65
MedHist	436
NaturalSelection	357
Psi-com	85

\* Łańcuch drzewa klasyfikacyjnego [...] jest to ciąg klas rozpoczynający się od dowolnej klasy i składający się z tej klasy i wszystkich jej klas nadrzędnych [...] Łańcuch wolny jest to łańcuch, którego ostatnim ogniwem jest klasa końcowa [Ungurian 1976, s. 122]

Średnia liczba kategorii	633
--------------------------	-----

Liczbę kategorii dla danego systemu porównano z wielkością jego zbioru informacyjnego. Ustalono brak jednoznacznej zależności pomiędzy wielkością zbioru wyszukiwawczego systemu, a całkowitą liczbą kategorii tematycznych w narzędziu. Wraz ze wzrostem wielkości zbiorów informacyjnych analizowanych systemów następowała jednak rozbudowa narzędzia. Nie odnotowano jednak stałej zależności pomiędzy tymi dwiema zmiennymi.

Liczba kategorii i ich zakres na pierwszym stopniu w strukturze to wynik tego, co Barbara Kwasnik określa mianem „pierwszego cięcia”. Identyfikacja liczby i zakresu kategorii głównych, rozpoczynających dalszy podział determinuje formę i „elokwencję” reprezentacyjną całej struktury. [Kwasnik 1999] Pierwsze cięcie, czyli pierwsza zasada podziału jest zazwyczaj wynikiem zastosowania podziału dyscyplinarnego dla systemów o zakresie ogólnym (np. BUBL) lub identyfikacją przedmiotów badawczych w ramach określonej dyscypliny naukowej dla systemów dziedzinowych. W grupie reprezentatywnej liczba kategorii wyodrębnianych na pierwszym stopniu podziału wynosiła od 3 do 40. Średnia wartość dla tej zmiennej to 14 kategorii na pierwszym stopniu podziału (Tab. 6.4).

Tab. 6.4. Liczba kategorii na pierwszym stopniu podziału w narzędziach. [Źródło: opracowanie własne]

Nazwa	Liczba kategorii na 1 stopniu podziału
Aerade	13
Agrifor	10
Altis	3
AnglistikGuide	5
Arkade	6
Artifact	7
BioethicsWeb	10
Biogate	12
BioResearch	10
BUBL	10
Chemdex	13
EdNa	6
EELS	6
EEVL Informatyka	11
EEVL Inżynieria	16
EEVL Matematyka	10
ForestryGuide	14
GEES	10
GEM	12
GeoGuide	18
Humbul	33
Infomine Bio, Ag Med. Sciences	21
Intute Arts&Humanities	40
Intute Sciences	10

Intute Social Sciences	19
MathForum	35
MathGuide	7
MedHist	11
NaturalSelection	16
Psi-com	9
Średnia liczba kategorii	14

Nie stwierdzono zależności pomiędzy liczbą kategorii wyodrębnianych na pierwszym stopniu podziału, a głębokością całej struktury oraz całkowitą liczbą zidentyfikowanych kategorii w układzie.

Kolejną cechą o charakterze ilościowym w analizowanych strukturach była dystrybucja kategorii na poszczególnych stopniach podziału. Wskaźnik ten charakteryzuje stopień nagromadzenia kategorii na poszczególnych stopniach podziału (Tab.6.5).

Tab. 6.5. Dystrybucja kategorii w strukturze. [Źródło: opracowanie własne]

Stopień podziału	1	2	3	4	5	6	7
Średnia liczba wyodrębnionych kategorii	14	120	296	151	44	7	2
Procentowy rozkład kategorii	4,66%	33,10%	39,09%	15,65%	4,03%	0,72%	0,30%

Dane przedstawione w Tab. 6.5. wskazują znaczący udział ilościowy kategorii na poziomach drugim i trzecim w stosunku do pozostałych stopni podziału. Na tych poziomach zidentyfikowano kolejno około 33% i 39% łącznej liczby wyodrębnianych kategorii tematycznych. Tym samym na wskazanych dwóch poziomach zidentyfikowano ponad 70% łącznej liczby kategorii. Najmniej kategorii grupuje się na końcowych stopniach podziału i wskaźnik ten nie przekracza 1% ogólnej liczby kategorii tworzących narzędzie. Średnia liczba kategorii na poszczególnych stopniach podziału jest mocno zróżnicowana i waha się od 2 do 296 w zależności od poziomu głębokości. Największą wartość reprezentuje poziom trzeci i wynosi on średnio 296 kategorii. Jest to konsekwencja tendencji do dystrybucji największej liczby kategorii na tym poziomie w strukturze hierarchicznej.

W Tab. 6.6. przedstawiono dane dotyczące dystrybucji charakterystyk wyszukiwawczych obiektów informacyjnych przyporządkowanych do kategorii na poszczególnych stopniach podziału. Podczas badania analizowano również zmienną dotyczącą rozproszenia ChWD w strukturze hierarchicznej kategorii tematycznych. Wartości tej zmiennej dla poszczególnego stopnia podziału uśredniono i przedstawiono

również w postaci rozkładu procentowego w Tab. 6.6.

Tab. 6.6. Dystrybucja ChWD w strukturze. [Źródło: opracowanie własne]

Stopień podziału	1	2	3	4	5	6	7
Średnia liczba zgrupowanych ChWD	356	2706	5829	1351	534	105	4
Procentowy rozkład ChWD	3,27%	24,86%	53,56%	12,41%	4,90%	0,97%	0,03%

Analiza danych zawartych w Tab. 6.6 wskazuje, że w kategoriach tematycznych na poziomach drugim i trzecim zgrupowano około 25% i 54% ChWD, co daje razem ponad 79% łącznej liczby ChWD tworzących zbiór wyszukiwawczy systemu. Biorąc pod uwagę dystrybucję ChWD w kategoriach na drugim, trzecim i czwartym stopniu podziału, mamy do czynienia z narzędziem, które grupuje w kategoriach na tych poziomach 90% elementów zbioru wyszukiwawczego. W kategoriach na pierwszym stopniu podziału grupuje się zaledwie 3% łącznej liczby ChWD. Taka dystrybucja ChWD w strukturze hierarchicznej świadczy o:

- małej liczbie obiektów informacyjnych w zbiorze reprezentujących szeroki lub ogólny zakres tematyczny – około 3% łącznej liczby ChWD przyporządkowanych do kategorii na pierwszym stopniu podziału,
- małej liczbie obiektów informacyjnych w zbiorze informacyjnym reprezentujących wąski lub bardzo specjalistyczny zakres przedmiotowy – około 18% łącznej liczby ChWD przyporządkowanych do kategorii zgrupowanych od piątego do siódmego stopnia podziału,
- charakterze polityki gromadzenia i selekcji zasobów internetowych, gdzie uwzględnia się szczegółowość zagadnień poruszanych w treści obiektów informacyjnych włączanych do kolekcji.

Przedstawione wnioski z dystrybucji kategorii tematycznych w strukturze potwierdzają hipotezę o wykorzystaniu tzw. „zasady trzech kliknięć” (ang. 3-click rule) podczas tworzenia narzędzi dostępu w wariacie hierarchicznym. Wyrosła ona na gruncie badań nad użytecznością serwisów hipertekstowych. Według niej, użytkownicy korzystający z hipertekstowych serwisów internetowych wykonują maksymalnie trzy „kliknięcia”, czyli trzy kroki w eksploracji takiej struktury w celu dotarcia do relewantnych informacji. Jeśli taki proces nie skutkuje odnalezieniem pożądaných informacji następuje porzucenie przez użytkownika narzędzia lub ścieżki eksploracji. [Wise 2001]

Kolejną zmienną, której wartość określono podczas badania była pojemność

kategorii na określonym stopniu podziału, czyli liczba przyporządkowanych do niej ChWD. Uśrednione dane dotyczące tej zmiennej przedstawia Tab. 6.7.

Tab. 6.7. Pojemność kategorii na poszczególnych stopniach w strukturze. [Źródło: opracowanie własne]

Stopień podziału	1	2	3	4	5	6	7
Średnia liczba ChWD przyporządkowanych do kategorii	23,32	31,25	20,70	12,94	8,82	3,86	0,35

Średnia pojemność kategorii tematycznej na poziomach od pierwszego do trzeciego przyjmuje wartość w przedziale od 20 do 30 ChWD. Wraz ze zwiększaniem się głębokości schematu, pojemność kategorii maleje, aż do wartości 0,35 ChWD\* na siódmym stopniu podziału. Na pierwszym stopniu podziału pojemność jest stosunkowo wysoka i wynosi średnio około 23 ChWD. Wynika to ze stosunkowo małej liczby kategorii wyodrębnianej na tym stopniu i konieczności przyporządkowania do nich określonej liczby ChWD. Ostatnią zmienną, której wartość określono podczas badania była liczba kategorii, do których przyporządkowano jedną ChWD (Tab. 6.8). Wskaźnik ten charakteryzuje liczbę punktów dostępu do jednej ChWD rozproszonych w strukturze hierarchicznej. Średnią liczbę punktów dostępu do ChWD w tego rodzaju narzędziach obliczono na podstawie wzoru:

$$P_D = L_O / L_R$$

$P_D$  – liczba punktów dostępu do jednej ChWD

$L_O$  – liczba ChWD przyporządkowanych do wszystkich kategorii

$L_R$  – całkowita liczba rekordów w systemie

Tab. 6.8. Liczba punktów dostępu do jednej ChWD dla wybranych systemów.

[Źródło: opracowanie własne]

Nazwa systemu	$L_O$	$L_R$	$P_D$
Altis	4802	3852	1,25
AnglitikGuide	3737	2024	1,85
BioethicsWeb	3512	654	5,37
Biogate	2984	1000	2,98
EEVL Computing	4848	5617	0,86
EEVL Engineering	18407	18895	0,97**
EEVL Mathematics	3442	2620	1,31
Forestry Guide	1113	992	1,12

\* Spadek wartości poniżej 1 ChWD na kategorię w przypadku siódmego stopnia podziału jest spowodowany uśrednieniem wartości dla całej grupy reprezentatywnej, w której nie wszystkie systemy stosowały tak głębokie struktury hierarchiczne. Dodatkowym powodem takiej wartości jest występowanie w schematach pustych kategorii, gdzie nie przyporządkowano żadnej ChWD. Szerzej zostanie to omówione w następnej sekcji tego rozdziału.

\*\* Obecność pustych kategorii.



GeoGuide	5451	3144	1,73
Humbul	39013	18000	2,17
Intute Science	65269	33309	1,96
MathGuide	1084	1126	0,96**
MedHist	5921	1105	5,36
Natural Selection	10137	5547	1,83
Psi-Com	1192	951	1,25
Wartość średnia			2,07

Wskaźnik ten dla systemów tworzących próbę reprezentatywną wynosi średnio 2,07. Oznacza to, że dostęp do jednej ChWD zapewniają średnio dwa punkty dostępu w postaci kategorii tematycznych.

Przedstawione dane ilościowe dotyczące budowy narzędzi dostępu przedmiotowego w wariancie hierarchicznym w DSH poprzez generalizację wyników badań pozwalają na konkluzję, że jest to struktura hierarchiczna:

- zbudowana średnio z 633 kategorii,
- wykorzystująca maksymalnie siedmiostopniowy system podziału,
- posiadająca na pierwszym stopniu podziału średnio 14 kategorii,
- gdzie 88% łącznej liczby kategorii zostało wyodrębnionych na poziomach od drugiego do czwartego,
- gdzie 90% wszystkich ChWD zostało przyporządkowanych do kategorii na poziomach od drugiego do czwartego,
- gdzie średnia pojemność kategorii wynosi około 14 ChWD i uwarunkowana jest od jej miejsca w strukturze,
- gdzie jedna ChWD została przyporządkowana średnio do dwóch kategorii w schemacie.

### 6.2.3. Analiza jakościowa

Celem analizy o charakterze jakościowym było wskazanie charakterystycznych cech w budowie analizowanych struktur, reprezentujących organizację pola tematycznego dziedzinowego systemu hipertekstowego. Analiza budowy tego rodzaju struktur organizujących dostęp do zasobów pozwoliła zidentyfikować różne rodzaje organizacji pól tematycznych systemów, których projekcją są układy kategorii tematycznych w narzędziach dostępu przedmiotowego. Szczegółowej analizie poddano rodzaj relacji porządkującej kategorie tematyczne oraz zasady ich tworzenia.

Ostatecznie, celem badań było określenie funkcji JIW wykorzystywanego do tworzenia wariantu hierarchicznego narzędzi dostępu przedmiotowego.

### 6.2.3.1. Porządek kategorii w układzie

Porządek kategorii wyodrębnionych na pierwszym stopniu podziału w strukturach hierarchicznych ma formę logiczną lub alfabetyczną. Porządek alfabetyczny na pierwszym stopniu podziału zidentyfikowano w 46% łącznej liczby zbadanych systemów wykorzystujących tego typu narzędzia dostępu. W nieznaczącej większości – 54%, stosowano porządek wynikający z przyjętego układu logicznego. Przyjęty sposób porządkowania kategorii na pierwszym stopniu (Rys. 6.11 i 6.12) jest stosowany konsekwentnie w dalszej rozbudowie struktury.



The screenshot shows the AERADE website header with the logo and tagline "Your quality portal to aerospace & defence resources on the Internet". Below the header is a navigation menu with links: Home, About, AERADE@yoursite, Contact us, Help. The main content area is divided into two columns. The left column contains three sections: "AERADE services" with links to Aerospace & defence resources, DEVISE, ESDU Series, Reports Archive, and Shrivenham Index; "Inform:" with links to ConferenceBrief, JobsBrief UK, and NewsBrief; "AERADE newsletter" with a link to More...; and "AERADE friends" with a link to Books Express. The right column is titled "Browse DEVISE by subject heading" and contains a "Browse headings" section with a list of subject headings: Armed Forces, Military Vehicle Technology, Surface Ship and Submarine Technology, Weapon Technology, Missile Technology, Ballistics, Materials Technology - Military Applications, Military Operations, Emerging Technology, Command, Control, Communications, Computers and Intelligence (C4I), Electronic and Information Warfare, Training and Simulation, and Information Centre.

Rys. 6.11. Logiczny porządek kategorii na pierwszym stopniu podziału w systemie Aerade. [Źródło: [http://aerade.cranfield.ac.uk/devise\\_index.html](http://aerade.cranfield.ac.uk/devise_index.html)]

Search  in Psci-com gateway Go

[Advanced search](#) [Latest resources](#) [Help](#)

**Browse**

Browse by [LC classification schedules](#).  
The Library of Congress (LC) Classification Schedules give you a very broad overview of the range of subject areas where we have identified and described resources

**Psci-com subject headings**

<b><a href="#">Associations &amp; Organisations</a></b> <a href="#">PES-focused orgs.</a> , <a href="#">Funding sources</a> , ...	<b><a href="#">Event listings</a></b> <a href="#">For professionals</a> , <a href="#">For the public</a>
<b><a href="#">Exemplar Science Web Sites</a></b> <a href="#">Health</a> , <a href="#">Science</a> , <a href="#">Technology</a> , ...	<b><a href="#">Nature of Science</a></b> <a href="#">History</a> , <a href="#">Philosophy</a> , ...
<b><a href="#">Publications</a></b> <a href="#">Reports</a> , <a href="#">Surveys</a> , <a href="#">Journals</a> , ...	<b><a href="#">Reference</a></b> <a href="#">Libraries</a> , <a href="#">Web gateways</a> , ...
<b><a href="#">Science Centres and Museums</a></b> <a href="#">Museum policy</a> , <a href="#">UK science centres</a> , ...	<b><a href="#">Science Communication</a></b> <a href="#">Public engagement</a> , <a href="#">Science &amp; media</a> , ...
<b><a href="#">Science Education</a></b> <a href="#">Associations</a> , <a href="#">Projects/Initiatives</a> ,	<b><a href="#">Science in Society</a></b> <a href="#">Ethics</a> , <a href="#">Science &amp; law</a> , <a href="#">Science &amp; art</a> , ...

Rys. 6.12. Alfabetyczny porządek kategorii na pierwszym stopniu podziału w systemie Psi-com. [Źródło: <http://www.intute.ac.uk/healthandlifesciences/psicom/>]

Kategorie wyodrębniane na pierwszym stopniu, w przypadku systemów o zakresie ogólnym (np. BUBL, Claygate), są wynikiem zastosowania podziału dyscyplinarnego wykorzystującego zazwyczaj porządek klas głównych znanych klasyfikacji nauk. W systemach, których zakres kolekcji jest ograniczony do jednej dziedziny wiedzy lub sfery działalności ludzkiej, na pierwszym stopniu podziału mamy do czynienia z wyodrębnianiem kategorii, których zakres odpowiada zagadnieniom szczegółowym wchodzącym w zakres danej dyscypliny. Suma zakresów kategorii tematycznych wyodrębnionych na pierwszym stopniu podziału wyznacza zakres całej struktury hierarchicznej i tym samym całego zbioru informacyjnego systemu (Rys. 6.13).

Open all categories | Close all categories

- ▼ **Forestry, Forest Products and Agroforestry (General) (202 records)**
- ▼ **Forests and Forest Trees (Biology and Ecology) (464 records)**
  - Silviculture and Forest Management (338 records)
  - Forest Mensuration and Management (Discontinued March 2000) (78 records)
  - Forest Fires (32 records)
  - Protection Forestry (Discontinued March 2000) (48 records)
  - Other Land Use (Discontinued March 2000) (6 records)
  - Ornamental and Amenity Trees (32 records)
- ▼ **Forest Products and Industries (General) (92 records)**
  - Wood Properties, Damage and Preservation (29 records)
  - Logging and Wood Processing (29 records)
  - Wood Utilization and Engineered Wood Products (80 records)
  - Chemical and Biological Processing of Wood (56 records)
  - Non-Wood Forestry Products (20 records)
  - Agroforestry and Multipurpose Trees, Community, Farm and Social Forestry (104 records)

Open all categories | Close all categories

Rys. 6.13. Hierarchiczne narzędzie dostępu w systemie ForestryGuide. [Źródło: <http://forestryguide.de>]

W przykładowym narzędziu dostępu przedmiotowego zastosowanym w systemie ForestryGuide, na pierwszym stopniu podziału zidentyfikowano trzy kategorie tematyczne. Są to:

- leśnictwo, lasy i agrobiologia (ogólnie),
- lasy i drzewa (biologia i ekologia),
- przemysł i produkcja związana z lasami (ogólnie).

Suma zakresów tych kategorii, poddanych dalszej rozbudowie wyznacza zakres organizowanej kolekcji. W narzędziu tym w precyzyjny sposób określono zakres kolekcji poprzez bezpośrednią identyfikację zakresów kategorii pierwszego stopnia. Brak jest tutaj niejednoznacznej kategorii tematycznej typu *Inne źródła*, która jest często wykorzystywana w katalogach portali internetowych. Jej użycie jest dyskusyjne ponieważ zasada podziału, na podstawie której ją wyodrębniono ma małą wartość organizującą i wyszukiwawczą. Dodatkowo jest wynikiem nieprecyzyjnej analizy kolekcji pod kątem tematów występujących w jej zasobach.

### 6.2.3.2. Przynależność kategoriałna punktów dostępu

Analiza zestawu punktów dostępu zidentyfikowanych na pierwszym stopniu podziału w narzędziach w wariacie hierarchicznym w zbiorze reprezentatywnym, pozwoliła na identyfikację sześciu kategorii reprezentujących określone wymiary pola

tematycznego. Są to:

1. Przedmiot – wskazanie na przedmiot dokumentu, nazwy zagadnień ogólnych oraz szczegółowych, np. *rolnictwo, mikrobiologia, przetwarzanie informacji*;
2. Użytkownik – wskazanie na grupę użytkowników, dla których przeznaczone są obiekty informacyjne, np. *szkolnictwo wyższe, dorośli, dzieci*;
3. Miejsce – identyfikacja fragmentu przestrzeni geograficznej, w której umiejscowiony jest przedmiot obiektu informacyjnego, np. *kultura Chin*.
4. Forma – wskazanie na formę obiektu informacyjnego, np. *informatory, portale, czasopisma*;
5. Instytucja – typy instytucji, będących przedmiotem treści obiektu informacyjnego, np. *organizacje rządowe, muzea, biblioteki*;
6. Niezidentyfikowane – kategoria o niesprecyzowanym zakresie reprezentowana najczęściej przez kategorie tematyczne o nazwach, np. *zasoby różne, inne źródła związane z medycyną, inne*;

Przynależność kategoryalną punktów dostępu na pierwszym stopniu podziału do opracowanego zestawu przedstawia Tab. 6.9.

Tab. 6.9. Przynależność kategoryalna punktów dostępu na pierwszym stopniu podziału.

[Źródło: opracowanie własne]

Nazwa kategorii	Przedmiot	Użytkownik	Miejsce	Forma	Instytucja	Niezidentyfikowane
Udział procentowy	78%	5%	7%	7%	2%	1%

Dane przedstawione w Tab. 6.9 wskazują na dominację kategorii reprezentujących treść obiektów informacyjnych. Badanie miało charakter uogólniający, tj. stworzono zestaw kategorii o stosunkowo dużym stopniu ogólności tak, aby uchwycić tendencję, którą można rozciągnąć na cały zbiór reprezentatywny. Analiza przynależności kategoryalnej poszczególnych systemów umożliwiłaby identyfikację kategorii specjalistycznych, czy też branżowych.

W rozdziale drugim pracy przedstawiono wyniki badań dotyczących m.in. analizy struktur taksonomicznych pod kątem identyfikacji kategorii leksykalnych wyodrębnianych na pierwszym stopniu podziału w portalach internetowych. W tym miejscu w Tab. 6.10. dokonano porównania typów kategorii odpowiednio w grupie reprezentatywnej dziedzinowych systemów hipertekstowych, grupie portali internetowych zbadanych przez Ch. Zisa oraz grupie portali internetowych zbadanych przez zespół pod kierunkiem K. Kyung-Sun.

Tab. 6.10. Porównanie rodzajów kategorii na pierwszym stopniu podziału. [Źródło: badania własne; Zins 2002; Kyung-Sun 2006]

DSH	Ch. Zins	K. Kyung-Sun
<ul style="list-style-type: none"> <li>▪ Przedmiot</li> <li>▪ Użytkownik</li> <li>▪ Miejsce</li> <li>▪ Forma</li> <li>▪ Instytucja</li> <li>▪ Niezidentyfikowane</li> </ul>	<ul style="list-style-type: none"> <li>▪ Przedmiot</li> <li>▪ Obiekt</li> <li>▪ Aplikacja</li> <li>▪ Użytkownicy</li> <li>▪ Lokalizacja</li> <li>▪ Informator</li> <li>▪ Medium</li> <li>▪ Język</li> </ul>	<ul style="list-style-type: none"> <li>▪ Przedmiot</li> <li>▪ Aplikacja</li> <li>▪ Lokalizacja</li> <li>▪ Medium/Typ/Format/Narzędzie</li> <li>▪ Organizacja</li> <li>▪ Ludzie</li> <li>▪ Informacje ogólne</li> <li>▪ Język</li> </ul>

Analiza danych przedstawionych w Tab. 6.10 wskazuje na mniejsze zróżnicowanie ilościowe kategorii tematycznych wyodrębnianych w DSH na pierwszym stopniu podziału. Tego rodzaju mniejsze rozproszenie kategoryjne, w którym dominującą kategorią jest *przedmiot* pozwala na ukierunkowanie całej struktury w stronę reprezentacji cech treściowych elementów kolekcji. Wpływ pierwszej zasady podziału, tzw. „pierwszego cięcia” [Kwasnik 1999] na konieczność reprezentacji treści wyznacza dalszy sposób systematyzacji słownictwa przyjętego za nazwy kategorii tematycznych. Pozostałe kategorie w DSH mają niewielki (12%) udział w wyborze rodzajów punktów dostępu na pierwszym stopniu podziału.

### 6.2.3.3. Budowa struktur hierarchicznych

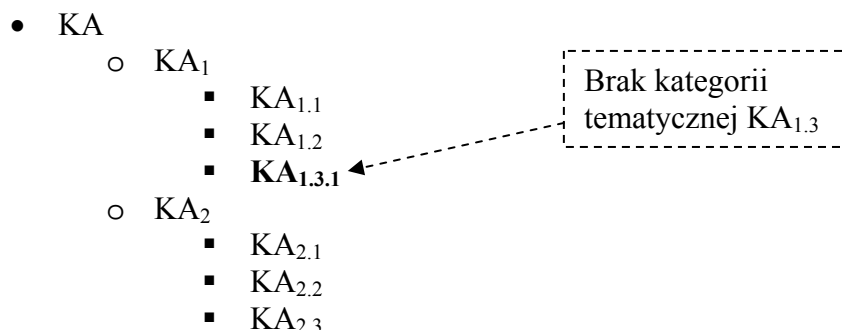
Porządkowanie rzeczywistości dokumentacyjnej będącej odbiciem świata informacji zawartego w obiektach informacyjnych tworzących kolekcję DSH odbywa się poprzez systematyzację słownictwa przyjętego za nazwy kategorii tematycznych w narzędziach dostępu przedmiotowego. W strukturach tego typu narzędzi mamy do czynienia z monohierarchicznym modelem organizacji pola tematycznego, ze zidentyfikowanym jednym typem relacji hierarchiczno-organizującej. Jest to relacja hierarchii tematycznej, którą definiuje się jako „zależność hierarchiczna między nazwą tematu szerszego i nazwą tematu węższego. Pozwala ona opisać hierarchiczną strukturę systemu leksykalnego wielu JIW. Zachodzić może między różnego rodzaju nazwami, które w określonych okolicznościach przyjmują funkcję tematu uznanego za szerszy lub nazwy tematu uznanego za węższy w strukturze pewnego zbioru informacyjnego (zbioru dokumentów lub opisów dokumentów). Najczęściej realizują go pary takich

nazw jak: nazwa dyscypliny naukowej > nazwa przedmiotu badawczego, nazwa obiektu > nazwa własności, nazwa obiektu > nazwa działania, nazwa procesu > nazwa poddanego mu obiektu, etc. [...] Relacja hierarchii tematycznej jest związkiem wyznaczanym arbitralnie dla potrzeb strukturyzacji określonego zbioru informacyjnego. W konsekwencji, tylko w określonym zbiorze informacyjnym spełnia ona logiczne warunki relacji hierarchicznej” [Sosińska-Kalata 2002a, s. 30]

Zestaw kategorii tematycznych, które grupują ChWD jest wyodrębniany według zasady *literary warrant*. Istota tej reguły opracowanej na początku ubiegłego wieku przez E. Wyndama Hulme’a [Hulme 1911] polega na analizie kolekcji pod kątem występujących w jej elementach tematów i ustalania ich nazw w procesie indeksowania. Norma ANSI/NISO Z39.19-2005 z zakresu tworzenia zbioru słownictwa kontrolowanego na potrzeby indeksowania pt. „Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies” definiuje tę zasadę jako „uzależnienie wyboru danego terminu dla reprezentacji danego pojęcia od jego częstego występowania w literaturze”. [ANSI/NISO Z39.19-2005, s. 162] Tym samym analiza pola tematycznego systemu pozwala na zastosowanie tej zasady w wyborze kategorii tematycznych w hierarchicznych narzędziach dostępu przedmiotowego. Jak odnotowano w rozdziale czwartym, w DSH występuje zjawisko polireprezentacji wiedzy. Jest ono przydatne w budowie hierarchicznych narzędzi dostępu przedmiotowego, które są projektowane w przeciwieństwie do automatycznie generowanego wariantu ahierarchicznego.

Polireprezentacja wiedzy obok funkcji heurystycznych w DSH, dzięki kumulacji zasobu leksykalnego wykorzystanego w reprezentacji wiedzy, pozwala na ustalenie ram ograniczających pole tematyczne systemu oraz na identyfikację pojęć reprezentowanych przez określone jednostki leksykalne. W praktyce oznacza to, że w hierarchicznym narzędziu dostępu przedmiotowego nie powinna mieć miejsce sytuacja, w której do danej kategorii tematycznej nie przyporządkowano pośrednio żadnego obiektu informacyjnego, a bezpośrednio żadnej jego charakterystyki wyszukiwawczej. Szytwe przestrzeganie zasady *literary warrant* w wyodrębnianiu kategorii tematycznych powoduje jednak pewne komplikacje. Może bowiem doprowadzić do naruszenia spójności całej struktury. Wybór relacji hierarchii tematycznej jako relacji systematyzującej kategorii tematyczne polega na wyodrębnianiu w zakresie określonej szerszej kategorii tematycznej zbioru kategorii tematycznych o węższym zakresie. Brak przestrzegania zasad podziału logicznego (rozłącznego i adekwatnego) w tego typu narzędziach powoduje sytuację, w której w określonej kategorii tematycznej na tym

samym poziomie zostaną wyodrębnione kategorie tematyczne o różnym stopniu szczegółowości. W zbiorze informacyjnym systemu może bowiem nie być dokumentów, których zakres tematyczny jest równy zakresowi pośredniczącej kategorii tematycznej (Rys. 6.14).



Rys. 6.14. *Literary warrant* jako źródło niespójności struktur hierarchicznych. [Źródło: opracowanie własne]

W hierarchicznym wariacie narzędzi dostępu przedmiotowego w DSH stosuje się odstępstwa od tej reguły w tych miejscach w strukturze narzędzia, gdzie brakuje kategorii tematycznych stanowiących ogniwa pośrednie. Stosuje się tam puste kategorie tematyczne nieposiadające egzemplarzy. Ich funkcją jest zachowanie spójności narzędzia.\* Taka sytuacja miała miejsce w omawianych w poprzednim rozdziale systemach MathGuide oraz Intute.

W odniesieniu do kilku DSH można mówić o wyodrębnianiu kategorii tematycznych w oparciu zasadę zbliżoną do *user warrant*. Zasada ta polega ona na wyborze terminów indeksowych z zapytań użytkowników. Z nich derywuje się określone wyrażenia, których zakres jest istotny dla efektywności procesów wyszukiwawczych. W przypadku struktur hierarchicznych narzędzi dostępu przedmiotowego, polega to na uwzględnieniu potrzeb informacyjnych grupy użytkowników docelowych w kontekście poszukiwania przez nich informacji o specyficznym zakresie. W większości grup tematycznych systemu Intute, skierowanego do użytkowników wywodzących się z sektora nauki i szkolnictwa Wielkiej Brytanii, istnieje wyraźna tendencja do wyodrębniania kategorii tematycznej (w różnych miejscach w strukturze), której zakres wiąże się z odniesieniem przedmiotu do lokalizacji geograficznej, w tym przypadku do Wielkiej Brytanii (Rys. 6.15 i 6.16).

\* W momencie kiedy zostanie przyporządkowany do nich jakkolwiek ChWD zostaje tym samym aktywowany ich udział w systematyzacji kategorii tematycznych.



- Architects and architectural practices
  - Asian architects and architectural practices
  - European architects and architectural practices
    - **British architects and architectural practices**
  - North and Central American architects and architectural practices
  - Oceanic architects and architectural practices

Rys. 6.15. *User warrant* w dostępie przedmiotowym systemu *Artifact*. [Źródło: <http://intute.ac.uk/>]

- Food and culture
  - African food and culture
  - American food and culture
  - Asian food and culture
  - Australian and Oceanic food and culture
  - European food and culture
    - **British food and culture**

Rys. 6.16. *User warrant* w dostępie przedmiotowym systemu *Intute: nauki społeczne*. [Źródło: <http://intute.ac.uk/>]

Na Rys. 6.15 i 6.16 przedstawiono rozbudowę przykładowych kategorii tematycznych w hierarchicznych narzędziach dostępu przedmiotowego, gdzie obok zasady *literary warrant* uwzględnia się również preferencje użytkowników. Przejawia się to wyodrębnianiem w zakresie danej kategorii tematycznej takich podkategorii, które lokalizują temat/przedmiot w przestrzeni geograficznej Wielkiej Brytanii. Takie podejście reprezentują następujące kategorie tematyczne:

- *British architects and architectural practices (Brytyjscy architekci i praktyka architektoniczna)* jako jedyna kategoria o zakresie węższym wyodrębniona z kategorii nadrzędnej *European architects and architectural practices (Europejscy architekci i praktyka architektoniczna)*,
- *British food and culture (Kultura i żywność brytyjska)* jako jedyna kategoria o zakresie węższym wyodrębniona z kategorii nadrzędnej *European food and culture (Kultura i żywność europejska)*

Podstawową relacją hierarchiczno-organizującą kategorie tematyczne jest relacja hierarchii tematycznej. Nie mniej, tego rodzaju struktury charakteryzuje również zastosowanie sieci relacji asocjacyjnych w odwzorowaniu powiązań znaczeniowych pomiędzy wybranymi wyrażeniami przyjętymi za nazwy kategorii grupujących ChWD. Wzbogacenie struktury o tego typu zależności semantyczne jest rzadko praktykowane. Najbardziej rozwinięty system relacji skojarzeniowych można spotkać w dużych DSH typu BUBL oraz Intute: nauki społeczne. Mamy tutaj do czynienia z zastosowaniem niespecyfikowanych relacji skojarzeniowych identyfikowanych pomiędzy dwoma

typami punktów dostępu. W Intute: nauki społeczne sieć relacji asocjacyjnych łączy wyłącznie zastosowane w schemacie kategorie tematyczne (Rys. 5.10, s. 236). Mają one formę odsyłaczy hipertekstowych, których aktywacja powoduje wyświetlenie ChWD przyporządkowanych do skojarzonych kategorii tematycznych i ich ewentualną rozbudowę. W BUBL (Rys. 5.3, s. 205) mamy do czynienia z zastosowaniem do wybranych kategorii tematycznych zestawu odsyłaczy uzupełniających, którego źródło stanowi leksyka LCSH. Określonym kategoriom tematycznym przyporządkowuje się skojarzone terminy LCSH, które użyto podczas indeksowania zbioru informacyjnego. Te również mają postać odsyłaczy hipertekstowych, których aktywacja powoduje wygenerowanie listy relewantnych ChWD, w których je użyto w procesie reprezentacji treści.

W rozbudowanych strukturach hierarchicznych w narzędziach dostępu przedmiotowego wykorzystuje się interesujący zabieg tzw. *quasifasetyzacji*. Umożliwia on użytkownikowi zreorganizowanie eksplorowanej kategorii tematycznej poprzez wyodrębnienie w jej zakresie innego rodzaju kategorii. Jest to zazwyczaj kategoria *typ zasobów* i/lub *czas*. Ma to miejsce w systemie Intute: sztuka i humanistyka. Ujawnienie kategorii *czas* oraz *typ zasobów* jest możliwe poprzez aktywację odesłań hipertekstowych oznaczonych:

- Browse by time period – przeglądamy według jednostki czasu,
- Browse by resource type – przeglądamy według typu zasobów.

*Quasifasetyzacja* polega tutaj na opracowaniu struktury kategorii tematycznych według zasady *literary warrant* z jednoczesnym zastosowaniem relacji hierarchii tematycznej do ich systematyzacji. Dodatkowo, w każdej kategorii z takiej struktury użytkownik ma możliwość wyodrębnienia kategorii z faset *czas* i/lub *typ zasobów* poprzez aktywację odpowiedniego hiperłącza (Rys. 6.17). Powoduje to wyświetlenie użytkownikowi podziału danej kategorii tematycznej według kryterium czasu i/lub formy obiektów informacyjnych.

[Intute](#) > [Arts and Humanities](#)

<b>Subject links</b>	<a href="#">Architecture</a> > <b>Architectural Education and Training</b>
<a href="#">About us</a>	<a href="#">Browse by time period</a>
<a href="#">A-Z of services</a>	<a href="#">Browse by resource type</a>
<b>Internet catalogue</b>	<b>Sub-categories</b>
<a href="#">Search</a>	<ul style="list-style-type: none"> <li><a href="#">Geography or Culture</a> [44]</li> <li><a href="#">Research and Development</a> [16]</li> </ul>
<a href="#">Browse</a>	Search <input type="text"/> in <a href="#">Architectural Education and Training</a> <input type="button" value="Go"/>
<a href="#">New resources</a>	<a href="#">Advanced search</a> <a href="#">Subject A-Z</a> <a href="#">New resources</a> <a href="#">Help</a>
<a href="#">Suggest a site</a>	Filter by <a href="#">All resource types</a> <input type="button" value="Go"/>
<b>Internet training</b>	<input type="button" value="Details"/> <input type="checkbox"/> Full record <input checked="" type="checkbox"/> Save Record
<a href="#">Virtual Training Suite</a>	Now showing: 1 - 11 of 11 records (Filter: All resource types)
<a href="#">Support materials</a>	Page: 1 <span style="float: right;">Order by <a href="#">Date</a>   <a href="#">Title</a></span>
<a href="#">Intute events</a>	<a href="#">Architectstudent.net</a> <input type="button" value="Details"/> <input checked="" type="checkbox"/>
<b>Features</b>	Architectstudent.net is a resource and showcase for architecture students from anywhere in the world. Upon registration (which is free), students are able to create their own portfolios, including a profile, 'manifesto' and uploaded images and texts. The website includes many student designs, including plans and 3D renderings. In addition to its student portfolios, Architecture.net includes a discussion area and links to other architecture resources on the Web. Each page can be reformatted as a PDF document for printing.
<a href="#">AHRC Projects</a>	
<a href="#">Blog</a>	
<a href="#">Conferences &amp; Events</a>	
<a href="#">Electronic journals</a>	
<a href="#">Limelight</a>	
<a href="#">MyIntute</a>	

Rys. 6.17. Quasifasetyzacja w Intute: sztuka i humanistyka.

[Źródło: <http://www.intute.ac.uk/artsandhumanities/>]

Tego typu „reorganizacja wiedzy na żądanie” charakteryzuje się dużą elastycznością i funkcjonalnością, gdyż nie obciąża całej struktury hierarchicznej. Może bowiem być aktywowana przez użytkownika w dowolnym miejscu podczas eksploracji schematu. Zabieg ten w DSH wykonuje się bardzo rzadko, ale z pewnością otwiera on nowe możliwości dla fasetowej organizacji punktów dostępu np. w strukturach taksonomicznych w środowisku sieciowym.

Wizualizacja struktury, w której zorganizowane są kategorie tematyczne jest ważnym zagadnieniem z punktu widzenia realizacji procesów pozyskiwania informacji przez użytkowników. Zagnieżdżenie jej w środowisku hipertekstowym wyznacza sposoby jej eksploracji, która opiera się na koncepcji węzłów i powiązań hipertekstowych. Każda kategoria tematyczna stanowi węzeł dla powiązania hipertekstowego, którego aktywacja powoduje wyświetlenie jej zawartości w postaci listy relewantnych ChWD oraz jej rozbudowy, czyli wyodrębnionych w jej zakresie bezpośrednio węższych zakresowo kategorii tematycznych. Do tych zalicza się kategorie tematyczne zidentyfikowane poprzez zastosowanie relacji hierarchii tematycznej, a także dodatkowo niespecyfikowanej relacji asocjacyjnej. Każda z nich posiada również status węzła hipertekstowego, którego uruchomienie spowoduje

analogiczne zachowanie systemu lub zwykłe wygenerowanie zbioru odpowiedzi w postaci listy ChWD. Przy czym kategorie tematyczne wyodrębnione w zakresie kategorii dzielonej lub skojarzone z nią są poprzedzone odpowiednim wskaźnikiem (np. ang. subcategories – podkategorie, ang. related sections – sekcje skojarzone) Eksploracja takiego układu jest zazwyczaj „monitorowana” w postaci tzw. wskaźnika miejsca.\* Ma on formę zbliżoną do ekstrahowanego łańcucha klasyfikacyjnego, którego pierwszym ogniwem jest kategoria tematyczna pierwszego stopnia podziału, a ostatnim aktualnie eksplorowana (Rys. 6.17).

The screenshot shows the header of 'The Math Forum @ Drexel' with a 'DONATE' button. Below the header is the text 'MATH FORUM INTERNET MATHEMATICS LIBRARY'. The main heading is 'Browse and Search the Library'. A breadcrumb trail is displayed: 'Home : Math Topics : Algebra : Modern Algebra : Rings/Ideals'. Below this is a navigation bar with links: 'Library Home || Search || Full Table of Contents || Suggest a Link || Library Help'. A red box highlights the 'Full Table of Contents' link. Below the navigation bar is a 'Selected Sites' section with the text '(see also All Sites in this category)'. The first site listed is 'Commutative Rings and Algebras - Dave Rusin; The Mathematical Atlas', followed by a short article description and a 'more>>' link.

Rys. 6.18. Wskaźnik miejsca w systemie *Math Forum*. [Źródło: <http://mathforum.org/library/topics/>]

Podstawową funkcją wskaźnika miejsca jest określenie miejsca w strukturze narzędzia dostępu przedmiotowego, jakie aktualnie eksploruje użytkownik. Każdy z elementów takiego łańcucha jest również wyposażony w odesłanie hipertekstowe, które pozwala w szybki sposób wybrać jedną z kategorii nadrzędnych w celach wyszukiwawczych.

\* Wskaźnik miejsca (ang. place indicator) jest narzędziem stosowanym w architekturze informacji i służy do wskazywania użytkownikowi miejsca w strukturze serwisu, które aktualnie eksploruje.

#### 6.2.4. Funkcje języka informacyjno-wyszukiwawczego

Rola JIW w organizacji kategorii tematycznych w narzędziach dostępu przedmiotowego w wariacie hierarchicznym jest uzależniona od wykorzystania organizacji jego pola semantycznego do ich porządkowania. Podstawową funkcją realizowaną na etapie dostępu do kolekcji jest jednak wykorzystanie zasobu leksykalnego JIW do nadawania nazw kategoriom tematycznym. Drugą, nie mniej ważną, jest wykorzystanie systemu relacji paradygmatycznych opracowanych dla konkretnych JIW do tworzenia układu kategorii tematycznych w strukturach hierarchicznych.

W DSH mamy do czynienia z narzędziami dostępu przedmiotowego w omawianym wariacie, które:

- są konstruowane na potrzeby danego systemu przy udziale leksyki JIW jako źródła słownictwa dla nazw kategorii tematycznych lub
- odwzorowują w określonym stopniu organizację pola semantycznego konkretnego JIW, który cechuje daleko posunięty redukcjonizm.

W pierwszym przypadku udział JIW jest zredukowany do pełnienia funkcji zasobu leksykalnego dla nazw wyodrębnianych kategorii tematycznych, identyfikowanych na podstawie zasady *literary warrant*. Tego rodzaju **indukcyjne** podejście do systematyzacji kategorii tematycznych powoduje, że cała struktura narzędzia dostępu do kolekcji jest dopasowana do stopnia rozproszenia tematów w polu tematycznym systemu. Niestety takie podejście ma również swoje wady. Jest ono obciążone małą spójnością całego układu, gdyż operuje wyłącznie kategoriami tematycznymi wyodrębnionymi na podstawie zasady *literary warrant*. Chociaż tego rodzaju struktury nie spełniają wymogów stawianych podziałowi logicznemu, to tutaj także pojawia się problem racjonalnego uzasadnienia obecności nowej kategorii tematycznej w ramach innej o szerszym zakresie. W związku z tym, często spotyka się podział kategorii tematycznych niespełniający wymogu rozłączności i adekwatności podziału. Relacją organizującą jest tutaj relacja hierarchii tematycznej, która wyznacza dosyć swobodne ramy interpretacyjne.

Funkcję zaplecza leksykalnego w tym wariacie pełnią JIW o notacji paranaturalnej, o niekoniecznie rozbudowanej paradygmatyce. Samo słownictwo JIW pełni wyłącznie funkcję pomocniczą i nie stosuje się go konsekwentnie, włączając w to wyrażenia języka naturalnego.

Wykorzystanie JIW w tworzeniu struktury hierarchicznych narzędzi dostępu

przedmiotowego, jest zdominowane przez udział języka informacyjnego typu klasyfikacji. W tym wariancie mamy do czynienia z wykorzystaniem organizacji pola semantycznego wybranej klasyfikacji piśmiennictwa do uporządkowania kategorii tematycznych w narzędziu. To **dedukcyjne** podejście polega na dystrybucji kategorii tematycznych w strukturze wyznaczonej przez podział uniwersum zastosowany w danej klasyfikacji. Organizacja kategorii tematycznych w hierarchicznym narzędziu dostępu przedmiotowego jest wypadkową i również niejako konsensusem przejętej systematyzacji klas z konkretnego JIW typu klasyfikacji oraz kierowania się zasadą *literary warrant*. Tego rodzaju struktury są mało elastyczne, ale trzon systematyzacji przejęty z klasyfikacji zapewnia im spójność.

Metodyka adaptacji schematów klasyfikacyjnych w hierarchicznych narzędziach dostępu w DSH zakłada przede wszystkim redukcję tego rodzaju struktur. Przejawia się to zgodnością z wykorzystanym schematem do określonego stopnia podziału (zazwyczaj do trzeciego lub czwartego poziomu - np. BUBL, MathGuide). Dalszy podział ma na celu zapobieganie nadmiernemu rozproszeniu tematów. W związku z tym, wyodrębnianie nowych kategorii tematycznych następuje wyłącznie według zasady *literary warrant* oraz z wykorzystaniem relacji hierarchii tematycznej.

Zaadaptowany schemat klasyfikacyjny wyznacza perspektywę strukturalizacji zbioru kategorii tematycznych. W DSH, gdzie nie stosuje się quasifasetyzacji, rozbudowa kategorii tematycznych poprzez dołączenie do przejętej klasy symbolu pomocniczego odbywa się na tym samym stopniu podziału (Rys. 6.18).

---

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

---

## 610 Medical sciences, medicine

[610 Medicine: general resources](#)

[610 Medicine: alternative medicine](#)

[610 Medicine: education](#)

[610 Medicine: discussion lists](#)

[610 Medicine: journals](#)

[610 Medicine: libraries](#)

[610 Medicine: national centres](#)

[610 Medicine: research](#)

[610 Medicine: societies](#)

[610 Medical and health information management](#)

[610 Medical imaging and illustration](#)

[610.28 Biomedicine](#)

[610.73 Nursing](#)

[611 Human anatomy, cytology \(cell biology\), histology \(tissue biology\)](#)

[612 Human physiology](#)

[613 Promotion of health](#)

[614 Forensic Medicine, Incidence of disease, preventative medicine](#)

[614.4 Epidemiology](#)

[615 Pharmacology](#)

[615.5 Therapeutics](#)

Rys. 6.19. Rozbudowa kategorii tematycznej 610 Nauki medyczne w systemie BUBL.

[Źródło: <http://bubl.ac.uk>]

Przedstawiona na Rys. 6.19 rozbudowa kategorii tematycznej *610 Nauki medyczne* oparta jest na podziale tej klasy w Klasyfikacji Dziesiętnej Deweya, która stanowi wzorzec dla struktury hierarchicznego narzędzia dostępu przedmiotowego w tym systemie. Zasada *literary warrant* jest tutaj widoczna w postaci zestawu klas, jakie przejęto z KDD. Nie stanowi to zestawu wszystkich klas, jakie wyodrębniono w klasie 610 w KDD. Nie mniej, wyróżniona kategoria tematyczna *610 Medicine: journals (Medycyna: czasopisma)* jest ewidentnym przykładem zastosowania symbolu pomocniczego formy, który dla tej klasy w KDD ma postać 610.5. Takie rozwiązanie zastosowano w wielu miejscach w strukturze narzędzia dostępu przedmiotowego.

Notacja, czyli forma nazw kategorii tematycznych w narzędziach dostępu przedmiotowego opartych na istniejących klasyfikacjach cechuje się dużą niekonsekwencją ich przejmowania. Polega to na wykorzystywaniu odpowiedników słownych symboli klasyfikacyjnych, samodzielnie lub w połączeniu z odpowiadającymi im symbolami. Nie ma pełnej zgodności w formie odpowiedników słownych

w narzędziach i tablicach klasyfikacyjnych. Przyjęta forma wyrażen ma w jak najkrótszy sposób, a jednocześnie precyzyjnie określać zakres danej kategorii tematycznej. Jest to zgodne z metodyką tworzenia taksonomii internetowych [por. Zhonghong, Chaudhry, Khoo 2007], której szkielet jest wyznaczany przez podział uniwersum przyjętej klasyfikacji, a forma nazw kategorii tematycznych jest przejmowana np. z tezaursów lub słowników innych JIW o notacji paranaturalnej. W omawianym typie narzędzi dostępu przedmiotowego praktykuje się takie rozwiązanie, ale w sposób niejawni. Dokumentacje DSH identyfikują jedynie konkretne języki informacyjno-wyszukiwawcze wykorzystane do tego celu, bez wskazania ich rzeczywistego udziału w tym zakresie.

Spotyka się również technikę zagnieżdżania symbolu klasyfikacyjnego dla odpowiedniej kategorii tematycznej w jej adresie URL i wywołaniu jej zawartości z bazy danych. Zjawisko to występuje jednak stosunkowo rzadko, ale z pewnością jest pomocne w interpretacji organizacji samej bazy danych.

Do budowy narzędzi dostępu przedmiotowego w wariacie hierarchicznym wykorzystuje się zarówno klasyfikacje uniwersalne oraz specjalistyczne. Wśród pierwszej grupy dominuje Klasyfikacja Dziesiętna Deweya (BUBL), Uniwersalna Klasyfikacja Dziesiętna (SOSIG, EdNa Online), Klasyfikacja Biblioteki Kongresu (Intute), a także popularna wśród niemieckich systemów Göttinger Online-Klassifikation (GeoGuide). Nie mniej, wśród DSH o zakresie ograniczonym do jednej dziedziny wiedzy wykorzystuje się również klasyfikacje specjalistyczne. Są to np. Mathematics Subject Classification (MathGuide), Physics and Astronomy Classification Scheme (ViFaPhys).

W omawianym rodzaju narzędzi dostępu do zasobów DSH stosuje się również łączenie opracowanych na potrzeby systemów schematów hierarchicznych z systematyzacją klas przejętą z klasyfikacji piśmiennictwa. Polega to na opracowaniu własnego schematu kategorii, w którym tylko wybrane wybrane kategorie tematyczne są rozbudowywane zgodnie z podziałem zbieżnych z nimi zakresowo klas ze schematów klasyfikacyjnych. Taka sytuacja ma miejsce w opracowanej na potrzeby systemu Intute: nauka, inżynieria i technologia strukturze hierarchicznej narzędzia dostępu przedmiotowego. Na pierwszym stopniu podziału wyodrębniono m.in. kategorię tematyczną *Computing* (przetwarzanie danych), której rozbudowa jest zgodna z podziałem klas, jaki zastosowano w ACM Computing Classification Systems. Rozbudowa pozostałych kategorii tematycznych nie jest oparta systematyzację tej klasyfikacji. W tej funkcji schematy klasyfikacyjne występują jako narzędzia



pomocnicze w organizowaniu kategorii tematycznych w narzędziach dostępu przedmiotowego.

Przyjęcie rozległej perspektywy interpretacji wykorzystania JIW w hierarchicznym wariacie narzędzi dostępu przedmiotowego pozwala na odniesienie dotychczasowych rozważań do realizacji przez JIW funkcji wyszukiwawczej i organizacyjnej. Przy czym to funkcja systematyzacji kategorii tematycznych poprzedza możliwość realizacji funkcji heurystycznej. Specyficzne wykorzystanie JIW polega na priorytetowym traktowaniu jego zasobu leksykalnego jako potencjalnego źródła nazw dla kategorii tematycznych. Wykorzystanie organizacji pola semantycznego wybranych klasyfikacji do opracowywania struktur narzędzi nacechowane jest w dużym stopniu redukcjonizmem, któremu towarzyszy dążenie do zapewnienia maksymalnej funkcjonalności struktur często kosztem ich spójności.

### **6.3. *Wariant hybrydowy***

W narzędziach dostępu przedmiotowego można zaobserwować również inne rozwiązania dystrybucji punktów dostępu, niż omówione dotychczas struktury płaskie i hierarchiczne. Występują one stosunkowo rzadko, ale prezentują interesujące rozwiązania dotyczące samej organizacji punktów dostępu jak i realizacji procesu pozyskiwania informacji z systemu.

Grupa reprezentatywna, którą poddano szczegółowej analizie pod kątem zastosowanych rozwiązań w narzędziach dostępu przedmiotowego zawierała jedynie dwa systemy, w których stwierdzono obecność takich narzędzi. Były to systemy CISMef ([www.chu-rouen.fr/cismef/](http://www.chu-rouen.fr/cismef/)) oraz GEM Gateway to 21st Skills ([www.thegateway.org/](http://www.thegateway.org/)). Szczegółowa charakterystyka tych systemów została przedstawiona w poprzednim rozdziale. W tym miejscu zostaną jedynie podsumowane przeprowadzone analizy.

Hybrydowy wariant narzędzia dostępu przedmiotowego polega tutaj na połączeniu podejścia hierarchicznego z płaskim. Narzędzia dostępu przedmiotowego niekwalifikujące się do zasygnalizowanych dwóch typów charakteryzuje zastosowanie zaawansowanych sposobów reprezentacji zależności semantycznych pomiędzy kategoriami tematycznymi, które wykorzystywane są podczas konstrukcji zapytania informacyjnego przez użytkownika.

W przypadku systemu GEM Gateway to 21st Skills zastosowano

sześciorozmiarową strukturalizacją zbioru wyszukiwawczego za pomocą kategorii leksykalnych, którym odpowiadały nazwy:

- typ,
- poziom,
- słowa kluczowe,
- pośrednik,
- beneficjent,
- koszt dostępu.

Dla każdej z nich opracowano na podstawie zasobu leksykalnego wybranych JIW oraz informatorów bezpośrednich listę kontrolowanych potencjalnych wartości. W ten sposób narzędzie dostępu przedmiotowego wykorzystywało jednocześnie sześć wspomnianych kategorii leksykalnych. Budowanie zapytania nie polega tutaj na eksploracji tej struktury, lecz na wyborze z każdej z „faset” punktu dostępu, którego wartość reprezentuje konkretny aspekt poszukiwań użytkownika. Dynamiczne dołączanie do zapytania terminów z faset powoduje wyodrębnianie tych ChWD, które w odpowiednim polu rekordu zawierają poszukiwane wartości. Relacją semantyczną łączącą wybrane podczas wyszukiwania terminy była ich kookurencja w ChWD w postaci określonej wartości, wyłącznie w ramach danego pola rekordu bazy danych. Specyfiką takiej struktury narzędzia dostępu przedmiotowego jest jego duża dynamika i wysoka interaktywność podczas wyszukiwania informacji. Zbiór terminów wybranych jako elementy strukturalne narzędzia powstał poprzez podział zbioru wyszukiwawczego według wspomnianych sześciu wymiarów. Siłą tego narzędzia jest dynamiczne modyfikowanie zakresu zapytania informacyjnego poprzez działanie iloczynu logicznego na zbiorach ChWD reprezentowanych przez określone terminy pełniące funkcje punktów dostępu. Pozwala to jednocześnie na bieżące analizowanie rezultatów wyszukiwania.

W drugim przypadku (CISMef), struktura narzędzia dostępu przedmiotowego opiera się na systematyzacji słownictwa MeSH, które wykorzystano do indeksowania. Tak, jak w wariacie ahierarchicznym dostęp do kolekcji realizowany jest poprzez alfabetyczny indeks wykorzystanych w indeksowaniu terminów. Różnica jednak polega na tym, że dla całego zbioru terminów pochodzących ze słownika MeSH opracowano zestaw kategorii tematycznych nadrzędnych, które nazwano metaterminami. Odpowiadają im zakresy głównych specjalności w ramach medycyny i biologii. Każde wyrażenie MeSH zostało przyporządkowane do jednego metaterminu na podstawie relacji hierarchii tematycznej. Powstała w ten sposób dwupoziomowa

struktura umożliwi eksplorację podobnie jak, w przypadku struktur hierarchicznych. Niewątpliwą zaletą narzędzia dostępu przedmiotowego CISMef jest wykorzystanie sieci relacji paradygmatycznych zachodzących pomiędzy jednostkami leksykalnymi w MeSH, które wyznaczają kierunki eksploracji narzędzia w poszukiwaniu przez użytkownika kategorii tematycznej, która grupuje relewantne dla użytkownika ChWD.

Wariant hybrydowy reprezentowany tutaj przez dwa wspomniane DSH cechuje heterogeniczność na płaszczyźnie strukturalnej. Dwa scharakteryzowane narzędzia zasadniczo różnią się od siebie. Wariant ten to jednak interesujący kierunek zastosowania JIW w organizacji punktów dostępu oraz realizacji procesu wyszukiwania informacji. Funkcje JIW pozostają niezmiennie, jednak modyfikacji ulegają sposoby i narzędzia ich realizacji.

#### **6.4. Podsumowanie**

Przedstawione w tej części rozprawy wyniki badań nad budową narzędzi dostępu przedmiotowego uwzględniały trzy warianty organizacji punktów dostępu. Każdy z nich reprezentował odmienne cechy strukturalne i tym samym różne sposoby realizacji funkcji wyszukiwawczych.

W wariancie ahierarchicznym (płaskim) punkty dostępu do kolekcji były generowane do postaci alfabetycznego wykazu jednostek leksykalnych użytych podczas indeksowania przedmiotowego dokumentów. Reprezentowały one wymiar treści oraz formy elementów zbioru informacyjnego. Dystrybucja punktów dostępu w narzędziu uwarunkowana była ich liczebnością, co powodowało tworzenie sekcji grupujących jednostki leksykalne w grupach odpowiadającym kolejnym literom alfabetu łacińskiego. Udział JIW w tym wariancie to przede wszystkim traktowanie jego słownika jako źródła leksyki do indeksowania i generowania na jego podstawie narzędzia dostępu do zasobów zagnieżdżonego w strukturze hipertekstu.

Narzędzia w wariancie hierarchicznym wykazują bardziej złożony obraz organizacji punktów dostępu. Tutaj kategorie tematyczne są włączane w struktury drzewiaste opracowane na potrzeby systemu lub dostosowywane na podstawie organizacji pola semantycznego wybranego JIW.

Relacją porządkującą zbiór kategorii tematycznych jest relacja hierarchii tematycznej, a zasadą ich wyodrębniania jest *literary warrant*. Chociaż w wyniku badań uzyskano do pewnego stopnia spójny obraz narzędzia dostępu przedmiotowego, to odnotowano również odstępstwa od zidentyfikowanych reguł. W wariancie tym

odnotowano znaczący udział JIW zarówno w warstwie „zaplecza leksykalnego” wykorzystywanego do tworzenia nazw dla kategorii tematycznych, jak i w wykorzystaniu organizacji jego pola semantycznego w systematyzacji kategorii tematycznych. Wykorzystuje się tutaj przede wszystkim język informacyjno-wyszukiwawczy typu klasyfikacji jako podstawę organizacji kategorii tematycznych w strukturze drzewiastej. Metodyka adaptacji klasyfikacji w tego rodzaju narzędziach dostępu jest różna w odniesieniu do poszczególnych systemów. Najczęściej jednak polega na redukcji systematyki klas przejętej z danej klasyfikacji piśmiennictwa do trzy- lub czteropoziomowej struktury, która dalej jest rozbudowywana zgodnie z zasadą *literary warrant*. W nielicznych systemach (np. AERADE) wariant hierarchiczny ma również narzędzie pomocnicze w postaci alfabetycznego wykazu kategorii tematycznych (indeksu rzeczowego), jednak tego typu rozwiązanie jest stosowane bardzo rzadko.

W odniesieniu do wariantu hybrydowego, wyniki badań nie wskazują na występowanie określonych zależności strukturalnych. Jego podstawową zaletą jest duża elastyczność, która jest wynikiem wielowymiarowej strukturalizacji pola tematycznego, które reprezentują jednostki leksykalne przejmowane ze słowników wybranych JIW. Udział JIW pokazuje jednak nowy sposób organizacji punktów dostępu opartej na fasetowym modelu organizacji wiedzy.

## Zakończenie

Celem rozprawy było określenie roli języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych. Przyjęta w pracy koncepcja organizacji informacji wymagała zidentyfikowania trzech obszarów zagadnień: formatowanie jednostki opisu, reprezentacja metainformacji oraz organizacja dostępu. Celem organizacji informacji jest udostępnienie użytkownikowi systemu informacyjnego strukturalnych narzędzi dostępu do jego zasobów. Weryfikacja hipotezy postawionej w rozprawie została dokonana na podstawie tak sformułowanych ram interpretacyjnych. W pracy wskazano rolę i sposoby wykorzystania języka informacyjno-wyszukiwawczego odpowiednio w identyfikacji jednostki opisu, charakterystyce jej cech treściowych i formalnych oraz w budowie narzędzi dostępu do zasobów dziedzinowych systemów hipertekstowych.

Heterogeniczny charakter zasobów WWW, którego elementy stanowią zbiór informacyjny omawianej klasy systemów informacyjno-wyszukiwawczych uniemożliwia zastosowanie „paradygmatu dokumentu”, w którym jest on jednostką poddawaną charakterystyce i organizacji. Termin ten utożsamia się z fizycznym utrwaleniem treści, co w przypadku WWW nie do końca odzwierciedla charakter jego zasobów. Przyjęty w rozprawie termin *obiekt informacyjny* pozwolił na uniknięcie niepożądanych konotacji oraz uwzględnił zróżnicowanie formalne i strukturalne gromadzonych przez systemy zasobów. Ma to swoje odzwierciedlenie w stosowanym formacie opisu danych. Formatowanie jednostki opisu w dziedzinowych systemach hipertekstowych polega na zastosowaniu języka opisu bibliograficznego uzupełnianego o dodatkowe atrybuty obiektów podnoszące wartość informacyjną ich charakterystyk. Systemy, o których mowa należą do klasy dokumentacyjnych systemów informacyjno-wyszukiwawczych udostępniających użytkownikowi na wyjściu metainformacje. Dlatego też rodzaj informacji odwzorowywanych podczas indeksowania ich zasobów ma duży wpływ zarówno na realizację procesów wyszukiwania informacji oraz na ocenę przez użytkownika rezultatów wyszukiwania. Rola języka informacyjno-wyszukiwawczego w tym zakresie (a dokładnie języka opisu bibliograficznego) polega na zastosowaniu standaryzowanych schematów metadanych, które stanowią o potencjale wyszukiwawczym elementów zbioru informacyjnego. Umożliwia to także nadanie wybranym atrybutom obiektów statusu punktów dostępu, utożsamianych z cechami wyszukiwawczymi, aktywizowanymi w mechanizmach pozyskiwania

informacji z systemu. Modelowanie danych z wykorzystaniem języka opisu bibliograficznego, jako typu języka informacyjno-wyszukiwawczego, ma wymiar semantyczny oraz syntaktyczny. Schemat metadanych wyznacza semantyczną strukturę jednostki opisu poprzez identyfikację jej atrybutów. Odwzorowanie wartości dla każdego z nich podczas indeksowania wiąże się przyjęciem reguł budowania takiej reprezentacji. To, wraz z zasadami przypisywania i formą wartości dla określonego atrybutu, określa syntaktyczny wymiar modelowania danych w dziedzinowych systemach hipertekstowych. Rola języka informacyjnego polega tutaj na wykorzystaniu standaryzowanych języków opisu bibliograficznego (często w postaci schematu metadanych Dublin Core) jako podstawy do tworzenia dla danego systemu jego zmodyfikowanej wersji. Sprowadza się to do dołączania do jego struktury dodatkowych atrybutów obiektów informacyjnych lub rezygnację z wybranych o małej wartości wyszukiwawczej.

Druga płaszczyzna interpretacji zastosowania JIW w organizacji informacji w dziedzinowych systemów hipertekstowych, wynika z realizacji jego podstawowej funkcji, jaką jest odwzorowanie treści i/lub formy obiektów informacyjnych. Rola JIW w reprezentacji informacji w DSH polega na wykorzystaniu ich zasobów leksykalnych do odwzorowania wartości dla wybranych kategorii atrybutów zidentyfikowanych przez język opisu bibliograficznego. Reprezentacja cech wymiaru treści odbywa się w DSH przy użyciu czterech typów JIW: języka haseł przedmiotowych, języka informacyjno-wyszukiwawczego typu klasyfikacji, języka deskryptorowego oraz języka słów kluczowych w postaci kontrolowanej oraz niekontrolowanej. Można tutaj zaobserwować znaczący udział klasyfikacji oraz języka słów kluczowych (zarówno w odmianie kontrolowanej jak i swobodnej). Reprezentacja wiedzy skodyfikowanej w obiektach informacyjnych DSH z zastosowaniem języków informacyjnych o notacji paranaturalnej charakteryzuje się:

- bezpośrednim oznaczaniem tematów i dodatkowych cech wymiaru treści,
- jednoczesną redukcją gramatyczną tych zdań JIW do postaci jednostek autosyntaktycznych i znacznym ograniczeniem roli jednostek pomocniczych.

Udział języków informacyjno-wyszukiwawczych typu klasyfikacji w odwzorowaniu treści obiektów informacyjnych cechuje specyficzne użycie jego leksyki. Polega to na wyznaczaniu zakresów tematycznych za pomocą odpowiednich symboli klasyfikacyjnych, ale przy priorytetowym traktowaniu ich odpowiedników słownych pobieranych z tablic. Oznacza to, że zakres tematyczny obiektu informacyjnego podczas katalogowania jest wyznaczany za pomocą symbolu klasyfikacyjnego,

a w charakterystyce wyszukiwawczej wyświetlanej użytkownikowi eksponuje się jego odpowiednik w języku naturalnym. Tutaj również występują uproszczenia syntaktyczne w tworzeniu symboli klasyfikacyjnych w postaci rezygnacji z symboli pomocniczych.

Udział JIW w katalogowaniu zasobów DSH to przede wszystkim wykorzystanie języków o szerokim zasięgu stosowania. Wśród znanych i ugruntowanych JIW są narzędzia leksykalne o charakterze uniwersalnym (np. LCSH, UKD, KBK, KDD, Göttinger Classification System), a także specjalistycznym (np. MeSH, Tezaurus CAB, Tezaurus HASSET, Royal College of Nursing Thesaurus).

Powszechnym zjawiskiem w DSH jest polireprezentacja wiedzy, która polega na reprezentacji cech treści za pomocą kilku narzędzi. Obok charakterystyki słownej są to teksty różnych JIW stosowanych w systemie. Polireprezentacja wiedzy w DSH przyjmuje formę:

- jawną z określonym *explicite* udziałem konkretnych JIW w indeksowaniu oraz
- ukrytą, gdzie zbiór słownictwa przyjęty do reprezentacji wiedzy jest połączeniem podzbiorów słowników różnych JIW.

Do odwzorowywania cech pozat treściowych w DSH stosuje się zbiory słownictwa kontrolowanego w postaci kartotek wzorcowych nazw osobowych, geograficznych oraz dokumenty normalizacyjne w postaci zbioru jednostek leksykalnych stosowanych do reprezentacji określonych cech, np. nazw języków etnicznych. Zastosowanie kartotek wzorcowych do kontroli wspomnianych cech ma jednak miejsce stosunkowo rzadko i dotyczy przede wszystkim systemów o dużych zbiorach wyszukiwawczych. Do oznaczenia typów obiektów informacyjnych stosuje się opracowywane na potrzeby systemów listy wartości, których źródłem są również odpowiednie wyrażenia pochodzące ze słowników języków informacyjno-wyszukiwawczych o notacji paranaturalnej. Rezygnacja z odwzorowania cech formalnych obiektów informacyjnych w zdaniach JIW podyktowana jest strukturą języka opisu danych, gdzie identyfikuje się odrębne atrybuty odpowiedzialne za te cechy. W związku z tym, w celu uniknięcia redundancji informacji, funkcję odwzorowania formy obiektów informacyjnych pełni odpowiedni atrybut w stosowanym schemacie metadanych.

Trzecia płaszczyzna interpretacji zastosowania JIW w organizacji informacji w dziedzinowych systemach hipertekstowych wynika z realizacji jego funkcji wyszukiwawczej. Pozyskiwanie informacji z zasobów DSH realizują mechanizmy wyszukiwawcze (w postaci wyszukiwarek) oraz narzędzia dostępu przedmiotowego. Udział JIW w mechanizmach wyszukiwawczych DSH jest ograniczony i sprowadza się

do „asystowania” podczas budowy zapytania informacyjnego. Polega to na odpowiedniej budowie interfejsu użytkownika, która umożliwia tworzenie zapytania informacyjnego poprzez wybór adekwatnych terminów ze stosowanych w systemie zbiorów słownictwa kontrolowanego. Zazwyczaj są to listy wartości dla cechy *typ obiektu* lub zestaw kategorii tematycznych umożliwiających wyszukanie informacji według przyjętego kryterium treściowego. W niewielu przypadkach w mechanizmach wyszukiwawczych adaptuje się słownik JIW stosowanego do indeksowania zasobów. Umożliwia to pobieranie przez użytkownika terminów i konstrukcję zapytania informacyjnego. Takie zastosowanie JIW miało miejsce w odniesieniu do języka deskryptorowego w systemach o dużym zbiorze wyszukiwawczym (np. <http://www.intute.ac.uk>). Jediną relacją wiążącą użycie pobranych terminów było ich współwystępowanie w odpowiednim polu rekordu.

Drugi rodzaj narzędzi dostępu polega na organizacji punktów dostępu w postaci kategorii tematycznych w środowisku hipertekstowym. Każda z nich jest węzłem dla odesłania hipertekstowego. Jego aktywacja powoduje wyświetlenie zgrupowanych tam odpowiednich charakterystyk wyszukiwawczych obiektów informacyjnych. Pozyskiwanie informacji za pomocą tego typu narzędzi odbywa się poprzez eksplorację takiej struktury z wykorzystaniem techniki przeglądania (ang. browsing). Ze względu na sposób porządkowania punktów dostępu w narzędziach dostępu przedmiotowego zidentyfikowano trzy warianty ich budowy: wariant ahierarchiczny, hierarchiczny oraz hybrydowy. W budowie każdego z nich zidentyfikowano i scharakteryzowano udział języka informacyjno-wyszukiwawczego.

Wariant ahierarchiczny polega na generowaniu alfabetycznego wykazu punktów dostępu na podstawie wartości odpowiedniego atrybutu w strukturze rekordu bazy danych, służącego do reprezentacji treści obiektu informacyjnego. Stworzone w ten sposób narzędzie ma postać zbliżoną do indeksu przedmiotowego lub katalogu przedmiotowego. Funkcja odsyłania jest tutaj realizowana automatycznie za pomocą hiperłączy. Udział JIW w budowie tego wariantu narzędzia dostępu polega na użyciu jego leksyki bezpośrednio podczas indeksowania, a pośrednio do generowania na jej podstawie wykazu punktów dostępu. Tym samym narzędzie to tworzy podzbiór słownika danego JIW, w takim zakresie, w jakim jego elementy zostały użyte w charakterystykach wyszukiwawczych.

Wariant hierarchiczny budowy narzędzi dostępu przedmiotowego charakteryzuje się dużym zróżnicowaniem strukturalnym i większym wpływem JIW na jego budowę niż w poprzednim przypadku. Tworzy go uporządkowany hierarchicznie



zbiór kategorii tematycznych pełniących funkcję punktów dostępu. W przeciwieństwie do wariantu ahierarchicznego jest on projektowany. Wyodrębnianie kategorii tematycznych następuje według zasady *literary warrant*. Prowadzi to do utworzenia pola tematycznego systemu, którego reprezentacją jest struktura kategorii tematycznych. Relacją organizującą zbiór kategorii jest relacja hierarchii tematycznej. Odnotowano również przypadki wykorzystania niespecyfikowanych relacji asocjacyjnych do tworzenia dodatkowych powiązań semantycznych między kategoriami. Udział JIW w tym wariantcie polega na pełnieniu funkcji źródła słownictwa do tworzenia nazw dla kategorii tematycznych. W wielu przypadkach sposób porządkowania kategorii tematycznych w narzędziu oparty był na organizacji pola semantycznego wybranego JIW. Polega to najczęściej na przejmowaniu z istniejących klasyfikacji piśmiennictwa podziału klas do określonego stopnia podziału i budowie na tej podstawie hierarchicznego układu kategorii tematycznych. W tym celu wykorzystuje się najbardziej znane klasyfikacje uniwersalne (KDD, KBK, UKD) oraz funkcjonujące klasyfikacje specjalistyczne (np. Mathematics Subject Classification, ACM Computing Classification Systems, Physics and Astronomy Classification Scheme (PACS)). Metodyka tworzenia hierarchicznego wariantu narzędzi dostępu przedmiotowego polega na tworzeniu własnej struktury (w oparciu o leksykę określonego JIW) lub na adaptacji systematyzacji danej klasyfikacji do potrzeb systemu. Analiza ilościowa budowy tego rodzaju narzędzi umożliwiła syntetyzację (poprzez uogólnienie) modelowej struktury narzędzia dostępu przedmiotowego reprezentującego ten typ. Jest to narzędzie dostępu:

- zbudowane średnio z 633 kategorii tematycznych,
- wykorzystujące maksymalnie siedmiostopniowy system podziału,
- posiadające na pierwszym stopniu podziału średnio 14 kategorii tematycznych,
- gdzie 88% łącznej liczby wyodrębnionych kategorii tematycznych zostało zgromadzonych na poziomach od drugiego do czwartego,
- gdzie 90% wszystkich ChWD zostało przyporządkowanych do kategorii tematycznych na poziomach od drugiego do czwartego,
- gdzie średnia pojemność kategorii tematycznej wynosi około 14 ChWD i wartość ta zależy od jej miejsca w strukturze,
- gdzie jedna ChWD została przyporządkowana średnio do dwóch kategorii tematycznych w schemacie.

Wariant hybrydowy budowy narzędzi dostępu przedmiotowego stosuje się bardzo rzadko. Został wyodrębniony w tej typologii z uwagi na brak możliwości

przyporządkowania niektórych z analizowanych narzędzi do dwóch wspomnianych wariantów. Nie ma podstaw do opracowania jego modelowej wersji z uwagi na heterogeniczność strukturalną narzędzi kwalifikujących się do tego typu. Jednakże jego zastosowanie, jako podstawy do tworzenia narzędzia dostępu, cechuje się interesującymi sposobami wykorzystania JIW. Jego charakterystyczną cechą jest wielowymiarowa charakterystyka pola tematycznego kolekcji, co przekłada się na wybór wielu typów punktów dostępu. Dodatkowo, JIW ma tutaj również zastosowanie zarówno jako źródło słownictwa oraz jako wzorzec do projektowania układu kategorii tematycznych.

Wariant ahierarchiczny reprezentuje ujęcie przedmiotowe, hierarchiczny - aspektowe, zaś wariant hybrydowy - aspektowo-przedmiotowe. Warianty ahierarchiczny i hierarchiczny to egzemplifikacja statycznych struktur organizacji wiedzy. [por. Vickery 2008] Alfabetyczny wykaz punktów dostępu nie reprezentuje żadnego porządku rzeczywistości dokumentacyjnej i nie odwzorowuje z założenia żadnych zależności między nimi. Układy hierarchiczne, chociaż są tworzone na podstawie segmentacji pola tematycznego systemu, to charakteryzują się małą elastycznością, zarówno pod względem reprezentacji zależności pomiędzy tematami, jak również w asymilacji nowych zagadnień (bez burzenia spójności dotychczasowego układu kategorii). Kierowanie się zasadą *literary warrant* w tworzeniu hierarchicznych układów kategorii tematycznych jest obarczone ryzykiem niespójności całego układu (problem pustych kategorii tematycznych). Jedyny element dynamiki systemu organizacji wiedzy reprezentowanego przez ten wariant to zjawisko *quasifasetyzacji*. Możliwość reorganizacji struktury kategorii tematycznej poprzez zastosowanie fasetowej organizacji wiedzy z pewnością poprawia efektywność narzędzia i czyni je bardziej elastycznym. Wariant hybrydowy, szczególnie reprezentowany przez narzędzie dostępu w systemie GEM – Gateway to 21st Century Skills, to przykład dynamicznych systemów organizacji wiedzy. W pełni fasetowa organizacja pola tematycznego i jej zastosowanie w narzędziu pozyskiwania informacji umożliwia wielowymiarową interpretację treści obiektu informacyjnego, co przekłada się na dodatkowe możliwości wyszukiwawcze. Poważnym ograniczeniem w tym przypadku jest niemożność ustanowienia zależności semantycznych pomiędzy kategoriami z różnych faset. Kookurencja w charakterystyce wyszukiwawczej jest kryterium reprezentującym wyłącznie aspekt techniczny (nie zaś znaczeniowy) zależności pomiędzy wyrażeniami wyszukiwawczymi.

W szerszym kontekście rozważań nad organizacją wiedzy w DSH można mówić

o postmodernistycznym ujęciu tego zjawiska [zob. np. Mai 2002] w omawianej klasie systemów informacyjno-wyszukiwawczych. Polega to na odrzuceniu zobiektywizowanego i neutralnego sposobu prezentacji struktury wiedzy skodyfikowanej w treści obiektów informacyjnych na rzecz pragmatyzmu w tego rodzaju obrazie. Zastosowanie tradycyjnych systemów leksykalnych, wywodzących się z bibliotek i ośrodków dokumentacji, do reprezentacji wiedzy zapisanej w obiektach informacyjnych ma zapewnić spójność pojęciową i terminologiczną. W tym celu wykorzystuje się tutaj również słownictwo z wyróżnionymi *explicite* relacjami paradygmatycznymi umożliwiające m.in. tworzenie narzędzi dostępu przedmiotowego. Z drugiej jednak strony mamy do czynienia z polireprezentacją wiedzy, zasadą *literary warrant* oraz redukcją gramatyczną tekstów JIW. Ma to służyć maksymalnemu uproszczeniu korzystania z tego rodzaju systemów i oferowanych przez nie funkcji. Uniwersum wiedzy [por. Miksa 1998] jest tutaj reprezentowane w skali mikro. Systemy organizacji wiedzy w DSH odwzorowują wyłącznie obraz takiej rzeczywistości, którą **aktualnie** tworzy suma przedstawiń w poszczególnych obiektach informacyjnych.

Przedstawione wnioski z badań przeprowadzonych przez autora pozwalają na pozytywną weryfikację postawionej we wstępnej części rozprawy tezy o istotnej roli języka informacyjno-wyszukiwawczego w organizacji informacji w dziedzinowych systemach hipertekstowych.

Wyniki badań przeprowadzonych przez autora mogą pomóc w praktycznym rozwiązywaniu wielu problemów dotyczących organizacji informacji w dziedzinowych systemach hipertekstowych. Dodatkowo opracowana metodyka badań narzędzi dostępu przedmiotowego może być wykorzystana w analizie taksonomii internetowych jako typu sieciowego systemu organizacji wiedzy.

Perspektywy dalszych badań na tym polu mogą zdaniem autora przyjąć dwa kierunki. Pierwszy może dotyczyć wykorzystania dorobku teoretycznego i praktycznego języków informacyjno-wyszukiwawczych w tworzeniu narzędzi dostępu. Celem tego rodzaju badań może być zastosowanie metodyki tworzenia indeksów oraz katalogów przedmiotowych i systematycznych w projektowaniu taksonomii internetowych. Drugi kierunek badań autor upatruje w formalizacji struktury konceptualnej sieciowych systemów organizacji wiedzy. Formalne modelowanie pola tematycznego systemu z wykorzystaniem leksyki i organizacji pola semantycznego JIW, a także formalnego języka reprezentacji wiedzy (np. SKOS – Simple Knowledge Organization System) pozwoliłoby na automatyczne przetwarzanie danych

i wprowadzenie dynamicznych struktur w narzędziach dostępu przedmiotowego. Dodatkowo interesującym kierunkiem badań jest połączenie tradycyjnych systemów organizacji wiedzy oraz tagowania\* w jednym narzędziu wyszukiwania i organizacji informacji. Przykładem tego kierunku badań jest np. projekt EnTag. [Golub, Jonem, Lykke Nielsen 2008]

---

\* Tagowanie - nadawanie elementom kolekcji etykiet w postaci swobodnych słów kluczowych przez użytkowników.

## Bibliografia załącznikowa

1. Abramowicz, W. Retoryka hipertekstu. *Informatyka*. 1996, nr 3, s. 4-7
2. Adams, M. (red.). *World Wide Web: beyond the basics*. New Jersey: Prentice-Hall, 1998, ISBN 0-13-954785-1
3. Anderson, J. *Organization of knowledge*. W: *International Encyclopedia of Information and Library Science*. London: Routledge, 2003 s. 471-490
4. Anderson, J.; Perez-Carballo, J. The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part 1: Research, and the nature of human indexing. *Information Processing & Management*. 2001, vol. 37, s. 231-254
5. Anderson, J.; Perez-Carballo, J. *Information Retrieval Design*. St. Petersburg: Ometeca Institute, 2005. ISBN 0976354705
6. *ANSI/NISO Z39.19-2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* [online]. 2005, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [www.niso.org/standards/resources/Z39-19-2005.pdf](http://www.niso.org/standards/resources/Z39-19-2005.pdf)
7. Antelman, K.; Lynema, E.; Pace, A. Toward Twenty-First Century Library Catalog. *Information Technology and Libraries*. 2006, nr 3, s. 128-139
8. Australian Subject Gateways: Definitions for Web-Based Services [online]. [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.nla.gov.au/initiatives/sg/servicetypes.html>
9. *Authority and Source Detail for "Resource Type" Element* [online]. 2000, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ifla.org/udt/dc8/resource.pdf>
10. Babik, W. Struktura pola semantycznego języka informacyjno-wyszukiwawczego dla materiałów kartograficznych w świetle potrzeb ich użytkowników. *Zagadnienia Informatyki Naukowej*. 1992, vol. 60, nr 2, s. 47-65
11. Babik, W. *Generowanie języków informacyjno-wyszukiwawczych ze słowników terminologicznych*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 1996. ISBN 83-233-0981-7
12. Babik, W. *Języki informacyjno-wyszukiwawcze - charakterystyka funkcjonalno-strukturalna i metodologiczna*. W: *Informacja naukowa: rozwój - metody - organizacja*. Warszawa: Wydaw. SBP, 2006 s. 191-218
13. Baeza-Yates, R. ; Ribeiro-Neto, B. *Modern Information Retrieval*. New York: ACM Press, 1999. ISBN 020139829X
14. Barthes, R. *S/Z*. Warszawa: Wydaw. KR, 1999. ISBN: 83-86989-38-6
15. Bates, M. Subject Access in Online Catalogs: A Design Model. *Journal of the American Society for Information Science*. 1986, vol. 37, nr 11, s. 357-376
16. Bates, M. The design of browsing and berrypicking techniques for the online search interface. *Online Review*. 1989, vol. 13, nr 5, s. 407-424
17. Bates, M. Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution. W: *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science*. Greenwood Village: Libraries Unlimited, 2002 s. 137-150.
18. Bates, M. What is browsing— really? A model drawing from behavioural science research. *Information Research*. 2007, vol. 12, nr 4. Dostępny w World Wide Web: <http://InformationR.net/ir/12-4/paper330.html>
19. Becker, H.; Braun, J.; Klaproth, F. *Evaluation report of partner subject gateways*. [online] 2000, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://renardus.sub.uni-goettingen.de/wp6/d6.1/index.html>
20. Belcher, M. [i in.]. *Information Gateways Handbook* [online]. 1999, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.desire.org/handbook/>
21. Bliss, H.E. *Organization of Knowledge in Libraries and the Subject Approach to Books*. New York: Wilson, 1933
22. Błaszczuk, H., Dziuk-Renik, E.; Kocznorowska, K. Inforol - katalog polskich rolniczych stron WWW. *Praktyka i Teoria Informatyki Naukowej i Technicznej*. 2001, vol. 9, nr 4, s. 41-43
23. Bojar, B. *Językoznawstwo dla studentów informatyki naukowej*. Warszawa: Wydaw. SBP, 2005. ISBN 83-89316-35-8
24. Borgman, Ch. Why are online catalogs still hard to use? *Journal of the American Society for Information Science*. 1996, vol. 47, nr 7, s. 493-503
25. Brooks, T.A. Where is meaning when the form is gone? Knowledge representation on the Web. *Information Research*. 2001, vol. 6, nr 2. Dostępny w World Wide Web: <http://informationr.net/ir/6-2/paper93.html>

26. Broughton, V.; Hansson, J.; Hjørland, B.; Lopez-Huertas, M. *Knowledge Organization*. W: European Curriculum Reflections on Library and Information Science Education. Copenhagen: Royal School of Library and Information Science, 2005 s. 133-148
27. Brümmer, A. *Subject Based Information Gateways* [online]. 1998, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://web.archive.org/web/20060204161407/http://www.lub.lu.se/desire/sbig.html>
28. BUBL Link. *Main subject menus*. [online] 2007, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.bubl.ac.uk/link/subjectbrowse.cfm>
29. Bush V.(1945). As we may think. *The Atlantic Monthly*. 1945, July. Dostępny w World Wide Web: <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml>
30. Campbell, D. Australian subject gateways: political and strategic issues. *Online Information Review*. 2000, nr 1, s. 73-77. ISSN 1468-4527
31. Capurro, R; Hjørland, B. The Concept of Information. *Annual Review of Information Science & Technology*. 2003, vol. 37, s. 343-411
32. Carr, L., Hall, W.; De Roure, D. The Evolution of Hypertext Link Services. *ACM Computing Surveys*. 1999, vol. 31, nr 4. Dostępny w World Wide Web: [http://www.cs.brown.edu/memex/ACM\\_HypertextTestbed/papers/19.html](http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/19.html)
33. Chapman, A. RDA: A New International Standard. *Ariadne* 2006, nr 49. Dostępny w World Wide Web: [www.ariadne.ac.uk/issue49/chapman](http://www.ariadne.ac.uk/issue49/chapman)
34. Charles, F. *CISMef: Catalog and Index of French-language Health Internet resources. A quality-controlled subject gateway*. [online] 2005, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.chu-rouen.fr/cismef/cismefeng.html>
35. Chowdhury, G. Progress in Documentation. *Digital Library Research: Major Issues and Trends. Journal of Documentation*. 1999, vol. 55, nr 4, s. 409-448
36. Chowdhury, G. ; Chowdhury, S. *Organizing Information: From the Shelf to the Web*. London: Facet Publishing, 2007. ISBN 978-1856045780
37. Chu, H. *Information Representation an Retrieval in the Digital Age*. New Jersey: Information Today, 2003. ISBN 1-57387-172-9
38. Cliff, P. *RDN Resource Types* [online]. 2002, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/pan/14883/20060912/www.rdn.ac.uk/publications/catalogue/types/index.htm>
39. Codes for the representation of names of languages [online] 2002, [dostęp: 2009-01-22] Dostępny w World Wide Web: <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=22109&ICS1=1&ICS2=140&ICS3=20>
40. Conway, S. ; Sligar, C. *Unlocking Knowledge Assets*. Redmond: Microsoft Press, 2002. ISBN 0-7356-1463-6
41. Daconta, M.; Orbst, L.; Smith, K. *The Semantic Web: a guide to the future of XML, web services and knowledge management*. Indianapolis: Wiley Publishing, 2003. ISBN 0-471-43257-1
42. Darmoni, S. *Net Scoring : crite`res de qualité de l'information de santé sur l'Internet*. [online] 1997, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.chu-rouen.fr/netscoring/>
43. Dawson, A., Simpson, J. How BUBL benefits academic librarians. *Ariadne* 1997, nr 10. Dostępny w World Wide Web: [www.ariadne.ac.uk/issue10/bubl/](http://www.ariadne.ac.uk/issue10/bubl/)
44. Day, M. *Renardus: Academic Subject Gateway Service Europe* [online]. 2000, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/metadata/renardus/>
45. Day, M.; Koch, T.; Neuroth, H. Searching and Browsing Multiple Subject Gateways in the Renardus Service. *Rc33 Sixth International Conference on Social Science Methodology, Amsterdam, Netherlands, 16-20 August 2004* [online] 2004, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [www.ukoln.ac.uk/metadata/publications/rc33-2004/renardus-paper.pdf](http://www.ukoln.ac.uk/metadata/publications/rc33-2004/renardus-paper.pdf)
46. Dempsey, L. ; Heery, R. *A review of metadata: a survey of current resource description format*. [online] 1997, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/metadata/desire/overview/>
47. Dempsey, L. A policy context - eLib and the emergence of the subject gateways. *Ariadne* 2000, nr 25. Dostępny w World Wide Web: <http://www.ariadne.ac.uk/issue25/subject-gateways/>
48. Dempsey, L.; Russel, R. A Utopian place of criticism: brokering access to network information. *Journal of Documentation*. 1999, vol. 55, nr 1, s. 33-70
49. Denton, W. *Putting Facets on the Web: An Annotated Bibliography* [online]. 2003, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.miskatonic.org/library/facet-biblio.html>
50. Derfert-Wolf, L. Serwisy tematyczne o kontrolowanej jakości w Internecie - subject gateways. *Biuletyn EBIB*. 2004, nr 6, Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2004/57/derfert.php>
51. Derfert-Wolf, L., Ganińska, H. Płoszajski, G. *Koncepcja Tworzenia Serwisu Typu Subject Gateway „BazTol - Polskie Zasoby Sieciowe z zakresu nauk technicznych”* [online] 2007,

- [dostęp: 2009-01-22]. Dostępny w World Wide Web:  
[http://www.biblos.pk.edu.pl/files/File/Koncepcja\\_BazTOL\\_wersja\\_2.pdf](http://www.biblos.pk.edu.pl/files/File/Koncepcja_BazTOL_wersja_2.pdf)
52. Drabenstott, K.M. Enhancing a New Design for Subject Access to Online Catalogs. *Library Hi Tech*. 1996, vol. 14, nr 1, s. 87-109
  53. Duval, E.; Hodgins, W.; Sutton, S.; Weibel, S. Metadata Principles and Practicalities. *D-Lib Magazine*. 2002, vol. 8, nr 4. Dostępny w World Wide Web:  
<http://www.dlib.org/dlib/april02/weibel/04weibel.html>
  54. Ellis, D; Vasconcelos, A. The Relevance of Facet Analysis for World Wide Web Subject Organization and Searching. *Journal Of Internet Cataloging*. 2000, vol. 2, nr 3/4, s. 97-114
  55. Filipek, A. Model opisu dokumentu według FRBR (Functional Requirements for Bibliographic Records). *Zagadnienia Informacji Naukowej*. 2005, nr 2, s. 13-30
  56. Fischer, T.; Neuroth, H. SSG-FI - special subject gateways to high quality Internet resources for scientific users. *Online Information Review*. 2000, vol. 24, nr 1, s. 64-68
  57. Follet, B. *Joint Funding Council's Libraries Review Group: Report (Follet Report)*. Bristol: HEFCE, 1993. Dostępny w World Wide Web:  
<http://www.ukoln.ac.uk/services/papers/follett/report/>
  58. Frohmann, B. Grounding a theory of documentation. Referat wygłoszony podczas DOCAM '03 The first annual meeting of the Document Academy. Sierpień 13-15, 2003 The School of Information Management and Systems (SIMS) at The University of California, Berkeley (nieopublikowany). Abstrakt dostępny w World Wide Web:  
<http://thedocumentacademy.org/resources/2003/abstracts/bermd.frohman.html>
  59. *Functional requirements for bibliographic records : final report*. IFLA Study Group on the Functional Requirements for Bibliographic Records. München: K.G. Saur, 1998. ISBN 978-3-598-11382-6
  60. Ganińska, H; Bajer, J.; Kołodzińska, E.; Ober, K.; Pomianowicz, J.; Tomaszczak, J. *TOL Technika Online – Informator o zasobach internetowych dla nauk technicznych : elektrotechnika: projekt Biblioteki Głównej Politechniki Poznańskiej* [online]. 2006, Dostępny w World Wide Web: <http://eprints.rclis.org/archive/00006492/>
  61. Garshol, L.M. Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*. 2004, vol. 30, nr 4, s. 378-391.
  62. Gateways Defined [online]. [dostęp: 2009-01-22]. Dostępny w World Wide Web:  
[http://web.archive.org/web/20060813164427/http://www.renardus.org/about\\_us/subject\\_gateways.html](http://web.archive.org/web/20060813164427/http://www.renardus.org/about_us/subject_gateways.html)
  63. Gliński, W. Projektowanie systemów hipermedialnych. Cz. 1 Teoria. *Praktyka i Teoria Informacji Naukowej i Technicznej*. 1994, nr 3, s. 22-27
  64. Gliński, W. *Ontologie - próba uporządkowania terminologicznego chaosu*. W: Od informacji naukowej do technologii społeczeństwa informacyjnego, red. B. Sosińska-Kalata, M. Przystek-Samokowa. Warszawa: Wydaw. SBP, 2005, s. 163-176. ISBN 83-89316-47-1
  65. Gliński, W. *Wybrane metodologie i metody budowania ontologii*. W: Informacja w sieci, red. B. Sosińska-Kalata. Warszawa: Wydaw. SBP, 2006, s. 157-206. ISBN 83-89316-66-8
  66. Głowacka, E. Kryteria i wskaźniki oceny jakości europejskich internetowych serwisów tematycznych. *Zagadnienia Informacji Naukowej*. 2007, nr 1, s. 3-14
  67. Gmiterek, G.; Kotuła, S.; Pacek, J.; Zieliński, P. *Katalogowanie Internetu na przykładzie polskiego subject gatewaya nauki o informacji* W: *Oblicza Internetu. Architektura komunikacyjna sieci*, red. Marek Sokołowski, Elbląg, 2007, s. 341-347
  68. Gold, J. Introduction to the BUBL Subject Tree. *Computers and Libraries*. 1995, vol.15, nr 6, s. 79-80
  69. Golub, K.; Jones, C.; Lykke Nielsen, M.; Matthews, B.; Moon, J.; Tudhope, D. Enhancing Social Tagging With a Knowledge Organization System. *ALISS Quarterly*. 2008, vol. 3, nr 4, s. 13-16
  70. Gorman, M.; Winkler, P. (red.). *Anglo-American Cataloguing Rules. 2nd ed.* Chicago: ALA, 1978
  71. Heery, R. Information gateways: collaboration on content. *Online Information Review*. 2000, vol. 24, nr 1 s. 40-45
  72. Heery, R. *Review of Metadata Formats*. [online] 1996, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/metadata/review.html>
  73. Heery, R., Carpenter, L., Day, M. Renardus Project Developments and the Wider Digital Library Context. *D-Lib Magazine*. 2001, vol.7, nr 4. Dostępny w World Wide Web:  
[www.dlib.org/dlib/april01/heery/04heery.html](http://www.dlib.org/dlib/april01/heery/04heery.html)
  74. Hickey, T. CORC: a system for gateway creation. *Online Information Review*. 2000, vol. 24, nr 1, s. 49-53

75. Hildreth, C.R. *Online Catalog Design Models: Are we Moving in the Right Direction?* [online]. 1995, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://myweb.cwpost.liu.edu/childret/clr-opac.html>
76. Hiom, D. SOSIG: an Internet hub for the social sciences, business and law. *Online Information Review* 2000, vol.24, nr 1, s. 54-58
77. Hiom, D. *Intute Collection Development Framework and Policy*. [online] 2006a, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.intute.ac.uk/Intute%20Collection%20Development%20Framework%20and%20Policy%20v1%204.doc>
78. Hiom, D. *Intute Cataloguing Guidelines*. [online] 2006b, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://www.intute.ac.uk/intute\\_cataloguing\\_guidelines\\_v4\\_1.doc](http://www.intute.ac.uk/intute_cataloguing_guidelines_v4_1.doc)
79. Hiom, D. Retrospective on the RDN. *Ariadne* 2006c, nr 47. Dostępny w World Wide Web: [www.ariadne.ac.uk/issue47/hiom](http://www.ariadne.ac.uk/issue47/hiom)
80. Hjørland, B. The Concept of "Subject" in Information Science. *Journal of Documentation*. 1992, vol. 48, nr 2, s. 172-200
81. Hjørland, B. Information Retrieval, Text Composition and Semantics. *Knowledge Organization*. 1998, vol. 25, nr 1/2, s. 16-31
82. Hjørland, B. Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of Documentation*. 2002, vol. 58, nr 4, s. 422-462
83. Hjørland, B. *Basic Units in Library and Information Science* [online]. 2004, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://web.archive.org/web/20051208011211/http://www.db.dk/bh/Units+in+IS\\_B.ppt](http://web.archive.org/web/20051208011211/http://www.db.dk/bh/Units+in+IS_B.ppt)
84. Hjørland, B. *Subject access points* [online]. 2007a, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://www.db.dk/bh/lifeboat\\_ko/CONCEPTS/subject\\_access\\_points.htm](http://www.db.dk/bh/lifeboat_ko/CONCEPTS/subject_access_points.htm)
85. Hjørland, B. *Units or entities in knowledge organization (KO). What is being organized?* [online]. 2007b, Dostępny w World Wide Web: [http://www.db.dk/bh/Lifeboat\\_KO/HISTORY%20&%20THEORY/units\\_in\\_knowledge\\_organization.htm](http://www.db.dk/bh/Lifeboat_KO/HISTORY%20&%20THEORY/units_in_knowledge_organization.htm)
86. Hjørland, B.; Kyllsbech Nielsen, L. Subject access points in electronic retrieval. *Annual Review of Information Science and Technology*. 2001, vol. 35, s. 249-299
87. Hodge, G. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington: Digital Library Federation, 2000. ISBN 1-887334-76-9
88. Hodgins, Wayne (red.). *Draft Standard for Learning Object Metadata* [online]. 2002, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)
89. Hofman, P.; Worsfold, E. *Selection Criteria for Quality Controlled Information Gateways. DESIRE deliverable D3.2*. [online] 1997. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/metadata/desire/quality/quality.rtf>
90. Hulme, E.W. *Principles of Book Classification*. Library Association Record, 1911
91. Huxley, L., Carpenter, L., Peereboom, M. The Renardus broker service: collaborative frameworks and tools. *The Electronic Library*. 2003, vol. 21, nr 1, s. 39-48
92. Huxley, L., Joyce, A. A social science gateway in a shifting digital world: shaping SOSIG for users' needs of the future. *Online Information Review*. 2004, nr 5, s. 328-337
93. *IMesh Toolkit: What is a subject gateway?* [online]. [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://scout.wisc.edu/Projects/PastProjects/IMeshToolkit/>
94. *Intute Cataloguing Guidelines*. [online] 2006, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [www.intute.ac.uk/intute\\_cataloguing\\_guidelines\\_v4\\_1.doc](http://www.intute.ac.uk/intute_cataloguing_guidelines_v4_1.doc)
95. *ISAAR (CPF). Międzynarodowy standard archiwalnych hasel wzorcowych stosowanych do archiwów ciał zbiorowych, osób i rodzin, ISAAR (CPF)*. 2006. Warszawa: Naczelna Dyrekcja Archiwów Państwowych. ISBN 83-89115-63-8
96. *ISBD(ER): International Standard Bibliographic Description for Electronic Resources*. London: IFLA Universal Bibliographic Control and International MARC Programme, 1997. ISBN 3-598-11369-2
97. Jachimczyk, A. *Webliografia - tematyczne zestawienia internetowych źródeł*. W: Zarządzanie informacją w nauce, red. D. Pietruch-Reizes, Katowice: Wydaw. Uniwersytetu Śląskiego, 2008, s. 158-169
98. Januszewska, M. Systemy hipertekstowe – rys historyczny. *Zagadnienia Informacji Naukowej*. 1992, nr 1(56), s. 85-100
99. Kirremuir, J. *Programme Background A brief history to, and background of, the eLib programme* [online] 1998. [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/services/elib/background/history.html>



100. Kirremuir, J. *A Brief Survey of Quality Resource Discovery Systems* [online]. 1999 [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://web.archive.org/web/20060911125110/http://www.rdn.ac.uk/publications/studies/survey/>
101. Knowlton, S. A Preview of the New Cataloging Code. *Michigan Library Association Forum*. 2006, vol. 6, nr 1. Dostępny w World Wide Web: <http://www.mlaforum.org/volumeV/issue1/article2.htm>
102. Koch, T. Spotlight on BUBL. *Ariadne*. 1996, nr 1. Dostępny w World Wide Web: [www.ariadne.ac.uk/issue1/bubl](http://www.ariadne.ac.uk/issue1/bubl)
103. Koch, T. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review*. 2000, vol.24, nr 1, s. 24-34
104. Koch, T. Report on the 5th European Networked Knowledge Organization Systems/Services (NKOS) Workshop. *D-Lib Magazine*. 2006, vol. 12, nr 10. Dostępny w World Wide Web: <http://www.dlib.org/dlib/october06/koch/10koch.html>
105. Koch, T.; Day, M. *The role of classification schemes in Internet resource description and discovery* [online]. 1997, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/metadata/desire/classification/>
106. Koch, T.; Golub, K.; Ado, A. Users Browsing Behaviour in DDC-based Web Services: A Log Analysis. *Cataloging & Classification Quarterly*. 2006, nr 3/4, s. 163-186
107. Koniger, P.; Janowitz, K. Drowning in Information, but Thirsty for Knowledge. *International Journal of Information Management*. 1995, vol. 15, nr 1, s. 5-16
108. Kowalski, G. *Information Retrieval Systems: Theory and Implementation*. Boston: Kluwer Academic Publishers, 1997. ISBN 0-7923-9926-9
109. Kwasnik, B. *A Descriptive Study of the Functional Components of Browsing*. W: *Engineering for Human-Computer Interaction*. Amsterdam: Elsevier Science Publishers, 1992, s. 191-202.
110. Kwasnik, B. The role of classification in knowledge representation and discovery. *Library Trends*. 1999, vol. 48, nr 1, s. 22-47.
111. Kyung-Sun, K.; Siei-Ching, J. S.; Soo-Jin, P. *Facet analysis of categories used in web directories: a comparative study*. W: *Libraries: Dynamic Engines for the Knowledge and Information Society: Proceedings of World Library and Information Congress: 72nd IFLA General Conference and Council*. 2006
112. Lancaster, F. W. Do Indexing and Abstracting Have a Future? *Annales de Documentacion*. 2003, nr 6, s. 137-144. Dostępny w World Wide Web: <http://eprints.rclis.org/2872/>
113. Larsen, B.; Ingwersen, P. *Cognitive overlaps along the polyrepresentation continuum*. W: *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, 2005 s. 43-60
114. Lassila, O.; Swick, R. R. *Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999* [online]. 1999, Dostępny w World Wide Web: <http://www.w3.org/TR/REC-rdf-syntax/>
115. Laundry, R. *GEM Browse Facets*. [online] 2006, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.thegateway.org/help/search-help/facetDescriptions/>
116. Laundry, R. *Information about the Gateway to Educational Materials (GEM) project including its mission, goals, organization and administration*. [online] 2006, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.thegateway.org/about/gemingeneral/about-gem/>
117. Leise, F.; Fast, K.; Steckel, M. Synonym Rings and Authority Files. *BoxesAndArrows*. 2003, nr 3. Dostępny w World Wide Web: [http://www.boxesandarrows.com/view/synonym\\_rings\\_and\\_authority\\_files](http://www.boxesandarrows.com/view/synonym_rings_and_authority_files)
118. Lesk, M. *Understanding Digital Libraries*. New York: Morgan Kaufmann, 2005. ISBN 1-55860-924-5
119. Logan, G. Cumulative progress in formal theories of attention. *Annual Review of Psychology*. 2003, nr 55, s. 207-234
120. Lyman, P.; Varian, H.R. *How much information 2003?* [online]. 2003, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>
121. Mai, J.-E. A postmodern theory of knowledge organization. W: ; *Proceedings of the 7th Annual International ISKO Conference*, red.M. J. Lopez-Huertas, Würzburg: Ergon-Verlag, 2002, s. 547-556
122. Marchionini, G. *Information Seeking In Electronic Environments*. New York: Cambridge University Press, 1995. ISBN 0521586747
123. Mauer, D; Warfel, T. Card sorting: a definitive guide. *BoxesAndArrows*. 2004, nr 4. Dostępny w World Wide Web: [http://www.boxesandarrows.com/view/card\\_sorting\\_a\\_definitive\\_guide](http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide)
124. McGuinness, D.L. *Ontologies Come of Age*. W: *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge: MIT Press, 2002 s. 171-196.

125. McKiernan, G. *Beyond bookmarks: schemes for organizing the Web* [online]. 1996, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.iastate.edu/~CYBERSTACKS/CTW.htm>
126. Miksa, F. *The DDC, the Universe of Knowledge, and the Post-Modern Library*. Albany, NY: Forest Press, 1998, ISBN 0-910608-64-4
127. Miller, P. *Web 2.0: Building the New Library*. *Ariadne*. 2005, vol. 45, nr 10. Dostępny w World Wide Web: <http://www.ariadne.ac.uk/issue45/miller/>
128. Monopoli, M.; Nicholas, D. *An Evaluation of the Use Subject Based Information Gateways: Case Study Adam*. W: *Proceedings 9th Panhellenic Conference of Academic libraries*. Athens: GR, 2000 s. 143-158
129. Nahotko, M. *Functional Requirements for Bibliographic Records (FRBR) - model opracowania zbiorów bibliotecznych IFLA*. *Bibliotekarz*. 2001, nr 1, s. 13-16
130. Nahotko, M. *Metadane. Sposób na uporządkowanie Internetu*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2004a. ISBN 83-233-1825-5
131. Nahotko, M. *Cyfrowa najmłodsza siostra bibliotek*. *Konspekt*. 2004b, lipiec-sierpień. Dostępny w World Wide Web: <http://www.wsp.krakow.pl/konspekt/19/nahotko.html>
132. Najjar, J.; Duval, E. *Actual Use of Learning Objects and Metadata*. *Bulletin of the IEEE Technical Committee on Digital Libraries* 2006, vol. 2, nr 2. Dostępny w World Wide Web: [www.ieee-tcdl.org/Bulletin/v2n2/najjar/najjar.html](http://www.ieee-tcdl.org/Bulletin/v2n2/najjar/najjar.html)
133. Neuroth, H.; Koch, T. *Cross-browsing and cross-searching in a distributed network of subject gateways: architecture, data model, and classification*. [online]. 2001, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.html>
134. Neveol, A.; Soulamia, L.; Magaly, D. *Using CISMef MeSH "encapsulated terminology and a categorization algorithm for health resources*. *International Journal of Medical Informatics*. 2004, nr 73, s. 57-64
135. Newman, Ch. *W3CDTF. Date and Time Formats* [online]. 1997, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.w3.org/TR/NOTE-datetime>
136. O'Connor, B. *Browsing: a framework for seeking functional information*. *Science Communication*. 1993, vol. 15, nr 2, s. 211-232
137. Olson, N. (red.). *Cataloging Internet Resources: A Manual and Practical Guide* [online]. 1997, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.oclc.org/support/documentation/worldcat/cataloging/internetguide/>
138. O'Reilly, T. *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software* [online]. 2005, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
139. Pacek, J. *W poszukiwaniu optymalnej jednostki opisu*. *Biuletyn EBIB*. 2007, vol. 86, nr 5. Dostępny w World Wide Web: <http://www.ebib.info/2007/86/a.php?pacek>
140. Pacek, J. *Bibliografia 2.0. Zagadnienia Informacji Naukowej*. 2008, nr 1, s. 35-44
141. Peereboom, M. *DutchESS Quality Criteria* [online]. 1999 [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://www.kb.nl/dutchess/manual/quality\\_eng.html](http://www.kb.nl/dutchess/manual/quality_eng.html)
142. Pejtersen, A.M. *New Model for Multimedia Interfaces to Online Public Access Catalogues*. *Electronic Library*. 1992, vol. 10, nr 6, s. 359-366
143. *PN-82/N-01152.01. Opis bibliograficzny. Książki*. Warszawa: PKN, 1982
144. Pepeł, A. *Zastosowanie UKD w dużych bazach danych oraz serwisach internetowych*. *Bibliotekarz*. 1999, nr 5, s. 6-9
145. Pomianowicz, J. *Baztol jako przykład serwisu typu subject gateway o kontrolowanej jakości*. W: *Informacja dla nauki a świat zasobów cyfrowych. Materiały z konferencji naukowej, Świnoujście, 11-13 września 2008 r.* (nieopublikowane)
146. Powell, A. *Expressing Dublin Core in HTML/XHTML meta and link elements* [online] 2003, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://dublincore.org/documents/dcq-html/>
147. Powell, A.; Heaney, M.; Dempsey, L. *RSLP Collection Description*. *D-Lib Magazine*. 2000, vol. 6, nr 9. Dostępny w World Wide Web: [www.dlib.org/dlib/september00/powell/09powell.html](http://www.dlib.org/dlib/september00/powell/09powell.html)
148. Próchnicka, M. *Człowiek i komputer: dialogowy model wyszukiwania informacji*. Kraków: Wydaw. UJ, 2004. ISBN 83-233-1874-3
149. Qin, J. *Representation and Organization of Information in the Web Space: From MARC to XML*. *Informing Science*. 2000, vol. 3, nr 2. Dostępny w World Wide Web: <http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>
150. Qin, J.; Paling, S. *Converting a controlled vocabulary into an ontology: the case of GEM*. *Information Research*. 2001, vol. 6, nr 2. Dostępny w World Wide Web: <http://informationr.net/ir/6-2/paper94.html>
151. Rada, R. *Hypertext. From Text to Expertext*. London: McGraw-Hill, 1991. ISBN 0077074017

152. Radwański, A. Biblioteka wirtualna - problemy definicyjne. *Biuletyn EBIB*. 2002, vol. 8. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/arc/e008-02.html>
153. Rajashekar, T.B. *IS 206 - Information and Knowledge Organization* [online]. 2003, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ncsi.iisc.ernet.in/raja/is206/index.htm>
154. Ratajowski, J. *Wybrane problemy metodologiczne informologii nauki*. Katowice: Wydaw. Uniwersytetu Śląskiego, 1994. ISBN 83-226-0546-3
155. RENARDUS Participating Gateways [online]. 2001, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://renardus.sub.uni-goettingen.de/info.html>
156. Rice, R.; McCreddie, M.; Chang, S.L. *Accessing and browsing information and communication*. Cambridge: MIT Press, 2001. ISBN 978-0-262-18214-0
157. Roberts, S.L. Practical taxonomies: hard-won wisdom for creating a workable knowledge classification system. *Knowledge Management*. 1999, January, s. 45-47
158. Rosenfeld, L.; Morville, P. *Architektura informacji w serwisach internetowych*. Gliwice: Helion, 2003. ISBN 83-7361-003-0
159. Roszkowski, M. Architektura informacji w serwisach hipertekstowych. *Zagadnienia Informacji Naukowej*. 2004, nr 2, s. 13-29
160. Roszkowski, M. Źródła słownictwa kontrolowanego dla dziedzinowych systemów hipertekstowych w Internecie. *Zagadnienia Informacji Naukowej*. 2005, nr 2, s. 31-44
161. Roszkowski, M. Dziedzinowe systemy hipertekstowe jako źródło informacji dla nauczycieli bibliotekarzy. *Warsztaty bibliotekarskie*. 2007a, vol. 21, nr 1. Dostępny w World Wide Web: <http://www.pedagogiczna.edu.pl/warsztat/2007/1/070105.htm>
162. Roszkowski, M. Folksonomie jako narzędzia społecznego tagowania. *Warsztaty bibliotekarskie*. 2007b, vol. 24, nr 4. Dostępny w World Wide Web: <http://www.pedagogiczna.edu.pl/warsztat/2007/4/070404.htm>
163. Rowley, J.; Farrow, J. *Organizing Knowledge: An Introduction to Managing Access to Information*. Hampshire: Ashgate, 2002. ISBN 0-566-08047-8
164. Rowley, J. Knowledge organization for a new millennium: Principles and practices. *Journal of Knowledge Management*. 2000, vol. 4, nr 3, s. 217-223
165. Sadowska, J.; Turowska, T. *Języki informacyjno-wyszukiwawcze. Katalogi rzeczowe*. Warszawa: Centrum Ustawicznego Kształcenia Bibliotekarzy. 1990. ISBN 83-00-03315-7
166. Saeed, H.; Chaudhry, A.S. Using Dewey Decimal Classification Scheme for building taxonomies for knowledge organization. *Journal of Documentation*. 2002, vol. 58, nr 5, s. 575-583
167. Salton, G. *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968
168. Shirky, C. *Ontology is Overrated: Categories, Links, and Tags* [online]. 2005, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [http://www.shirky.com/writings/ontology\\_overrated.html](http://www.shirky.com/writings/ontology_overrated.html)
169. Sicilia, M.A.; Garcia, E.; Aedo, I.; Paloma, D. A literature-based approach to annotation and browsing web resources. *Information Research*. 2003, vol. 8, nr 2. Dostępny w World Wide Web: <http://informationr.net/ir/8-2/paper149.html>
170. Skórka S. Architektura informacji. Nowy kierunek rozwoju informacji naukowej. *Biuletyn EBIB*. 2002, nr 11/40 [online] <http://ebib.oss.wroc.pl/2002/40/skorka.php> 02-12-2002
171. Skórka, S. *Użytkownicy systemów hipertekstowych. Strategie poszukiwania informacji w edukacyjnym serwisie internetowym*. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej w Krakowie, 2006. ISBN 83-7271-384-7
172. Slavic, A. UDC in subject gateways: experiment or opportunity? *Knowledge Organization*. 2006, vol. 33, nr 2, s. 67-85
173. *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych*. B. Bojar (oprac.) Warszawa: Wydaw. SBP, 2002. ISBN 83-87629-84-7
174. Smiraglia, R. *The nature of "a work". Implications for the Organization of Knowledge*. Lanham, Md.: Scarecrow Press, 2001. ISBN 0-8108-4037-5
175. Smiraglia, R. The progress of theory in knowledge organization. *Library Trends*. 2002, nr 12. Dostępny w World Wide Web: [http://findarticles.com/p/articles/mi\\_m1387/is\\_3\\_50/ai\\_88582618](http://findarticles.com/p/articles/mi_m1387/is_3_50/ai_88582618)
176. Smith, T. *Measuring the Efficacy of a Folksonomy for Subject Analysis*. W: Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research. Milwaukee, Wisconsin: 2007. Dostępny w World Wide Web: <http://dlist.sir.arizona.edu/2061/>
177. Sosińska, B. Funkcje języków informacyjno-wyszukiwawczych. *Zagadnienia Informacji Naukowej*. 1987, nr 1, s. 37-52
178. Sosińska-Kalata, B. *Modele organizacji wiedzy w systemach wyszukiwania informacji o dokumentach*. Warszawa: Wydaw. SBP, 1999, ISBN 83-87629-22-7
179. Sosińska-Kalata, B. *Klasyfikacja: struktury organizacji wiedzy, piśmiennictwa i zasobów informacyjnych*. Warszawa: Wydaw. SBP, 2002a. ISBN 83-87629-87-1

180. Sosińska-Kalata, B. *Struktury klasyfikacyjne w organizacji zasobów informacyjnych Internetu*. W: *Multimedialne i Sieciowe Systemy Informacyjne*. Wrocław: Zakład Systemów Informacyjnych. Politechnika Wroclawska, 2002b, s. 403-412
181. Sosińska-Kalata, B. *Języki informacyjno-wyszukiawcze. Współczesne tendencje w badaniu i zastosowaniu*. W: *Społeczeństwo informacyjne i jego technologie*. Red. B. Sosińska-Kalata. Warszawa: Wydaw. SBP, 2004, s. 79-91
182. Sosińska-Kalata, B. *Systemy organizacji wiedzy w środowisku sieciowym*. W: *Od informacji naukowej do technologii społeczeństwa informacyjnego*. Red. B. Sosińska-Kalata, M. Przystek-Samokowa. Warszawa: Wydaw. SBP, 2005a, s. 141-162
183. Sosińska-Kalata, B. *Tezaurusy w zmieniającym się środowisku sieciowym*. W: *Informacja w sieci : problemy, metody, technologie*. Red. B. Sosińska-Kalata. Warszawa: Wydaw. SBP, 2005b, s. 140-156
184. Sparck-Jones, K. *Retrieving Information or Answering Questions*. London: The British Library Board, 1990. ISBN 0712332154
185. Sperberg-McQueen, C. M. ; Burnard, L. *A gentle introduction to SGML. TEI Guidelines for Electronic Text Encoding and Interchange* [online]. 2001, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.isgmlug.org/sgmlhelp/g-index.htm>
186. Spink, A. (red.); Cole, Ch. (red.). *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, 2005. ISBN 978-1-4020-4014-6
187. Spiteri, L. Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*. 2007, vol. 4, nr 2. Dostępny w World Wide Web: <http://www.webology.ir/2007/v4n2/a41.html>
188. *SIG-FI Project Documentary*. [online] 1999, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.sub.uni-goettingen.de/ssgfi/projekt/doku/index.html>
189. Sutton, S. Conceptual Design and Deployment of a Metadata Framework for Educational Resources on the Internet. *Journal of the American Society for Information Science*. 1999, vol. 50, nr 13, s. 1182-1192
190. Sutton, S. Gateway to Educational Materials (GEM): Metadata for Networked Information Discovery and Retrieval. *Computer Networks and ISDN Systems*. 1998, nr 30, s. 691-693
191. Sutton, S. GEM: Using metadata to enhance Internet retrieval by K-12 teachers. *Bulletin of the American Society for Information Science*. 1997, s. 21-22
192. Svenonius, E. *The Intellectual foundation of information organization*. Cambridge: MIT Press Massachusetts Institute of Technology, 2000. ISBN 0-262-19433-3
193. Ścibor, E. *Wybrane zagadnienia teorii języków informacyjno-wyszukiawczych*. Olsztyn: Wyższa Szkoła Pedagogiczna, 1998. ISBN 83-87315-58-3
194. Ścibor, E.; Bielicka, L. *Wprowadzenie do teorii języków informacyjnych*. Warszawa: CİNTE, 1981
195. Taylor, A.G. *The Organization of Information*. Englewood, Colorado: Libraries Unlimited, Inc., 1999. ISBN 1-56308-493-7
196. The Projects [eLib Programme] [online], [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/services/elib/projects/>
197. Thirion, B., Darmoni, S. Simplified access to MeSH Tree Structures on CISMef. *Bulletin of the Medical Library Association*. 1999, vol. 87, nr 4, s. 480-481
198. Tomaszczyk, J. Taksonomia jako narzędzie organizacji informacji. *Zagadnienia Informacji Naukowej*. 2007, nr 1, s. 40-49
199. Tudhope, D. Report on the 4th European Networked Knowledge Organization Systems/Services (NKOS) Workshop. Mapping Knowledge Organization Systems. *D-Lib Magazine*. 2005, vol. 11, nr 11. Dostępny w World Wide Web: <http://www.dlib.org/dlib/november05/tudhope/11tudhope.html>
200. Tudhope, D. *Networked Knowledge Organization Systems and Services. The 6th European Networked Knowledge Organization Systems (NKOS) Workshop* [online]. 2007, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007/>
201. Tudhope, D.; Koch, T. New Applications of Knowledge Organization Systems: introduction to a special issue. *Journal of Digital Information*. 2003, vol. 4, nr 4. Dostępny w World Wide Web: <http://jodi.tamu.edu/Articles/v04/i04/editorial/>
202. Tudhope, D.; Koch, T.; Heery, R. *Terminology Services and Technology: JISC state of the art review* [online]. 2006, [dostęp: 2009-01-22]. Dostępny w World Wide Web: [www.ukoln.ac.uk/terminology/JISC-review2006.html](http://www.ukoln.ac.uk/terminology/JISC-review2006.html)
203. Ungurian, O. *Elementy teorii języków informacyjno-wyszukiawczych*. Warszawa: OIN PAN, 1976
204. Vickery, B. Ontologies. *Journal of Information Science*. 1997, nr 23, s. 277-286

205. Vickery, B. *On Knowledge Organisation* [online]. 2008, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.lucis.me.uk/knowlorg.htm>
206. Vizine-Goetz, D.; Beall, J. Using literary warrant to define a version of the DDC for automated classification services. *Advances in Knowledge Organization*. 2004, nr 9. Dostępny w World Wide Web: <http://www.oclc.org/research/publications/archive/2004/vizine-goetz-beall.pdf>
207. Vizine-Goetz, D. Classification schemes for internet resources revisited. *Journal of Internet Cataloging*. 2002, vol. 5, nr 4, s. 5-18
208. Wereszczyńska-Cisło, B. Charakterystyka i możliwości eksploatacji hipertekstu. *Zagadnienia Informacji Naukowej*. 1997, nr 2, s. 17-30
209. Wheatley, A. Subject Trees on the Internet: A New Role for Bibliographic Classification? *Journal Of Internet Cataloging*. 2000, vol. 2, nr 3/4, s. 115-141
210. Whitelaw, A.; Joy, G. *Summative evaluation of phases 1 and 2 of the elib initiative: final report* [online]. 2000, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/elib-fr-v1-2.pdf>
211. Will, L. *Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval* [online]. 2006, [dostęp: 2009-01-22]. Dostępny w World Wide Web: <http://www.willpowerinfo.co.uk/glossary.htm>
212. Wise, R. The 3 Click Rule. *WebSiteOwner*. 2001, 15th December. Dostępny w World Wide Web: <http://websiteowner.info/articles/design/3clickrule.asp>
213. Wolfe, J. Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin & Review*. 1994, vol. 1, nr 2, s. 202-238
214. Worsfold, E. Subject gateways: fulfilling the DESIRE for knowledge. *Computer Networks and ISDN Systems*. 1998, nr 30, s. 1479-1489
215. Woźniak, J. *Kategoryzacja - studium z teorii języków informacyjno-wyszukiwawczych*. Warszawa: Wydaw. SBP, 2000. ISBN 83-87629-32-4
216. Woźniak-Kasperek, J. Internet - spojrzenie z perspektywy organizacji informacji. *Zagadnienia Informacji Naukowej*. 2004, vol. 83, nr 1, s. 29-45
217. Woźniak-Kasperek, J. *Organizacja informacji w internetowych serwisach kontrolowanej jakości*. W: *Opracowanie przedmiotowe dokumentów z zakresu nauk ścisłych: matematyczno-przyrodniczych i technicznych. Język haseł przedmiotowych KABA: teoria, praktyka, przyszłość*. Kazimierz Dolny, 20-22 września 2006 roku. Warszawa: Stowarzyszenie Bibliotekarzy Polskich, 2006. Dostępny w World Wide Web: <http://www.ebib.info/publikacje/matkonf/kaba/wozniak-kasperek.php>
218. Zhonghong, W, Chaudhry, A.S.; Khoo, C. Potential and Prospects of Taxonomies for Content Organization. *Knowledge Organization*. 2006, vol. 33, nr 3, s. 160-169
219. Zins, C. Models for Classifying Internet Resources. *Knowledge Organization*. 2002, vol. 29, nr 1, s. 20-28

## Wykaz tabel

Tab. 1.1. Taksonomia przedmiotowych punktów dostępu.....	34
Tab. 2.1. Zestaw podstawowych elementów Dublin Core.....	47
Tab. 2.2. Porównanie wyszukiwania i przeglądania.....	68
Tab. 2.3. Potencjalne źródła wartości dla cech formalnych.....	74
Tab. 2.4. Wyniki badań D. Vizine-Goetz.....	89
Tab. 2.5. Porównanie modelu Zinsa i jego zmodyfikowanej wersji.....	92
Tab. 3.1. Charakterystyka grup użytkowników docelowych w wybranych dziedzinowych systemach hipertekstowych.....	108
Tab. 3.2. Rodzaj zasobów w systemach SOSIG i OMNI.....	109
Tab. 3.3. Struktura rekordu IAFA/ROADS dla źródeł internetowych.....	113
Tab. 3.4. Zestaw metadanych w systemie Intute.....	115
Tab. 3.5. Schemat metadanych dla systemów sieci SSG-FI.....	116
Tab. 3.6. Zestaw metadanych dla systemu Edna.....	120
Tab. 3.7. Schemat metadanych dla sieci Renardus.....	122
Tab. 3.8. Typologia dziedzinowych systemów hipertekstowych w postaci macierzy.....	124
Tab. 4.1. Typ zasobów dla systemów sieci RDN.....	141
Tab. 4.2. Udział JIW w reprezentacji wiedzy w DSH.....	145
Tab. 4.3. Punkty dostępu w wybranych DSH.....	160
Tab. 5.1. Rozkład kategorii tematycznych w strukturze hierarchicznej BUBL.....	187
Tab. 5.2. Organizacja kategorii tematycznej 020 Bibliotekoznawstwo i informacja naukowa w BUBL.....	187
Tab. 5.3. Różnice w budowie nazwy kategorii BUBL i symbolu KDD.....	189
Tab. 5.4. Dystrybucja odesłań do ChWD w strukturze hierarchicznej BUBL.....	191
Tab. 5.5. Średnia pojemność kategorii tematycznej w BUBL.....	191
Tab. 5.6. Wartości dla cechy typ źródła w katalogu BUBL.....	193
Tab. 5.7. Wartość cechy typ zasobów w GEM.....	202
Tab. 5.8. Organizacja kategorii tematycznej <i>Chemia</i> w systemie Intute.....	211
Tab. 5.9. Zagnieżdżanie symbolu KDD w strukturze odesłania hipertekstowego.....	212
Tab. 5.10. JIW jako narzędzie organizacji informacji w grupie tematycznej Intute: nauka, inżynieria i technologia.....	212
Tab. 5.11. Rozkład kategorii tematycznych w grupie Intute: nauka, inżynieria i technika.....	214
Tab. 5.12. Ilościowy rozkład odesłań na poszczególnych poziomach w grupie Intute: nauka, inżynieria i technologia.....	214
Tab. 5.13. Średnia pojemność klas w Intute: nauka, inżynieria i technologia.....	215
Tab. 5.14. Rozkład kategorii w grupie Intute: nauki społeczne.....	217
Tab. 5.15. Dystrybucja ChWD na poszczególnych poziomach w układzie hierarchicznym w grupie Intute: nauki społeczne.....	219
Tab. 5.16. Średnia pojemność kategorii na poszczególnych stopniach podziału.....	219
Tab. 5.17. Ilościowy rozkład kategorii tematycznych w klasie głównej Historia Naturalna.....	226
Tab. 5.18. Ilościowy rozkład odesłań do ChWD w klasie głównej Historia Naturalna.....	227
Tab. 5.19. Rozkład kategorii w klasie głównej Historia medycyny.....	230
Tab. 5.20. Rozkład odesłań w klasie głównej Historia medycyny.....	230
Tab. 5.21. Liczba klas wyodrębnionych na pierwszym stopniu podziału.....	231
Tab. 5.22. Głębokość struktur hierarchicznych w Intute.....	231
Tab. 5.23. Liczba wyodrębnionych kategorii w poszczególnych grupach Intute.....	232
Tab. 5.24. Organizacja kategorii tematycznych w MathGuide z wykorzystaniem Mathematics Subject Classification.....	234
Tab. 6.1. Metodyka identyfikacji miejsca kategorii w schemacie.....	256
Tab. 6.2. Głębokość struktur hierarchicznych.....	258
Tab. 6.3. Liczba kategorii w strukturze.....	259
Tab. 6.4. Liczba kategorii na pierwszym stopniu podziału w narzędziach.....	260
Tab. 6.5. Dystrybucja kategorii w strukturze.....	261
Tab. 6.6. Dystrybucja ChWD w strukturze.....	262
Tab. 6.7. Pojemność kategorii na poszczególnych stopniach w strukturze.....	263
Tab. 6.8. Liczba punktów dostępu do jednej ChWD dla wybranych systemów.....	263
Tab. 6.9. Przynależność kategorialna punktów dostępu na pierwszym stopniu podziału.....	268
Tab. 6.10. Porównanie rodzajów kategorii na pierwszym stopniu podziału.....	269

## Spis ilustracji

Rys. 1.1. Proces organizacji informacji.....	17
Rys. 1.2. Rodowód procesu informacyjnego.....	20
Rys. 2.1. Struktura hipertekstu.....	43
Rys. 2.2. Nawigacja globalna w systemie Intute.....	45
Rys. 2.3. Punkty dostępu w interfejsie wyszukiwarki Google.....	54
Rys. 2.4. Struktura katalogu internetowego dmoz.org.....	56
Rys. 2.5. Chmura tagów w serwisie CiteULike.....	61
Rys. 2.6. Interfejs serwisu KO-SWICKI.....	63
Rys. 2.7. Technika ewoluującego zapytania informacyjnego - berrypicking M. Bates.....	69
Rys. 2.8. Modelowanie przykładowej jednostki opisu.....	73
Rys. 2.9. Dostęp do listy wartości dla cechy kategoria tematyczna w NIST Virtual Library.....	76
Rys. 2.10. Dostęp do słownika JIW poprzez mechanizm wyszukiwawczy w systemie Virtual Health Library.....	76
Rys. 2.11. Alfabetyczny wykaz wartości punktów dostępu dla cechy <i>rodzaj źródła</i> w Digital Library of Earth System Education.....	79
Rys. 2.12. Wykaz słów kluczowych w narzędziu dostępu przedmiotowego w Śląskiej Bibliotece Cyfrowej.....	80
Rys. 2.13. Tworzenie taksonomii na podstawie KDD.....	88
Rys. 3.1. Model analityczny dziedzinowego systemu hipertekstowego.....	127
Rys. 4.1. Użycie Klasyfikacji Dziesiątej Deweya do odwzorowania treści w systemie BUBL.....	149
Rys. 4.2. Użycie Göttinger Online Klassifikation do odwzorowania treści w systemie GeoGuide.....	149
Rys. 4.3. Łączenie symboli klasyfikacyjnych z ich odpowiednikami słownymi w systemie ViFaPhys..	150
Rys. 4.4. Technika łączenia symboli klasyfikacyjnych z odpowiednikami słownymi w systemie Slavistik-Portal.....	151
Rys. 4.5. Polireprezentacja wiedzy w systemie MathGuide.....	153
Rys. 4.6. Polireprezentacja wiedzy w systemie EdnaOnline.....	154
Rys. 4.7. Ukryta polireprezentacja wiedzy w grupie tematycznej Social Science systemu Intute.....	155
Rys. 4.8. Identyfikacja cech wyszukiwawczych w dziedzinowych systemach hipertekstowych.....	159
Rys. 4.9. Tryb prosty modułu wyszukiwawczego w systemie Aerade.....	162
Rys. 4.10. Tryb prosty modułu wyszukiwawczego w systemie MathGuide.....	162
Rys. 4.11. Tryb zaawansowany modułu wyszukiwawczego w systemie MathGuide.....	164
Rys. 4.12. Zawężanie zapytania informacyjnego w systemie ForestryGuide.....	166
Rys. 4.13. Wyszukiwanie informacji wspomagane tezaurusami w Intute: Social Sciences.....	167
Rys. 4.14. Artykuł hasłowy w tezaurusie HASSET.....	168
Rys. 4.15. Rezultaty wyszukiwania z wykorzystaniem tezaurusa.....	169
Rys. 4.16. Ahierarchiczne narzędzia dostępu przedmiotowego wykorzystujące słowa kluczowe do reprezentacji treści dokumentów.....	171
Rys. 4.17. Ahierarchiczne narzędzia dostępu przedmiotowego w systemie AERADE.....	172
Rys. 4.18. Narzędzie dostępu przedmiotowego w postaci struktury drzewiastej.....	173
Rys. 4.19. Narzędzie dostępu przedmiotowego w postaci struktury drzewiastej w systemie BiziGate....	173
Rys. 4.20. Tryb prosty wyświetlania rezultatów wyszukiwania w systemie Aerade.....	175
Rys. 4.21. Tryb prosty wyświetlania rezultatów wyszukiwania w systemie MathGuide.....	176
Rys. 4.22. Tryb prosty wyświetlania rezultatów wyszukiwania w systemie Edna Online.....	177
Rys. 4.23. Tryb zaawansowany wyświetlania rezultatów wyszukiwania w systemie Intute.....	178
Rys. 4.24. Tryb zaawansowany wyświetlania rezultatów wyszukiwania w systemie MathGuide.....	179
Rys. 4.25. Relacje intertekstualne w systemie Intute.....	180
Rys. 5.1. Narzędzia dostępu do zasobów systemu BUBL.....	186
Rys. 5.2. Katalog przedmiotowy w BUBL.....	192
Rys. 5.3. Powiązanie katalogu przedmiotowego ze schematem hierarchicznym w BUBL.....	193
Rys. 5.4. Struktura kategorii punktów dostępu w systemie CISMef.....	196
Rys. 5.5. Punkty dostępu do zbioru informacyjnego w CISMef.....	197
Rys. 5.6. Artykuł hasłowy w CISMef.....	199
Rys. 5.7. Organizacja punktów dostępu w GEM.....	201
Rys. 5.8. Narzędzie dostępu przedmiotowego w systemie GEM.....	205
Rys. 5.9. Hybrydowe narzędzie dostępu przedmiotowego w systemie GEM.....	206
Rys. 5.10. Artykuł hasłowy w grupie tematycznej Intute: nauka, inżynieria i technologia.....	213
Rys. 5.11. Specyfikacja relacji semantycznych w grupie tematycznej Intute: nauki społeczne.....	218
Rys. 5.12. Podkolekcje w grupie tematycznej Intute: zdrowie i nauki o życiu.....	221
Rys. 5.13. Relacje paradygmatyczne w narzędziu dostępu przedmiotowego.....	224

Rys. 5.14. Dostęp do informacji w kategorii <i>Kalendarium wydarzeń</i> .....	229
Rys. 5.15. Struktura rekordu w systemie MathGuide.....	236
Rys. 6.1. Dostęp do informacji z wykorzystaniem struktury płaskiej w HealthWeb .....	242
Rys. 6.2. Dostęp do informacji z wykorzystaniem struktury płaskiej w systemie Intute: weterynaria. ...	242
Rys. 6.3. Tematy JHP jako autosyntaktyczne jednostki organizujące w systemie Intute: zdrowie i nauki o życiu – Bioetyka.....	243
Rys. 6.4. Organizacja informacji w systemie Infomine z wykorzystaniem języka haseł przedmiotowych .....	244
Rys. 6.5. Wykorzystanie tezauryza Royal College of Nursing Thesaurus w systemie Intute: zdrowie i nauki o życiu – Pielęgniarstwo.....	246
Rys. 6.6. Library of Congress Name Authority File jako źródło słownictwa w systemie MedHist.....	248
Rys. 6.7. Indeks tytułów w systemie Infomine .....	250
Rys. 6.8. Dostęp do zbioru w systemie MathGuide z wykorzystaniem indeksu rodzaj zasobu .....	251
Rys. 6.9. Indeks alfabetyczny rodzaju zasobu w systemie CISMef.....	252
Rys. 6.10. Wskaźnik kategorii leksykalnej <i>rodzaj zasobu</i> w systemie Intute: zdrowie i nauki o życiu – Medycyna.....	254
Rys. 6.11. Logiczny porządek kategorii na pierwszym stopniu podziału w systemie Aerade .....	265
Rys. 6.12. Alfabetyczny porządek kategorii na pierwszym stopniu podziału w systemie Psi-com .....	266
Rys. 6.13. Hierarchiczne narzędzie dostępu w systemie ForestryGuide .....	267
Rys. 6.14. <i>Literary warrant</i> jako źródło niespójności struktur hierarchicznych .....	271
Rys. 6.15. <i>User warrant</i> w dostępie przedmiotowym systemu Artifact.....	272
Rys. 6.16. <i>User warrant</i> w dostępie przedmiotowym systemu Intute: nauki społeczne .....	272
Rys. 6.17. Quasifasetyzacja w Intute: sztuka i humanistyka .....	274
Rys. 6.18. Wskaźnik miejsca w systemie Math Forum.....	275
Rys. 6.19. Rozbudowa kategorii tematycznej 610 Nauki medyczne w systemie BUBL .....	278



## Indeks autorski

- Abramowicz, Witold, 38  
Adams, Marc, 40  
Anderson, Jack, 18, 19, 86  
Antelman, Kristin, 32  
Artowicz, Elżbieta, 104  
Babik, Wiesław, 29, 30, 71  
Baeza-Yates, Ricardo, 52  
Barthes, Roland, 41  
Bates, Marcia J., 32, 65, 66, 67, 68, 69, 171  
Bawden, David, 96  
Beall, Julianne, 185, 235  
Bechler, Martin, 81  
Becker, Hans, 122  
Belcher, M., 98, 140, 142  
Bliss, Henry Evelyn, 18  
Błaszczyk, Hieronim, 132  
Bojar, Bożenna, 28, 29, 105, 106  
Borgman, Christine, 31  
Broughton, Vanda, 19, 22  
Brümmer, Anna, 96, 98, 99  
Burnard, Lou, 49  
Bush, Vannevar, 38  
Campbell, Debbie, 98  
Capurro, Rafael, 21  
Carr, Leslie, 38  
Champan, Ann, 208  
Chang, Shaun-Ju, 65  
Charles, Françoise, 195, 253  
Chaudhry, Abdus Sattar, 86, 87, 88  
Chowdhury, Gobinda, 16, 57, 96  
Chu, Heting, 47, 52, 65, 66, 67, 68  
Cliff, Pete, 142  
Cole, Charles, 21  
Conway, Susan, 87  
Cutter, Charles Ammi, 31  
Daconta, Michael, 86  
Darmoni, Stefan, 195  
Dawson, Alan, 185, 192  
Day, Michael, 81, 82, 84  
Dempsey, Lorcan, 21, 113  
Denton, William, 92  
Derfert-Wolf, Lidia, 10, 103  
Douyere, Magaly, 194, 195  
Drabenstott, Karen, 32  
Duval, Erik, 47, 119  
Eilbeck, Chris, 235  
Ellis, David, 36, 55, 56  
Engelbart, Douglas, 42  
Farrow, John, 28  
Fast, Karl, 77  
Fischer, Thomas, 116, 233  
Follet, Brian, 129  
Frohmann, Bernd, 21  
Ganińska, Halina, 11, 133  
Garshol, Lars Maria, 19  
Gliński, Wiesław, 38, 95  
Gmiterek, Grzegorz, 133  
Gold, Joanne, 185  
Heery, Rachel, 77, 100, 113, 122  
Hildreth, Charles, 32, 69  
Hiom, Debra, 207, 208, 215, 253  
Hjørland, Birger, 12, 18, 19, 20, 21, 33, 34  
Hodgins, Wayne, 72  
Hofman, Paul, 98, 110, 111, 185, 208  
Hulme, Wyndham E., 270  
Huxley, Lesly, 122, 216  
Ingwersen, Peter, 153  
Jachimczyk, Adam, 59  
Janowitz, Karl, 37  
Januszewska, Maria, 38  
Joyce, Angela, 216  
Khoo, Christopher, 86, 87  
Kirremuir, John, 100, 101, 107, 128  
Knowlton, Steven, 208  
Koch, Traugott, 77, 81, 82, 84, 93, 96, 99, 100, 102, 106, 107, 124, 125, 185, 216  
Koniger, Paul, 37  
Kotula, Sebastian, 133  
Kowalski, Gerard, 66  
Kwasnik, Barbara, 260, 269  
Kylllesbech Nielsen, Lykke, 12, 33, 34  
Kyung-Sun, Kim, 90, 91, 92  
Lancaster, Frederick Wilfred, 7, 8, 37, 57  
Larsen, Birger, 153  
Lassila, Ora, 49  
Laundry, Ryan, 200, 203  
Leise, Fred, 77  
Lesk, Michael, 56  
Lickliger, Joseph Carl Robnett., 56  
Logan, Gordon, 65  
Lyman, Peter, 36  
Lynema, Emily, 32  
Mai, Jens-Erik, 290  
Marchionini, Gary, 66, 67  
McCreadie, Maureen, 65  
McKiernan, Gerry, 84  
Miksa, Francis, 21, 290  
Miller, Paul, 60  
Monopoli, Maria, 96  
Morville, Peter, 44, 46  
Nahotko, Marek, 25, 47, 48, 49, 57, 113  
Najjar, Jehad, 119  
Neuroth, Heike, 100, 116, 233  
Neveol, Aurelie, 194, 195  
Newman, Chris, 144  
O'Reilly, Tim, 59  
Orbst, Leo, 86  
Pace, Andrew, 32  
Pacek, Jarosław, 25, 133  
Paling, Stephen, 206  
Pejtersen, Annelise Mark, 32  
Place, Emma, 81  
Powell, Andy, 123  
Próchnicka, Maria, 52  
Qin, Jian, 51, 95, 206  
Rada, Roy, 42  
Radwański, Aleksander, 56  
Rajashekar, Tarikere Basappa, 17  
Ribeiro-Neto, Berthier, 52  
Rice, Ronald, 65, 67  
Rosenfeld, Louis, 46  
Roszkowski, Marcin, 45, 46, 59, 100  
Rowley, Jennifer, 8, 18, 28, 297  
Russel, Rosemary, 21  
Sadowska, Jadwiga, 147  
Saeed, Hamid, 86, 87, 88  
Salton, Gerard, 18  
Shirky, Clay, 95

Simpson, Jan, 185, 192  
Skórka, Stanisław, 42, 44  
Slavic, Aida, 216, 222  
Sligar, Char, 87  
Smiraglia, Richard, 21  
Smith, Tiffany, 59, 86  
Sosińska-Kalata, Barbara, 11, 18, 21, 22, 23, 37, 42,  
83, 85, 86, 94, 95, 103, 124, 240, 270  
Soulamia, Lina, 194, 195  
Sparck-Jones, Karen, 153  
Sperberg-McQueen, Michael, 49  
Spink, Amanda, 21  
Spitieri, Louise, 62  
Steckel, Michael, 77  
Sutton, Stuart, 201  
Svenonius, Elaine, 8, 15  
Swick, Ralph, 49  
Ścibor, Eugeniusz, 29, 30  
Taylor, Arlene, 16, 21, 26  
Thirion, Benoit, 195  
Tomaszczyk, Jacek, 77, 88  
Tudhope, Douglas, 77, 93  
Turowska, Teresa, 147  
Ungurian, Olgierd, 33, 196  
Varian, Hal, 36  
Vasconcelos, Ana, 36, 55, 56  
Vickery, Brian C., 95, 289  
Vizine-Goetz, Diane, 88, 89, 137, 185, 235  
Wereszczyńska-Cisło, Barbara, 38, 39, 40, 41  
Wheatley, Alan, 88, 89, 90, 96, 137  
Whitelaw, Alan, 129  
Will, Leonard, 65  
Wise, Rosemarie, 262  
Wolfe, Jeremy, 65  
Worsfold, Emma, 96, 98, 110, 111, 185, 208  
Woźniak, Jadwiga, 33, 196  
Woźniak-Kasperek, Jadwiga, 94  
Zhonghong, Wang, 86, 87  
Zins, Chaim, 90, 91

# Indeks rzeczowy

- AACR, 26, 112, 115, 140, 141
- Access to Network Resources, 128
- ADAM, 125, 141
- Aerade, 161
- AERADE, 147, 176, 265
- AgriFor, 108
- AgriGate, 125
- Alexandria Digital Library, 143
- algebra Boole'a, 32, 52, 53, 75, 164
- Altis, 148
- architektura informacji, 16, 44
  - systemy nawigacyjne, 45, 46
  - systemy organizacyjne, 44
  - systemy oznaczania, 44
  - systemy wyszukiwawcze, 46
- Artifact, 146
- Arts and Architecture Thesaurus, 148
- Basisklassifikation, 152
- BazTOL, 132
- berrypicking, technika, 68, 69
- Bibliografia 2.0, 134
- biblioteki cyfrowe, 56, 58
- BIOME, 142, 207, 220, 225
- BioethicsWeb, 243
- Biz/Ed, 125
- BiziGate, 174
- British Library Name Authority File, 140
- BUBL, 125, 149, 184, 185, 187, 188, 191, 192, 194, 256, 260, 266, 272, 273, 279
- CAB Thesaurus, 148, 156, 167
- cecha wyszukiwawcza, 139
- CISMef, 184, 194, 195, 196, 199, 249, 253, 280, 281, 282
- Claygate, 266
- DESIRE - Development of a European Service for Information on Research and Education, 10, 84, 85, 98, 110, 134, 140, 185, 208
- Digital Library of Earth System Education, 79
- dostęp przedmiotowy, 64
  - narzędzia, 78, 79, 80, 81, 184, 185, 194, 200, 206, 209, 220
  - budowa, 187, 189, 195, 198, 201, 202, 205, 210, 223, 230
  - wariant ahierarchiczny, 251
  - wariant hierarchiczny, 255, 258, 264, 266, 268, 271, 272, 273, 275, 276, 280
  - wariant hybrydowy, 282
- drzewa przedmiotowe, 88
- DTD - Document Type Definition, 16, 49, 50
- Dublin Core, 47, 49, 71, 143
- DutchESS, 125
- dziedzinowy system hipertekstowy
  - cechy wyszukiwawcze, 158, 159, 160
  - cel i zakres, 105, 106
  - charakterystyka treściowa i formalna zasobów, 112
  - definicje, 96, 97, 98, 99, 102
  - dostęp przedmiotowy, 123, 161, 171, 172, 174
  - geneza, 128, 129, 130, 131
  - mechizmy wyszukiwawcze, 161, 162, 163, 164, 165, 167, 168, 169
  - model analityczny, 125, 127, 128
  - polskie, 132
  - reprezentacja informacji, 140, 142, 144, 145
  - rezultaty wyszukiwania, 176, 177, 179, 182
  - schematy metadanych, 113, 122, 123
    - Edna metadata standard, 119, 120, 121, 122
    - Intute metadata element set, 115
    - ROADS/IAFA, 113
    - schemat SSG-FI, 116, 118
  - struktura, 100, 101, 105
  - typologia, 124
  - zarządzanie kolekcją, 107, 108, 110, 111, 112
- Edna Online, 106, 119, 145, 147, 155, 177
- EELS, 125, 131
- EEVL, 106, 125
- Ekonomia Online, 132
- eLib, The Electronic Libraries Programme, 57, 114, 128, 207, 215
- ELISAD, 133
- e-Przewodnik, 133
- ERIC Thesaurus, 143, 157, 204
- fasety, 91, 203, 204, 205, 281, 289
- folksonomie, 59, 60, 61, 62
- ForestryGuide, 116, 130, 141, 166
- FRBR, 25, 26
- GEM, 106, 125, 143, 157, 200, 280
- GeoGuide, 116, 130, 141, 149
- Getty Thesaurus of Geographic Names, The, 115, 144, 247
- Göttinger Online Klassifikation, 84, 150
- hasło korporatywne, 28
- hasło osobowe, 28
- hasło tytułowe, 28
- HASSET, Thesaurus, 148, 156, 167, 168, 170, 216, 286
- hipertekst
  - budowa, 40
  - geneza, 38, 39
- hipertekstowy model organizacji informacji, 37
- Historicus, 132
- HistoryGuide, 161
- HTML, 49, 50
- IAFA, 113
- IMesh Toolkit, 97
- Infomine, 147, 148, 161, 244, 249
- InfoRol, 132
- interoperacyjność, 48, 82, 116
- Intute, 45, 115, 116, 129, 141, 144, 147, 148, 149, 153, 156, 160, 167, 181, 182, 206, 207, 208, 210, 212, 214, 215, 219, 222, 225, 230, 231, 232, 243, 246, 247, 252, 253, 271, 272, 273, 279
- ISAAR, 142
- ISAD(G), 27
- ISBD, 26
- ISBD(ER), 26
- język deskryptorowy, 30, 146, 148, 240, 245, 285, 287
- język haseł przedmiotowych, 30, 146, 147, 153, 195, 240, 243, 244, 245, 285
- język informacyjno-wyszukiwawczy
  - funkcje, 28, 29
  - typologia, 30
- język słów kluczowych, 30, 53, 60, 62, 69, 80, 123, 146, 147, 152, 154, 155, 156, 162, 163, 165, 169, 185, 204, 240, 241, 285
- katalog alfabetyczny, 31
- katalog biblioteczny
  - funkcje, 31
- katalog rzeczowy, 31

katalogi internetowe, 55  
 kategoria tematyczna, 55, 56, 64, 70, 83, 86, 88, 89, 90, 101, 123, 124, 128, 137, 155, 159, 165, 173, 190  
 KBK, 83, 84, 122, 286, 288  
 KDD, 84, 149, 152, 188, 190, 255  
 KINIA, 133  
 klasyfikacja, 16, 146, 149, 153  
 Klasyfikacja Bibliograficzna Blissa, 84  
 Klasyfikacja Biblioteki Kongresu, *Patrz* KBK  
 Klasyfikacja Dwukropkowa Ranganathana, 84  
 Klasyfikacja Dziesiąta Deweya, *Patrz* KDD  
 kodowanie metainformacji, 24, 27, 73, 74, 122  
 LCSH, 148  
 Learning Object Metadata, 72, 143  
 Library of Congress Name Authority File, 140, 141, 142, 230, 247, 248, 250  
 literary warrant, 190, 219, 235, 270, 271, 273, 276, 277, 278, 282, 283, 288, 289  
 MARC, 16  
 Mathematics Subject Classification, 154, 234  
 MathGuide, 106, 116, 124, 130, 141, 154, 160, 163, 165, 177, 251  
 MedHist, 147, 220, 221, 227, 247, 259, 261  
 MeSH, 147, 148, 153, 196, 197, 198, 243, 253, 281  
 metadane, 16, 32, 47, 48, 50, 71, 72, 73, 100, 101, 112, 113, 114, 116, 119, 122, 123, 141, 159, 178, 201, 208, 284  
 metatermin, 195, 196, 198, 199, 200, 281  
 National Library of Medicine Classification, 85, 153  
 netografia, 134  
 NISS, 125  
 NIST Virtual Library, 76  
 NMAP, 152  
 obiekt informacyjny, 8, 17, 18, 21, 22, 25, 27, 31, 35, 47, 48, 65, 71, 73  
 OMNI, 106, 109, 125, 222  
 ontologie internetowe, 16, 70, 94, 206  
 OPAC, 31, 32  
 opis bibliograficzny, 24, 26  
 organizacja informacji  
   a organizacja wiedzy, 20  
   a organizacja wiedzy, 22  
   definicja, 23  
   elementy, 16  
   jednostki, 19  
   procesy, 24  
   przedmiot, 15  
 organizacja wiedzy, 22  
   a organizacja informacji, 22  
   definicja, 18  
   jednostki, 19  
   systemy organizacji wiedzy, 22, 23  
 PACS, 151, 288  
 paradoks Konigera i Janowitza, 37  
 Physics and Astronomy Classification Scheme, *Patrz* PACS  
 pierścienie synonimów, 77  
 polireprezentacja wiedzy, 145, 146, 153, 154  
   jawna, 155  
   ukryta, 156  
 przeglądanie, technika, 12, 65, 287  
 Psi-com, 266  
 punkty dostępu, 27, 28, 31  
   organizacja, 139  
 punkty dostępu przedmiotowego  
   koncepcja, 33  
   taksonomia, 34  
 quasifasetyzacja, 273, 277, 289  
 raport Folleta, 129  
 RDF, 49  
 relacja generyczna, 86  
 relacja hierarchii tematycznej, 217, 269, 270, 272, 273, 274, 276, 277, 281, 282, 288  
 relacja mereologiczna, 86  
 Renardus, 97, 122, 123, 130  
 reprezentacja informacji, 28, 139  
 reprezentacja wiedzy, 157  
 Research Methods & Tools Thesaurus, 148, 156  
 Resource Discovery Network (RDN), 129, 142  
 ROADS, 114  
 Royal College of Nursing Thesaurus, 148, 153, 246, 247  
 Science LinkHouse, 106, 125  
 Semantyczny Web, 16  
 serendipity (przypadkowe odkrywanie informacji), 68  
 SGML, 16, 49, 50  
 sieciowe systemy organizacji wiedzy, 93, 94  
   klasyfikacja, 94  
   listy relacyjne, 94  
   listy terminów, 94  
   typologia, 94  
 Slavistik-Portal, 151  
 Social Research Methodology Thesaurus, 167  
 SOSIG, 106, 142, 169  
 system informacyjno-wyszukiwawczy, 103  
   dokumentacyjny, 104  
 taksonomia internetowa, 85, 86, 88, 90, 94, 212, 231, 268, 274, 277, 279, 290  
   funkcjonalność, 86  
   struktura, 87, 88, 89, 90, 91, 92  
 Thesaurus for Graphic Materials, 149  
 Thesaurus Social Care Online, 149  
 TOL - Informator Technika Online, 133  
 twierdzenia wiedzy, 21  
 UKAT UK Archival Thesaurus, 149  
 UKD, 83, 84, 122, 216, 222, 255, 286, 288  
 UNIMARC, 27  
 Union List of Artist Names, 141, 247  
 Uniwersalna Klasyfikacja Dziesiąta, *Patrz* UKD  
 uniwersum wiedzy, 290  
 user warrant, 272  
 USMARC, 27  
 VetGate, 106, 108  
 ViFaPhys, 151, 279  
 Web 2.0, 59  
 wyszukiwarki internetowe, 51, 52  
 XHTML, 50  
 XML, 16, 247

## Aneksy

### Załącznik 1. Lista dziedzinowych systemów hipertekstowych w Internecie (stan na 24-11-2008)

Lp.	Nazwa	URL
1.	Access to Law	<a href="http://www.accesstolaw.com/">http://www.accesstolaw.com/</a>
2.	ADAM	<a href="http://www.adam.ac.uk/">http://www.adam.ac.uk/</a> (włączone do Intute) wersja archiwalna: <a href="http://web.archive.org/web/20060101180917/http://adam.ac.uk/">http://web.archive.org/web/20060101180917/http://adam.ac.uk/</a>
3.	AECPORTICO	<a href="http://www.thenbs.com/resources/index.asp">http://www.thenbs.com/resources/index.asp</a>
4.	AERADE	<a href="http://aerade.cranfield.ac.uk/devise_index.html">http://aerade.cranfield.ac.uk/devise_index.html</a>
5.	AGRIGATE	<a href="http://www.agnic.org/browse">http://www.agnic.org/browse</a>
6.	Anglo-American Culture	<a href="http://www.sub.uni-goettingen.de/ssgfi/anglo-america.html">http://www.sub.uni-goettingen.de/ssgfi/anglo-america.html</a>
7.	Arkade	<a href="http://arkade.dk/">http://arkade.dk/</a>
8.	BazTol	<a href="http://centr5.ml.put.poznan.pl/baztol_czytelnik/baztol">http://centr5.ml.put.poznan.pl/baztol_czytelnik/baztol</a>
9.	BeCal - Belief, Culture and Learning Information Gateway:	<a href="http://www.becal.net/datapool/index.php">http://www.becal.net/datapool/index.php</a>
10.	Biogate	<a href="http://biogate.lub.lu.se/">http://biogate.lub.lu.se/</a> (nieczynny) wersja archiwalna: <a href="http://web.archive.org/web/20010721101318/http://biogate.lub.lu.se/">http://web.archive.org/web/20010721101318/http://biogate.lub.lu.se/</a>
11.	Bizigate	<a href="http://www.bizigate.dk/?lang=en">http://www.bizigate.dk/?lang=en</a>
12.	BUBL Information Service	<a href="http://bubl.ac.uk/">http://bubl.ac.uk/</a>
13.	Chemdex	<a href="http://www.chemdex.org/">http://www.chemdex.org/</a>
14.	Child and Family Web Guide	<a href="http://www.cfw.tufts.edu/">http://www.cfw.tufts.edu/</a>
15.	CISMef	<a href="http://www.chu-rouen.fr/cismef/">http://www.chu-rouen.fr/cismef/</a>
16.	ClayGate 2008	<a href="http://library.bendigo.latrobe.edu.au/irs/WEBCAT/ddc_index.htm">http://library.bendigo.latrobe.edu.au/irs/WEBCAT/ddc_index.htm</a>
17.	Digital Librarian	<a href="http://www.digital-librarian.com/">http://www.digital-librarian.com/</a>
18.	Dvm	<a href="http://dvm.nu/?lang=en&amp;?format=medium&amp;type=all">http://dvm.nu/?lang=en&amp;?format=medium&amp;type=all</a>
19.	EDNA Education Network Australia	<a href="http://www.edna.edu.au/edna/go">http://www.edna.edu.au/edna/go</a>
20.	EELS - Engineering E-Library	<a href="http://eels.lub.lu.se/">http://eels.lub.lu.se/</a> (nieczynny) wersja archiwalna: <a href="http://web.archive.org/web/20070625160503/http://eels.lub.lu.se/">http://web.archive.org/web/20070625160503/http://eels.lub.lu.se/</a>
21.	EEVL	<a href="http://www.eevl.ac.uk/">http://www.eevl.ac.uk/</a> (włączony do Intute) wersja archiwalna: <a href="http://web.archive.org/web/20060215070253/http://www.eevl.ac.uk/">http://web.archive.org/web/20060215070253/http://www.eevl.ac.uk/</a>
22.	EGIL – Electronic Gateway for Icelandic Literature	<a href="http://www-db.library.nottingham.ac.uk/egil/">http://www-db.library.nottingham.ac.uk/egil/</a>
23.	Ekonomia Online	<a href="http://kangurold.uek.krakow.pl/Biblioteka/Ekonomia/">http://kangurold.uek.krakow.pl/Biblioteka/Ekonomia/</a>
24.	ELISAD	<a href="http://www.addictionsinfo.eu/startpage.php?select_lang=pol">www.addictionsinfo.eu/startpage.php?select_lang=pol</a>
25.	Energiportal	<a href="http://www.energiportal.dk/?lang=en">http://www.energiportal.dk/?lang=en</a>
26.	food-i	<a href="http://www.food-i.dk/?lang=en">http://www.food-i.dk/?lang=en</a>
27.	Forestry Guide	<a href="http://www.forestryguide.de/">http://www.forestryguide.de/</a>
28.	GEES Electronic Information Gateway for Geography, Earth and Environmental Sciences with search engine	<a href="http://www.gees.ac.uk/db/">http://www.gees.ac.uk/db/</a>
29.	GEM Gateway to Educational Materials	<a href="http://www.thegateway.org/">http://www.thegateway.org/</a>
30.	GeoGuide	<a href="http://www.geo-guide.de/">http://www.geo-guide.de/</a>

31.	GROW Geotechnical, Rock and Water Resources Library	<a href="http://www.grow.arizona.edu/SPT--BrowseResources.php">http://www.grow.arizona.edu/SPT--BrowseResources.php</a>
32.	Hardin MD : Medical Information + Pictures	<a href="http://www.lib.uiowa.edu/hardin/md/">http://www.lib.uiowa.edu/hardin/md/</a>
33.	HealthWeb	<a href="http://healthweb.org/">http://healthweb.org/</a> (nieczynny) wersja archiwalna: <a href="http://web.archive.org/web/20071013095110/http://www.healthweb.org/">http://web.archive.org/web/20071013095110/http://www.healthweb.org/</a>
34.	Higher Education Academy Psychology Network	<a href="http://www.psychology.heacademy.ac.uk/html/resources.asp">http://www.psychology.heacademy.ac.uk/html/resources.asp</a>
35.	Historicus	<a href="http://historicus.umk.pl/">http://historicus.umk.pl/</a>
36.	History On-line	<a href="http://www.history.ac.uk/ihr/Resources/index.html">http://www.history.ac.uk/ihr/Resources/index.html</a>
37.	IMarEST - Maritime Links Directory	<a href="http://www.imarest.org/links/">http://www.imarest.org/links/</a>
38.	Infolaw	<a href="http://www.infolaw.co.uk/lawfinder/">http://www.infolaw.co.uk/lawfinder/</a>
39.	Infomine	<a href="http://infomine.ucr.edu/">http://infomine.ucr.edu/</a>
40.	InfoRol	<a href="http://www.cbr.edu.pl/inforol_form.htm">http://www.cbr.edu.pl/inforol_form.htm</a>
41.	Intute Arts and Humanities	<a href="http://www.intute.ac.uk/artsandhumanities/">http://www.intute.ac.uk/artsandhumanities/</a>
42.	Intute Health and Life Sciences	<a href="http://www.intute.ac.uk/healthandlifesciences/">http://www.intute.ac.uk/healthandlifesciences/</a>
43.	Intute Science & Technology	<a href="http://www.intute.ac.uk/sciences/">http://www.intute.ac.uk/sciences/</a>
44.	Intute Social Sciences	<a href="http://www.intute.ac.uk/socialsciences/">http://www.intute.ac.uk/socialsciences/</a>
45.	Juraportal	<a href="http://www.juraportal.dk/?lang=en">http://www.juraportal.dk/?lang=en</a>
46.	KINIA	<a href="http://www.kinia.czytelnia.net/">http://www.kinia.czytelnia.net/</a>
47.	Librarian's Index to the Internet	<a href="http://lii.org/">http://lii.org/</a>
48.	Materials	<a href="http://www.materials.dk/">http://www.materials.dk/</a>
49.	MathForum	<a href="http://mathforum.org/library/topics/">http://mathforum.org/library/topics/</a>
50.	MathGuide	<a href="http://www.mathguide.de/">http://www.mathguide.de/</a>
51.	MCS-Media and Communication Studies	<a href="http://www.aber.ac.uk/media/index.html">http://www.aber.ac.uk/media/index.html</a>
52.	Medweb	<a href="http://www.medweb.emory.edu/MedWeb/">http://www.medweb.emory.edu/MedWeb/</a>
53.	Mental Health & Psychology Resources Online	<a href="http://psychcentral.com/resources/">http://psychcentral.com/resources/</a>
54.	Moving Image Gateway - BUFVC	<a href="http://www.bufvc.ac.uk/gateway/">http://www.bufvc.ac.uk/gateway/</a>
55.	Nestor. Informationsplattform	<a href="http://nestor.sub.uni-goettingen.de/nestor_on/index.php?lang=en">http://nestor.sub.uni-goettingen.de/nestor_on/index.php?lang=en</a>
56.	PADI - Preserving Access to Digital Information	<a href="http://www.nla.gov.au/padi/">http://www.nla.gov.au/padi/</a>
57.	PORT	<a href="http://www.port.nmm.ac.uk/">http://www.port.nmm.ac.uk/</a> (włączone do Intute) wersja archiwalna: <a href="http://web.archive.org/web/20060927045842/http://www.port.nmm.ac.uk/">http://web.archive.org/web/20060927045842/http://www.port.nmm.ac.uk/</a>
58.	Ppportal	<a href="http://www.ppportal.dk/?lang=en">http://www.ppportal.dk/?lang=en</a> (nieczynny) wersja archiwalna: <a href="http://web.archive.org/web/20070712182803/http://www.ppportal.dk/">http://web.archive.org/web/20070712182803/http://www.ppportal.dk/</a>
59.	DePaul University Libraries	<a href="http://library.depaul.edu/Browse%20Subjects/bycategory.aspx">http://library.depaul.edu/Browse%20Subjects/bycategory.aspx</a>
60.	Research Guide for Students	<a href="http://www.aresearchguide.com/index.html">http://www.aresearchguide.com/index.html</a>
61.	SAPLING: The Architecture, Planning & Landscape Information Gateway	<a href="http://www.sapling.org.uk/gateway/">http://www.sapling.org.uk/gateway/</a>
62.	Serwis WiMBP im. J. Piłsudskiego w Łodzi	<a href="http://mak.wimbp.lodz.pl/informacja/">http://mak.wimbp.lodz.pl/informacja/</a>
63.	Social Care Online	<a href="http://www.scie-socialcareonline.org.uk">http://www.scie-socialcareonline.org.uk</a>
64.	SocioSite	<a href="http://www.sociosite.net/index.php">http://www.sociosite.net/index.php</a>
65.	SWAP (Social Policy and Social Work)	<a href="http://www.swap.ac.uk/Links/linkindex.asp">http://www.swap.ac.uk/Links/linkindex.asp</a> (nieczynny) wersja archiwalna: <a href="http://web.archive.org/web/20071008161245/http://www.swap.ac.uk/Links/linkindex.asp">http://web.archive.org/web/20071008161245/http://www.swap.ac.uk/Links/linkindex.asp</a>

		<a href="http://www.swap.ac.uk/Links/linkindex.asp">www.swap.ac.uk/Links/linkindex.asp</a>
66.	The WWW Virtual Library	<a href="http://www.vlib.org.uk/">http://www.vlib.org.uk/</a>
67.	Tieteen linkkitalo - Science Linkhouse	<a href="http://www.linkkitalo.fi/">http://www.linkkitalo.fi/</a>
68.	TOL - Informator Technika Online	<a href="http://www.ml.put.poznan.pl/TOL/index.html">http://www.ml.put.poznan.pl/TOL/index.html</a>
69.	Vifaphys Physics Web	<a href="http://www.vifaphys.de/fi.php?lang=en&amp;skin=gui&amp;mode=classification">http://www.vifaphys.de/fi.php?lang=en&amp;skin=gui&amp;mode=classification</a>
70.	WebEc - World Wide Web Resources in Economics	<a href="http://www.helsinki.fi/WebEc/">http://www.helsinki.fi/WebEc/</a>
71.	WebLaw	<a href="http://www.weblaw.edu.au/">http://www.weblaw.edu.au/</a>

## Załącznik 2. Grupa reprezentatywna dziedzinowych systemów hipertekstowych

Nazwa	URL
Aerade	<a href="http://aerade.cranfield.ac.uk/devise_index.html">http://aerade.cranfield.ac.uk/devise_index.html</a>
Agrifor	<a href="http://www.intute.ac.uk/healthandlifesciences/agriculture/">http://www.intute.ac.uk/healthandlifesciences/agriculture/</a>
Altis	<a href="http://www.altis.ac.uk/">www.altis.ac.uk/</a> obecnie jako część <a href="http://www.intute.ac.uk/socialsciences/">http://www.intute.ac.uk/socialsciences/</a>
AnglistikGuide	<a href="http://www.anglistikguide.de/">http://www.anglistikguide.de/</a>
Arkade	<a href="http://arkade.dk/">http://arkade.dk/</a>
Artifact	<a href="http://www.artifact.ac.uk">www.artifact.ac.uk</a> obecnie jako część <a href="http://www.intute.ac.uk/artsandhumanities/">http://www.intute.ac.uk/artsandhumanities/</a>
BioethicsWeb	<a href="http://www.intute.ac.uk/healthandlifesciences/bioethicsweb/">http://www.intute.ac.uk/healthandlifesciences/bioethicsweb/</a>
Biogate	<a href="http://biogate.lub.lu.se/">http://biogate.lub.lu.se/</a>
BioResearch	<a href="http://www.intute.ac.uk/healthandlifesciences/bioresearch/">http://www.intute.ac.uk/healthandlifesciences/bioresearch/</a>
BUBL	<a href="http://bubl.ac.uk/">http://bubl.ac.uk/</a>
Chemdex	<a href="http://www.chemdex.org">http://www.chemdex.org</a>
EdNa	<a href="http://www.edna.edu.au">www.edna.edu.au</a>
EELS	<a href="http://eels.lub.lu.se/">http://eels.lub.lu.se/</a>
EEVL Informatyka	<a href="http://www.eevl.ac.uk/computing/index.htm">http://www.eevl.ac.uk/computing/index.htm</a> obecnie jako część <a href="http://www.intute.ac.uk/socialsciences">http://www.intute.ac.uk/socialsciences</a>
EEVL Inżynieria	<a href="http://www.eevl.ac.uk/engineering/index.htm">http://www.eevl.ac.uk/engineering/index.htm</a> obecnie jako część <a href="http://www.intute.ac.uk/socialsciences">http://www.intute.ac.uk/socialsciences</a>
EEVL Matematyka	<a href="http://www.eevl.ac.uk/mathematics/index.htm">http://www.eevl.ac.uk/mathematics/index.htm</a> obecnie jako część <a href="http://www.intute.ac.uk/socialsciences">http://www.intute.ac.uk/socialsciences</a>
ForestryGuide	<a href="http://www.forestryguide.de/">www.forestryguide.de/</a>
GEES	<a href="http://www.gees.ac.uk/db/">http://www.gees.ac.uk/db/</a>
GEM	<a href="http://www.thegateway.org">www.thegateway.org</a>
GeoGuide	<a href="http://www.geo-guide.de/">http://www.geo-guide.de/</a>
Humbul	<a href="http://www.humbul.ac.uk">www.humbul.ac.uk</a> obecnie jako część <a href="http://www.intute.ac.uk/artsandhumanities/">http://www.intute.ac.uk/artsandhumanities/</a>
Infomine Bio, Ag Med. Sciences	<a href="http://infomine.ucr.edu/">http://infomine.ucr.edu/</a>
Intute Arts&Humanities	<a href="http://www.intute.ac.uk/artsandhumanities/">http://www.intute.ac.uk/artsandhumanities/</a>
Intute Social Sciences	<a href="http://www.intute.ac.uk/socialsciences">http://www.intute.ac.uk/socialsciences</a>
Intute Science, Engineering and Technology wraz z wcześniejszym systemem SOSIG	<a href="http://www.intute.ac.uk/sciences/">http://www.intute.ac.uk/sciences/</a>
MathForum	<a href="http://mathforum.org/library/topics/">http://mathforum.org/library/topics/</a>
MathGuide	<a href="http://www.mathguide.de/">http://www.mathguide.de/</a>
MedHist	<a href="http://www.intute.ac.uk/healthandlifesciences/medhist/">http://www.intute.ac.uk/healthandlifesciences/medhist/</a>
NaturalSelection	<a href="http://www.intute.ac.uk/healthandlifesciences/nature/">http://www.intute.ac.uk/healthandlifesciences/nature/</a>
Psi-com	<a href="http://www.intute.ac.uk/healthandlifesciences/psicom/">http://www.intute.ac.uk/healthandlifesciences/psicom/</a>