



Comparative Legilinguistics
vol. 40/2019

DOI: <http://dx.doi.org/10.14746/cl.2019.40.2>

ISTOTA SIECI SEMANTYCZNYCH – PUŁAPKI I KORZYŚCI ZASTOSOWANIA W SĄDOWNICTWIE

dr hab. Monika ODLANICKA-POCZOBUTT

Politechnika Śląska
Wydział Organizacji i Zarządzania
Instytut Zarządzania, Administracji i Logistyki
Poland
modlanicka@polsl.pl

ORCID: <https://orcid.org/0000-0001-7834-1188>

Abstrakt: Sieci semantyczne (ang. Semantic Web) pozwalają w coraz większym stopniu na przetwarzanie informacji za pośrednictwem Internetu zgodnie z ich zawartością i przeznaczeniem. Rozpowszechnienie standardów metadanych spowodowało jednak problemy z interoperacyjnością między różnymi systemami. Umożliwienie użytkownikom przeszukiwania różnych języków wymaga zasobów tłumaczeniowych, aby przekroczyć barierę językową, może stanowić poważne wyzwanie. Zaprezentowana tendencja do korzystania z sieci semantycznych jest szansą dla prawodawstwa, ułatwiającą procesy opracowania i udostępniania przepisów prawnych oraz zarządzania postępowaniem prac legislacyjnych i realizacji procedur. Celem artykułu jest przegląd teorii sieci semantycznych w odniesieniu do pułapek związanych z definiowaniem i stosowaniem ontologii na potrzeby zrozumienia

maszynowego treści przechowywanych w sieci. Zakres artykułu obejmuje wstępną analizę aspektów współpracy grup interdyscyplinarnych na etapie tworzenia metadanych – indeksowane według analiz koncepcyjnych (ontologii) danej dziedziny. Obszarem zainteresowań badawczych były również korzyści i pułapki zastosowania Open Source, XML i Legal Semantic Web w sądownictwie powszechnym.

Słowa kluczowe: sieci wiedzy; ICT; WWW; DBpedia; sądownictwo powszechne; XML; Legal Semantic Web.

SIGNIFICANCE OF SEMANTIC WEB – PITFALLS AND BENEFITS OF USE IN THE COMMON JUDICIARY

Abstract: Semantic Webs increasingly allow the processing of information over the Internet according to their content and purpose. The widespread use of metadata standards has, however, caused interoperability problems between different systems. Enabling users to search for different languages requires translation resources to cross the language barrier can be a serious challenge. The presented tendency to use semantic webs is an opportunity for legislation, facilitating the process of developing and providing legal regulations and managing the progress of legislative work and implementation of procedures. The aim of this article is to review semantic web theory in relation to the pitfalls of defining and applying ontology to the machine-based understanding of content stored in the network. The scope of this article is a preliminary analysis of aspects of interdisciplinary cooperation in the metadata creation phase – indexed according to the conceptual analysis (ontology) of a given field. The area of research interest was also the benefits and pitfalls of using open source, XML and the Legal Semantic Web in the common judiciary.

Key words: semantic web; ICT; WWW; DBpedia; common judiciary; XML; Legal Semantic Web.

1. Wprowadzenie

Wykorzystanie rozwiązań ICT (ang. *Information and Communication Technologies*) w zunifikowanej komunikacji (Itknowledgeexchange.techtarget.com; Silverstone 1991: 204–227) pozwala użytkownikom na dostęp, przechowywanie, przesyłanie i przetwarzanie informacji, obejmując całą gamę aplikacji

informatycznych umożliwiających przetwarzanie informacji i obsługę bieżących funkcji na bazie wspólnej technologii cyfrowej (Sallai 2012: 5-15). Pojawia się coraz więcej obszarów, w których rozwiązania oparte o technologie ICT znajdują zastosowanie – bankowość, telewizja cyfrowa, sieci społeczne, administracja elektroniczna oraz sądownictwo powszechne (Odlanicka-Poczobutt 2013: 553–561; Odlanicka-Poczobutt 2014: 93–101).

W miarę rozwoju Internetu oraz jego dynamicznej ekspansji – coraz większym problemem okazuje się jego najprostsza i najbardziej użyteczna funkcja – wyszukiwanie informacji. Strony WWW i język HTML umożliwiają bardzo łatwą publikację oraz prezentację informacji, ale niestety wykorzystują nieefektywne sposoby jej kategoryzacji. Wyszukiwarki internetowe zwracają mnóstwo niepotrzebnych informacji, co z kolei wymaga czasochłonnej ręcznej filtracji dokonywanej przez użytkownika.

Największy zatem problem to – problem z integracją informacji. Użytkownik szukający informacji, przebrnąwszy przez szereg stron, znajdzie poszukiwane dane, jednak o wiele trudniejszym do rozwiązania problemem jest komunikacja i inteligentna wymiana informacji pomiędzy maszynami komputerowymi. Nie chodzi tutaj o współpracę sieci serwerów, które są do tego przystosowane, ale o bardziej wyrafinowaną wymianę informacji, która umożliwiłaby komunikację maszyn, systemów, oprogramowania, które nigdy wcześniej ze sobą nie współpracowały, nie wiedząc często o swoim współistnieniu. Wzrost siły Internetu nastąpi, kiedy systemy będą mogły dynamicznie wyszukiwać inne jednostki obliczeniowe oraz komunikować się z nimi w zrozumiałym sposób.

Istotne znaczenie ma tutaj zgodność semantyczna (sjp.pwn.pl/), w opozycji do stosowanej przez maszyny komputerowe zgodności syntaktycznej. Zgodność syntaktyczna wymusza poprawność informacji pod względem ich prawidłowej budowy – składni (komputer weryfikuje czy wszystkie elementy wchodzące w skład informacji są prawidłowe, czy istnieją, np. w zdaniu weryfikowane są wyrazy pod względem ich istnienia). Tego typu weryfikacja nie jest w stanie potwierdzić czy dane, pomimo poprawności syntaktycznej, mają sens. Semantyka jest działem logiki, odpowiadającym za przypisanie znaczenia symbolom i elementom języka oraz sensu zbiorom tych symboli. Zgodność semantyczna wymaga mapowania oraz weryfikowania relacji pomiędzy elementami informacji, co z kolei wymusza analizę zawartości informacji. Semantyka wymusza

formalną oraz wyraźną specyfikację modelu dziedziny, który definiuje pojęcia oraz relacje pomiędzy składowymi tego modelu. Takie formalne modele, nazywane ontologiami¹ będą odgrywały decydującą rolę podczas umożliwiania przetwarzania oraz współdzielenia wiedzy pomiędzy programami w sieci WWW.

Problemy związane z niezgodnością protokołów oraz standardów zauważono już kilka lat temu. W celu umożliwienia integracji stworzono nową technologię tzw. Usługi Sieciowe (ang. Web Services), które umożliwiają komunikację oraz wymianę informacji pomiędzy niekompatybilnymi platformami (4programmers.net). To duże osiągnięcie w kierunku ujednoczenia oraz standaryzacji protokołów komunikacji (SOAP). Aby maszyny mogły samodzielnie wyszukiwać, tworzyć kompozycje usług oraz je wywoływać – istotne znaczenie będą miały aspekty semantyczne, ludzkie, prawne i międzynarodowe związane z ich wykorzystaniem. Pod względem wykorzystania semantyki, przeprowadzone przez G. Alemu badania sugerują, że rozpowszechnienie standardów metadanych spowodowało problemy z interoperacyjnością między różnymi systemami (Alemu, Stevens and Ross 2012: 38-54). Dodatkowo istotną kwestią jest fakt, że umożliwienie użytkownikom przeszukiwania różnych języków wymaga zasobów tłumaczeniowych, aby przekroczyć barierę językową, co może stanowić poważne wyzwanie (Diekema 2012: 165-181).

Celem artykułu jest przegląd teorii sieci semantycznych w odniesieniu do pułapek związanych z definiowaniem i stosowaniem ontologii na potrzeby zrozumienia maszynowego treści przechowywanych w sieci. Zakres artykułu obejmuje wstępną analizę aspektów współpracy grup interdyscyplinarnych na etapie tworzenia meta danych – indeksowane według analiz koncepcyjnych (ontologii) danej dziedziny. Obszarem zainteresowań badawczych były również korzyści i pułapki zastosowania Open Source, XML i Legal Semantic Web w sądownictwie powszechnym.

¹ **Ontologia** – w sensie pierwotnym — termin używany od XVII w. (J. Clauberg, Ch. Wolff) zamiennie ze starszą nazwą metafizyka; odnosi się do nauki o bycie; w fenomenologii R. Ingardena — odróżniona od metafizyki nauka aprioryczna, która w analizie zawartości idei odkrywa i ustala czyste możliwości oraz związki konieczne między czystymi jakościami idealnymi. (<https://encyklopedia.pwn.pl/haslo/ontologia;3951174.html>); w informatyce – zbiór pojęć i opisów relacji zachodzącymi pomiędzy tymi pojęciami w zakresie danej i ustalonej dziedziny.

2. Teoria sieci semantycznej

Teoria sieci semantycznej (ang. Semantic Web) odnosi się do sposobu organizacji reprezentacji pojęciowych zakładając, że pojęcia są przechowywane trwale w pamięci jako struktury sieciowe o hierarchicznym charakterze (Berners Lee, Hendler, Lassila 2001: 41–49). Sieci semantyczne są jedną z metod reprezentacji wiedzy powstałą na wzór sposobu postrzegania rzeczywistości przez człowieka i działania pamięci.

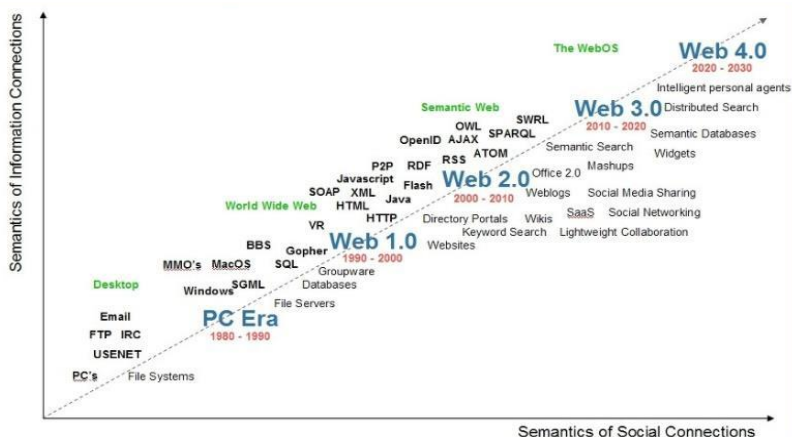
Sieć semantyczna składa się z punktów węzłowych i relacji pomiędzy nimi. Relacja semantyczna dwóch reprezentacji pojęciowych odnosi się do sumy wszystkich połączeń między desygnatami i właściwościami. Z formalnego punktu widzenia, można je przedstawić jako graf skierowany etykietowany. Węzłami tego grafu są pojęcia, które mogą być ze sobą powiązane krawędziami (Lula, Stal, Tadeusiewicz, Morajda, Paliwoda-Pękosz, Wilusz 2012: 193).

Pojęcia blisko związane cechuje wielość połączeń między nimi. Połączenia, nazywane także ścieżkami sieci cechuje różna waga – silniejszy związek cechuje większa waga połączenia, co skutkuje łatwiejszą wzajemną aktywacją pojęć. Teorię sieci semantycznej charakteryzują dwa rodzaje relacji semantycznych – połączenia pozytywne (np. „kos jest to ptak”) i ścieżki zaprzeczające (np. „pies nie jest to kot”) (Nęcka, Orzechowski, Szymura 2006: 126-128).

Sieci semantyczne, określane również jako sieci wiedzy, pozwalają w coraz większym stopniu na przetwarzanie informacji za pośrednictwem Internetu zgodnie z ich zawartością i przeznaczeniem, nie tylko jako czysty tekst, czyli sekwencja słów do odczytu przez człowieka.

Wynik ten osiąga się zazwyczaj poprzez osadzenie w naturalnych językach tekstowych specjalnych komputerowych specyfikacji, które mogą być wykorzystywane w różny sposób: do pobierania dokumentów, dostępu do odpowiednich informacji, w celu określania zawartości dokumentu itd. (system połączonego dokumentu jest dostępny w Internecie, dzięki wykorzystaniu protokołu HTTP). Stanowi to znaczny postęp w stosunku do sytuacji, w której dokumenty internetowe (strony internetowe) w większości zawierają teksty w języku naturalnym lub w formie plików multimedialnych.

Sieć semantyczna jest wizją Tima Bernersa-Lee (twórcy standardu WWW i pierwszej przeglądarki internetowej, a także przewodniczącego konsorcjum W3C). W swoich założeniach sieć semantyczna ma korzystać z istniejącego protokołu komunikacyjnego, na którym bazuje dzisiejszy Internet. Różnica miałaby polegać na tym, że przesyłane dane mogłyby być „rozumiane” także przez maszyny. Owo „rozumienie” polegałoby na tym, że dane przekazywane byłyby w postaci, w której można by powiązać ich znaczenia między sobą a także w ramach odpowiedniego kontekstu. Przewidywany rozwój sieci semantycznych przedstawiono na rysunku 1.



Rysunek. 1. Rozwój sieci semantycznych

Źródło: Radar Networks & Nova Spivack. 2007 <https://www.radarnetworks.com> [11.2017]

Aby Web w wersji 3.0 był w stanie zrozumieć użytkownika i dopasować, przeanalizować i wyświetlić odpowiednie dla niego treści wymaga czegoś, co pozwoliłoby działać na danych. Tym czymś może być właśnie sieć semantyczna. Określeniem Web 3.0 można nazwać więc sposób korzystania z Internetu opartego o sieć semantyczną (Rudman R., Bruwer 2006: 132 - 154).

Informacje przekazywane w ramach sieci semantycznej wymagają zatem nie tylko samych danych, ale także informacji o nich, czyli tzw. meta-danych, które zawierałyby sformułowania dotyczące relacji między danymi oraz prawa logiki, które można do nich zastosować. Dzięki wprowadzeniu takiego opisu można:

- powiązać różne dane znajdujące się w Internecie w ramach wspólnych jednostek znaczeniowych (np. strony dotyczące dziedzin nauki, muzyki, kuchni, etc.),
- rozróżnić dane, które dla maszyn są w tej chwili nierozróżnialne ze względu na identyczny zapis tekstowy (np. mysz – urządzenie komputerowe, zwierzątko; zamek – urządzenie do zamykania drzwi, element ubrania, okazała budowla mieszkalno-obronna),
- przeprowadzać na tych danych wnioski, tzn. otrzymywać informacje na ich temat, które nie są zawarte explicite (np. na podstawie danej „Ewa jest żoną Adama”, możemy też dowiedzieć się, że Ewa jest kobietą, Adam mężczyzną, Adam jest mężem Ewy, żaden inny mężczyzna nie jest mężem Ewy zgodnie z polskim prawem, etc.)

Z jednej strony Internet stanowi potężną sieć informacyjną, w której przechowywana jest ogromna wiedza i funkcjonują miliony maszyn, które umożliwiają wymianę informacji, ale pomimo swej potężnej mocy obliczeniowej nadal wykorzystywane są jako urządzenia do prezentacji, przesyłania oraz sztucznego przetwarzania danych, których nie rozumieją. Dla komputera słowo mysz, czy zamek nie ma żadnego znaczenia, nie jest semantycznie powiązane z żadnym pojęciem. Z tego powodu jego potężna moc obliczeniowa jest ograniczona jedynie do prezentacji niezrozumiałych dla niego znaków.

Internet rozwijał się gwałtownie jako medium służące przede wszystkim do publikacji dokumentów przeznaczonych dla ludzi, a nie dla informacji, które mogłyby być automatycznie przetwarzane przez maszyny. Celem sieci semantycznych jest umożliwienie tego typu funkcjonalności.

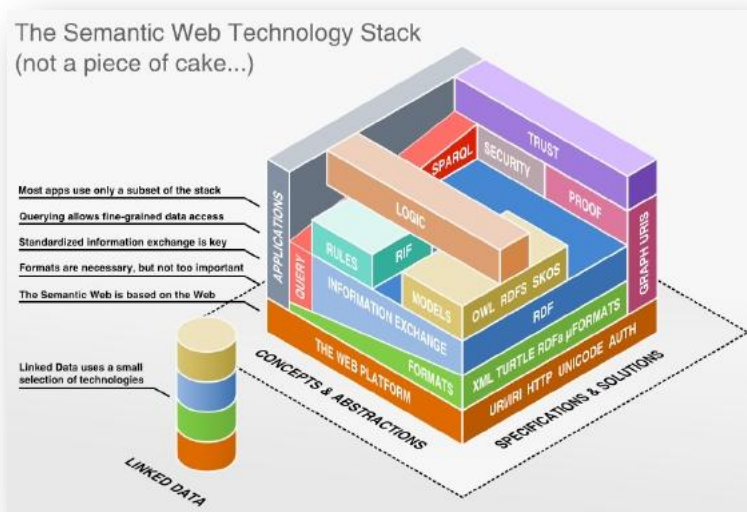
3. Techniczne podstawy sieci semantycznych

Semantic Web zbudowany powinien być na bazie już istniejących, wykorzystywanych i sprawdzonych standardów internetowych, nadbudowanych przez kilka kolejnych. Ponieważ każdy kolejny standard nakłada się na kolejny, dotyczący innego poziomu abstrakcji,

konstrukcja nosi nazwę warstwowego ciastka (ang. layer cake). Inne popularne określenie to semantyczny stos (ang. semantic stack).

Projekt sieci semantycznych powinien się przyczynić do utworzenia i rozpowszechnienia standardów opisywania treści w Internecie w sposób, który umożliwi maszynom i programom (np. tzw. agentom) przetwarzanie informacji w sposób odpowiedni do ich znaczenia. Wśród standardów sieci semantycznych znajdują się m.in. OWL, RDF, RDF Schema (inaczej RDFS). Znaczenia zasobów informacyjnych określa się za pomocą tzw. ontologii.

Układ sieci semantycznych – semantyczny stos – przedstawiono na rysunku 2.



Rysunek. 2. Układ sieci semantycznych - semantyczny stos

Źródło: Kruk S., Prezentacja Sieć Semantyczna w teorii i praktyce, <https://www.slideshare.net/skruk/sie-semantyczna-w-teorii-i-praktyce> [11.2017]

Kolejne warstwy stosu to (od dołu):

- **Unicode** - standard pozwalający na wyrażenie w języku maszyn dowolnego znaku pisanego, dowolnego języka znanego na Ziemi. Standard ten rozwiązuje problem niedogodności związanych z prymatem w świecie komputerów podstawowego alfabetu łańciskiego;
- **URI** - standard zapewniający unikatowość zasobów internetowych, które wymagają określenia dla nich

identyfikatora. Identyfikator będzie mógł składać się z zestawu znaków Unicode;

- **XML i XML Schema** XML (ang. Extensible Markup Language) – uniwersalny język znaczników przeznaczony do reprezentowania różnych danych w strukturalizowany sposób. Jest niezależny od platformy, co umożliwia łatwą wymianę dokumentów pomiędzy heterogenicznymi (różnymi) systemami i znacząco przyczyniło się do popularności tego języka w dobie Internetu. XML jest pozwalającym na zapis danych standardem rekomendowanym oraz specyfikowanym przez organizację W3C. Schematy XML wprowadzają ograniczenia dotyczące typu i struktury danych – ich zachowanie daje gwarancję, że dane XML są poprawne w sensie syntaktycznym (np. w polu, w którym oczekiwana jest wartość liczbowa, wartość taka się pojawi);
- **RDF i RDF Schema** (ang. Resource Description Framework) – standard, który pozwala na zapis danych w postaci grafu skierowanego, w którym dane zawarte są w wierzchołkach a relacje pomiędzy nimi i ich własności znajdują się w krawędziach. Schematy RDF wprowadzają do grafów takie pojęcia jak klasy i podklasy, pozwalające na wspólne grupowanie danych mających cechy wspólne. Dowolna dana może znajdować się w wielu klasach;
- **OWL** (ang. Web Ontology Language) – jest standardem pozwalającym na definiowanie klas na podstawie własności danych, a także na definiowanie logicznych charakterystyk relacji. OWL jest więc standardem formalnie zapisującym ontologię (Rysunek 3);



Rysunek. 3. Przykład ontologii

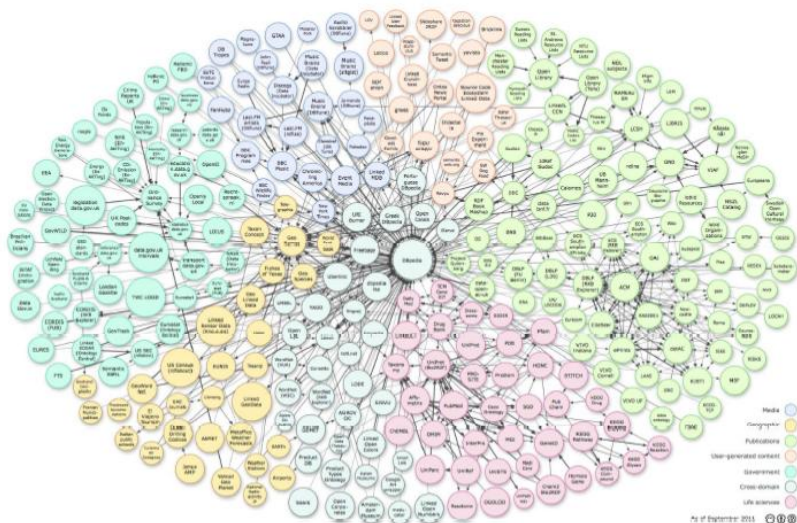
Źródło: Semantyczna Sieć: Boty uczą się kojarzyć. <https://www.chip.pl> [11.2017]

- **mechanizmy wnioskowania** - język OWL pozwala na definiowanie zależności między danymi, dzięki czemu można przeprowadzać wnioskowanie. Jednakże osobną sprawą jest przygotowanie odpowiednich mechanizmów, które takie wnioskowanie przeprowadzą poprawnie, a jednocześnie w odpowiednio szybkim czasie. Narzędziem programowym – silnikiem, który jest w stanie wnioskować na podstawie zdań, przesłanek, aksjomatów w standardach sieci semantycznej jest semantic reasoner (ang. reasoning – rozumowanie) (Elenius, Denker, Stehr 2008: 135-149). Sposób działania takiego silnika umożliwiają semantyka języków – ontologii (OWL) i opisowa (RDFS). Narzędzia programowe typu reasoner automatyzują proces wnioskowania na podstawie języka OWL. Zbierają i analizują aksjomaty (gotowe prawdziwe opisy) zebrane w ontologii i klasy, ich instancje, podklasy, nadklasy. W wyniku zadania wnioskowania wykonują klasyfikację istniejących obiektów, przypisują im właściwości, modyfikują i tworzą hierarchie klas zgodnie z własnościami, regułami i ograniczeniami OWL (Mishra, Kumar 2011: 339-368). Obecnie zaawansowanie techniczne reasonerów jest wciąż ograniczone i przeprowadzanie wnioskowań nawet na

nieskomplikowanych ontologiach trwa zbyt długo jak na uwarunkowania Internetu;

- **mechanizmy certyfikacji i zaufania** – Istnieją w tej chwili w sferze rozważań teoretycznych. Mechanizmy te pozwoliłyby na zestandaryzowanie i rozwiązanie problemów autoryzacji użytkowników, identyfikacji ich zasobów, a także określenia praw, na jakich te zasoby są przesyłane i mogą być udostępniane (Daconta M. C., Obrst L., Smith 2003).

Aktualną, częściową realizację wizji Internetu Semantycznego jako sieci połączonych danych opisanych w formacie RDF, jest inicjatywa Linked Data, która stanowi zbiór wskazówek jak publikować dane semantycznie oznaczone w sieci WWW i jak je łączyć z już istniejącymi zasobami (www.w3.org). Obecnie w projekcie bierze już udział wiele instytucji, w tym organizacje rządowe, media, wydawnictwa. Centralnym punktem ogromu połączonych danych jest baza DBPedia (Rysunek 4), czyli baza danych powstała poprzez zapis informacji z Wikipedii w formacie RDF. Dzięki udostępnieniu i połączeniu danych różnych organizacji w formacie RDF możliwe jest przeprowadzanie wyszukiwań na dużą skalę i automatyczne wnioskowanie (Lula, Stal, Tadeusiewicz, Morajda, Paliwoda-Pękosz, Wilusz 2012: 217).).



Rysunek. 4. Graf DBpedii

Źródło: https://www.w3.org/Talks/2012/1108-LOD-Rabat/lod-datasets_2011-09-19_colored.png

DBpedia (<http://wiki.dbpedia.org/>) jest projektem mającym na celu oznaczenie, ustrukturyzowanie i powiązanie danych (reprezentacja w postaci RDF) w największej interaktywnej encyklopedii – Wikipedii. Rezultatem tego działania jest utworzenie wielkiego grafu – reprezentacji wiedzy. Zasada funkcjonowania DBpedii polega na stosowaniu złożonych zapytań zwracających konkretne informacje z użyciem języka SPARQL. W przypadku klasycznej Wikipedii stosuje się jedynie wyszukiwanie na podstawie słów-kluczy.

4. Korzyści i pułapki takiej reprezentacji wiedzy • Open Source, XML i Legal Semantic Web w sądownictwie

Upowszechnienie Internetu, jako medium umożliwiającego kontakt, wymianę oraz wyszukiwanie informacji, jest obecnie jednym z głównych czynników mających wpływ na szybki rozwój gospodarczy, doskonalenie usług, rozwój nowych technologii i wymianę doświadczeń w środowiskach naukowych i biznesowych. W dobie nadmiaru informacji, mających swoje źródło nie tylko w samej organizacji, ale również poza nią, szczególne zainteresowanie budzi aspekt integracji systemów informatycznych obejmujący jednolitą definicję danych, będących przedmiotem wymiany w ramach organizacji. Rozwiązaniem powiązanych z nim problemów może być zastosowanie sieci semantycznych, których koncepcja ma swoje źródło w rozwiązaniach odnoszących się do globalnej sieci (Piasecki A., Socha 2010: 606-615).

Wśród korzyści zastosowania sieci semantycznych można wymienić:

- wyszukiwanie informacji na podstawie znaczenia haseł wpisywanych w wyszukiwarce, a nie tylko po słowach kluczowych, co często skutkuje błędnymi odpowiedziami i uniemożliwia korzystanie z synonimów,
- rozróżnianie danych na podstawie kontekstu ich znaczenia,
- ekstrakowanie informacji z różnych źródeł i ich automatyczna integracja,

- prezentowanie tylko istotnych danych z punktu widzenia użytkownika wraz z możliwością wyjaśnienia podjętych przez komputer działań/wniosków,
- przeprowadzanie wnioskowania na danych opisanych semantycznie, co skutkuje uzyskaniem nowych informacji (które nie były dotychczas jawnie zapisane), a te z kolei poddane procesom wnioskowania mogą prowadzić do uzyskania kolejnych nowych faktów itd.,
- wprowadzanie wiedzy zawartej w dokumentach danego typu, przez co ułatwiłoby to zarządzanie dokumentami oraz informacją, a także prowadziło do utrzymywania niesprzeczności i aktualności gromadzonych danych,
- automatyczną generację dokumentów opisanych semantycznie,
- automatyczną generację stron WWW opisanych semantycznie bez udziału użytkownika (wiedza służąca do opisu strony czerpana jest z semantyki zawartej w ontologii) (Bąk J., Jędrzejek 2009: 237).

Zastosowanie tego typu rozwiązań pojawiło się jako systemy wielkoskalowe, które opracowano dla administracji publicznej. Takie systemy wspomaganie pracowników i obywateli w zakresie stosowania przepisów prawnych, generują automatycznie odpowiednie wnioski (ustalenie wysokości podatków, ocena uprawnień, itd.). Systemy automatycznego zawierania umów często korzystają z własnych specyfikacji/oprogramowania (języków) reprezentujących informacje normatywne, które nie są najczęściej kompatybilne z innymi. Zastosowanie takich rozwiązań może utrudnić rozwój nowych aplikacji i uniemożliwić komunikację pomiędzy różnymi systemami (Odlanicka-Poczobutt, Kulińska 2014: 103-111).

Rozwiązaniem proponowanym jako optymalne jest wprowadzanie norm opartych na XML (ang. *Extended Markup Language*) do wyrażenia struktury dokumentu i uzupełnienia go o informacje meta-tekstowe. Tagowanie XML jest zwykle używane do umieszczenia metatekstowej informacji w dokumentach, uzupełnionej o OWL dla określania struktury pojęciowej i rozszerzeń logicznych oraz RULEXML do rejestrowania logicznej struktury przepisów, np. prawnych (Sartor 2011: 16-20). Zostały również opracowane sposoby określania odstępstw i zmian w dokumentach. Dalsze meta-informacje osadzone w tekstach dostępnych za pośrednictwem Internetu w celu wyszukiwania informacji pojęciowej

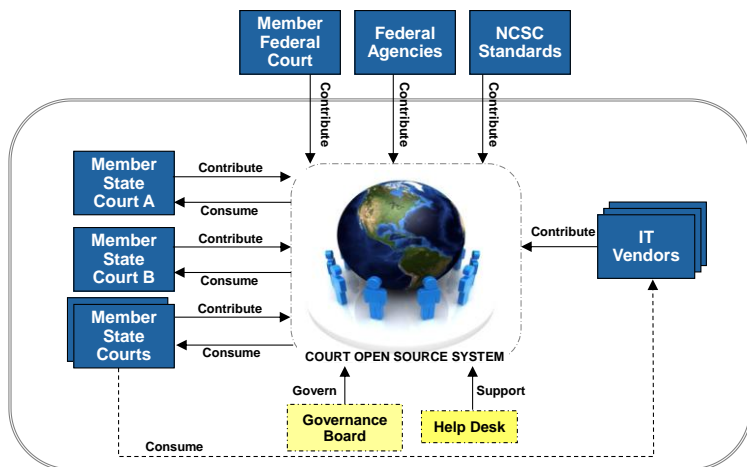
zawierają w szczególności dokumenty, które są indeksowane według analiz koncepcyjnych (ontologii) danej dziedziny. Dzięki standaryzacji możliwa stanie się swobodna wymiana danych oraz formalizacja i unifikacja dotychczasowej już zelektronizowanej wiedzy. Dzięki dobrze określonym strukturom reprezentacji będzie można w łatwiejszy sposób wyszukiwać informacje lub wnioskować w poszukiwaniu nowych faktów i powiązań.

W ciągu ostatnich kilku lat technologia informatyczna w sądownictwie przeszła wiele zmian. Obecnie sądom w wielu krajach zależy bardziej niż kiedykolwiek na aplikacjach ICT, które służyłyby do zarządzania sprawami i prowadzenia dokumentacji w celu wspierania ich codziennej działalności (Odlanicka-Poczobutt, Kulińska 2015: 541-550). Najbardziej poszukiwane są rozwiązania w obszarach takich jak e-kartoteki, zarządzanie sprawami, zarządzanie dokumentami, rejestrowanie obrazu, digitalizacja zapisu rozpraw oraz publiczny dostęp przez Internet do wszystkich informacji.

Przygotowanie oprogramowania dla sądów wiąże się jednak przede wszystkim z problemem braku ujednoczonych reguł postępowania. O ile przepisy prawne wskazują zasady postępowania w procesie rozpoznawania spraw, o tyle brak jest szczegółowych zasad postępowania dotyczących działalności administracyjnej wspierającej proces orzecznictwa. Modele proponowane w ramach rozwiązań informatycznych wymagają precyzyjnie zdefiniowanych reguł biznesowych, dotyczących działań, a w tym przypadku nie zostały one znacząco udokumentowane i nie są stosowane w sposób jednolity (Odlanicka-Poczobutt 2015: 531-540). Do tworzenia i utrzymania systemów informatycznych sądy posiadają własne aplikacje komputerowe opracowane wewnętrznie w ramach jednostek albo licencje na oprogramowanie działające na bazie własnych systemów. Często występuje połączenie tych dwóch rozwiązań. Wymienione opcje posiadają jednak istotne wady. Własne oprogramowanie lub in-sourcing często prowadzą do niemożności utrzymania sądowych systemów IT na odpowiednim poziomie, jeżeli chodzi o bieżącą aktualizację lub szybką odpowiedź na nowe wymagania funkcjonalne i proces technologicznego starzenia oprogramowania.

Każdego roku do sądów wpływają miliony spraw, które generują około miliarda dokumentów. Występujące problemy dotyczące przetwarzania i indeksowania, stają się głównym powodem zwrócenia się w kierunku zintegrowanych systemów elektronicznych

w celu zmniejszenia ilości dokumentów oraz usprawnienia codziennych operacji (Beard 2004). Przykładem dążenia do ułatwienia elektronicznej rejestracji dokumentów jest zastosowanie rozwiązań Open Source. Open Source jest modelem pozyskiwania aplikacji wykorzystywanych do wspólnego rozwijania i utrzymywania ich w obszarach, gdzie samodzielne działania byłyby zbyt drogie. Dotyczy to głównie środowiska, w którym ogólnie dostępne produkty oprogramowania nie pasują dokładnie do wymagań użytkowników (Rebo, Roper, Harvey 2009). Przykład ekosystemu Court Open Source zaproponowanego dla sądownictwa amerykańskiego przez firmę Gartner przedstawiono na rysunku 5.



Rysunek 5. Court Open Source - Ecosystem

Źródło: opracowanie Understanding the Open-Source Ecosystem, Gartner 2008, <https://www.gartner.com>. [12.2015]

Współpraca w ramach takiego systemu między sądami w całym kraju powinna zapewniać niezawodną sieć wsparcia, repozytorium na bazie solidnych i elastycznych komponentów oprogramowania i artefaktów, zastosowanie aktualnych i sprawdzonych najlepszych praktyk oraz szybkie tworzenie aplikacji z wykorzystaniem udostępnionych składników oprogramowania (Tracy, Guevara, Stegman 2008). Należałoby utworzyć i utrzymywać repozytorium składające się z solidnych, modyfikowalnych i elastycznych komponentów technicznych oprogramowania, do których można łatwo uzyskać dostęp. Wymaga to jednak wspierania

krajowych standardów i najlepszych praktyk w zakresie analizy procesowej, rozwoju oprogramowania i rozwoju infrastruktury systemów. Istotne znaczenie ma tutaj również stworzenie odpowiedniego języka do zapisu sposobu postępowania, zgodnego z odpowiednią jakością przekładu prawniczego, uwzględniającego cechy języka prawa (Matulewska 2008: 53–63), na przykład na podobieństwo DBpedii. Zasadą i warunkiem jest współpraca i współdzielenie rozszerzonych zasobów, które mogą być używane do budowy wspólnych rozwiązań technologicznych dla sądów. Należałoby wspólnie zdefiniować pojęcia, rozwijać je i dzielić się wiedzą, tworząc elementy komponentów oprogramowania do wielokrotnego użytku na podstawie najlepszych praktyk i doświadczeń. W przypadku braku możliwości pokonania tych barier współpraca systemów stanie się nieefektywna, będzie ograniczona i może nie spełniać formalnych wymogów. W ramach przepływu danych wymagana jest standaryzacja. Wymienione rozwiązania wymagają realizacji wdrożenia krok po kroku, z uwzględnieniem specyfiki organizacji, jaką jest sądownictwo powszechne.

W krajach, gdzie system sądowniczy jest w tym zakresie znacznie bardziej rozwinięty występują już semantyczne sieci prawne (ang. Legal Semantic Web), które tworzone są przez prawne treści dostępne w Internecie, które są wzbogacone o komputerowe przetwarzanie informacji. Semantycznie wzbogacone dokumenty mogą być automatycznie dopasowane do potrzeb ich użytkownika na podstawie automatycznego przetwarzania informacji, które zawierają (na przykład treści legislacyjne mogą być prezentowane w preferowanym formacie, można będzie je wizualizować w wersji będącej obecnie w użyciu, albo w poprzedniej wersji; widoczne są terminy obowiązywania – kiedy akty prawne weszły w życie, jakie są następstwa ich niedotrzymania, itp.).

Zaprezentowana tendencja do korzystania z sieci semantycznych jest szansą dla prawodawstwa, ułatwiająca procesy opracowania i udostępniania przepisów prawnych, utrzymanie odpowiedniej struktury źródeł informacji prawnej i zarządzania postępem prac legislacyjnych i realizacji procedur. Pozwala również na zwiększenie interakcji z podmiotami zewnętrznymi (jawność procedur i informacji, komunikacja z obywatelami, dialog z instytucjami krajowymi i międzynarodowymi). Określenie odpowiednich standardów dla dokumentów może także zapewnić efektywne relacje pomiędzy tworzeniem przepisów prawnych oraz ich

upowszechnianiem i stosowaniem w środowisku prawniczym i gospodarczym.

Ważnym elementem jest również integracja teorii prawa i informatyki prawniczej. Dane dostarczane przez teorię prawa w odniesieniu do aspektów, takich jak logika prawna, teoria norm i stanowisk normatywnych, modele argumentacji prawnej, są tłumaczone na język modeli obliczeniowych. Zadaniem informatyki prawniczej jest zapewnienie nowych modeli argumentacji prawnej do teorii prawa, biorąc pod uwagę osiągnięcia w nauce poznawczej i sztucznej inteligencji. Poprzez integrację tych dwóch dziedzin badań możliwe jest uzyskanie wartościowych wyników, takich jak lepsze zrozumienie prawa, lepsza specyfikacja informatycznych systemów prawnych, krytyczne podejście do informatyzacji prawa, identyfikacja nowych sposobów przetwarzania informacji prawnej (Wierczyński 2006). Ma to szczególne znaczenie w odniesieniu do standardów dotyczących informacji prawnej, które muszą być dostosowane do rodzaju prawa oraz do potrzeb instytucji i obywateli, tak, aby były racjonalne i możliwe do zaakceptowania zarówno przez dostawców jak i odbiorców informacji prawnej.

Informatyka prawnicza nie ogranicza się jedynie do badań i rozwoju aplikacji komputerowych, ale obejmuje także koordynowanie działań aplikacji i systemów już istniejących w systemach prawnych i poza nimi (w różnych działach administracji publicznej oraz w sektorze prywatnym i organizacjach non-profit) (Wiewiórowski, Wierczyński 2012). Jest to tym bardziej istotne w dobie usług internetowych, aby system oprogramowania zaprojektowany był do obsługi interakcji pomiędzy urządzeniami w sieci, zgodnie z definicją zawartą w W3C Web Services Glossary (Web Services Glossary). Celem W3C (ang. World Wide Web Consortium) jest stworzenie jednej ogólnosiwiatowej sieci, będącej otwartą platformą, posiadającą dużą ilość informacji, i gdzie usługi mogą być powszechnie dostępne (Sartor 2011).

5. Podsumowanie

Sieć semantyczną najprościej jest przedstawić jako rozwinięcie Internetu o zrozumienie intencji użytkownika, na podstawie haseł

i czynności, które w niej wykonuje. Agent (maszyna) po stronie użytkownika, jest w stanie porozumieć się z agentem po stronie innego klienta i uzyskać na tej podstawie konkretne informacje związane z zapytaniem. Autorem wizji sieci semantycznej jest Tim Berners-Lee (twórca World Wide Web). Wymaganymi standardami sieci semantycznej są sposoby opisu zasobów sieciowych, aby były one zrozumiałe dla silników wnioskujących (reasonerów). Wynikiem tworzenia sieci semantycznej jest ustrukturyzowanie zasobów i połączenie ich w logiczną sieć. Najistotniejszą informacją dotyczącą sieci semantycznej jest to, że przy zachowaniu takich standardów, można prowadzić wnioskowanie. Umożliwia to zrozumienie maszynowe treści przechowywanych w sieci, co pozwala na tworzenie olbrzymiej, dynamicznej bazy wiedzy.

Zastosowanie sieci semantycznej pozwoli pokonać granice obecnej sieci, ułatwi dostęp do informacji, jak również działania w Internecie (od e-commerce do e-administracji). Udostępnienie maszynom możliwości operowania na poziomie semantyki umożliwi stworzenie sieci, która zaoferuje nowy poziom usług. Powstanie ogromnej sieci wiedzy, wzbogaconej o możliwość inteligentnego przetwarzania informacji przez maszyny wymaga intensywnej współpracy interdyscyplinarnej na etapie tworzenia meta-danych. Bardzo ważną rolę do spełnienia będą mieli językoznawcy, którzy przy współpracy ekspertów z wyspecjalizowanych dziedzin będą tworzyć rzeczywiste opisy kategorii pojęć, aby uchronić się przed błędami tłumaczenia powierzchniowego, wynikającymi z niedostatecznej analizy tekstu źródłowego i braku umiejętności wyszukiwania ekwiwalentów translacyjnych.

Sieci semantyczne mogą być przydanym narzędziem dla różnych organizacji, jednak sposób ich wykorzystania obecnie jest dopiero stopniowo definiowany i precyzowany. Mimo sukcesów w wykorzystaniu sieci semantycznych należy pamiętać, że jest to dziedzina wymagająca dalszych prac badawczych, a w realizacji związanych z nimi projektów niezbędny jest udział jednostek naukowych, posiadających doświadczenie w definiowaniu i stosowaniu ontologii.

Bibliografia

- Alemu, G., Stevens, B. i Ross, P. 2012. Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: a social constructivist approach. *New Library World*, Vol. 113 Nos 1/2, ss. 38-54.
- Bąk J., Jędrzejek Cz. 2009. Semantic Web – technologie. XV Konferencja PLOUG, Kościelisko 2009. Politechnika Poznańska. s. 237.
- Beard J. 2004. An Open-Source System for Electronic Court Filing, *Linux Journal*, Issue #122/June 2004.
- Berners Lee T. Linked Data.[Online] [12. 2015] <https://www.w3.org/DesignIssues/LinkedData.html>.
- Berners Lee T., Hendler J., Lassila O. 2001. Sieć Semantyczna, *Świat Nauki*. ss. 41-49.
- Cloud network architecture and ICT – Modern Network Architecture. 2013. itknowledgeexchange.techtarget.com.
- Daconta M. C., Obrst L. J., Smith K. T. 2003. *The Semantic Web: A guide to the Future of XML, Web Services and Knowledge Management*. John Wiley&Sons.
- Diekema, A.R. 2012. Multilinguality in the digital library: a review. *The Electronic Library*, Vol. 30 No. 2, pp. 165-181.
- Elenius D., Denker G., Stehr M.-O. 2008. A Semantic Web Reasoner for Rules, Equations and Constraints, *Web Reasoning & Rule System*. ss. 135-149.
- <https://encyklopedia.pwn.pl/haslo/ontologia;3951174.html> [11.2017].
- <https://sjp.pwn.pl/slowniki/semantyczny.html>
- <https://www.gartner.com>. [12.2015]
- https://www.w3.org/Talks/2012/1108-LOD-Rabat/lod-datasets_2011-09-19_colored.png
- Kruk S., Senior Technical Product Manager at Dynatrace, Prezentacja Sieć Semantyczna w teorii i praktyce, <https://www.slideshare.net/skruk/sie-semantyczna-w-teorii-i-praktyce> [11.2017]
- Lula P., Stal J., Tadeusiewicz R., Morajda J., Paliwoda-Pękosz G., Wilusz W. 2012. Komputerowe metody analizy i przetwarzania danych. Uniwersytet Ekonomiczny w Krakowie. Kraków. s. 193.

- Matulewska A. 2008. Jakość przekładu prawniczego a cechy języka prawa [w:] *Język, Komunikacja, Informacja*, P. Nowak, P. Nowakowski (red.) 3/2008. ss. 53-63.
- Mishra R. B., Kumar S. 2011. Semantic web reasoners and languages, *The Artificial Intelligence Review*. Springer. ss. 339-368.
- Nęcka E., Orzechowski J., Szymura B. 2006. *Psychologia poznawcza*. Warszawa: PWN.
- Odlanicka-Poczobutt M. 2015. Court Technology jako przykład zastosowania ICT w sądownictwie, *Zeszyty Naukowe Uniwersytetu Szczecińskiego, Ekonomiczne Problemy Usług* 2015 nr 117. ss. 531-540.
- Odlanicka-Poczobutt M., Kulińska E. 2015. Elektronicznie systemy zarządzania sprawami jako przykład dobrych praktyk. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług* 2015 nr 117. ss. 541-550.
- Odlanicka-Poczobutt M., Kulińska E. 2014. Kierunki rozwoju informacji prawnej w ramach gospodarki elektronicznej, *Zesz. Nauk. Uniwersytetu Szczecińskiego, Ekonomiczne Problemy Usług* 2014 nr 113. ss. 103–111.
- Odlanicka-Poczobutt M. 2014. Rola technologii informacyjno-komunikacyjnych (ICT) w sądownictwie powszechnym – wyzwania i możliwości, [w] *Ekonomiczno-społeczne i techniczne wartości w gospodarce opartej na wiedzy*. Buko J., (red.). *Zeszyty Naukowe Uniwersytetu Szczecińskiego* nr 809. *Ekonomiczne Problemy Usług* nr 113. Tom II. Szczecin. ss. 93–101.
- Odlanicka-Poczobutt M. 2013. Zastosowanie nowoczesnych technik informacyjnych w sądownictwie powszechnym, [w] *Europejska przestrzeń komunikacji elektronicznej*. Buko J., (red.), *Zeszyty naukowe Uniwersytetu Szczecińskiego* nr 763, *Ekonomiczne Problemy Usług* nr 105, Tom II, Szczecin. ss. 553–561.
- Piasecki A., Socha M. 2010. Możliwości praktycznego zastosowania sieci semantycznych w przemyśle, *Pomiary Automatyka Robotyka* 2/2010, s. 606–615.
- Radar Networks & Nova Spivack. 2007 <https://www.radarnetworks.com> [11.2017].
- Rebo J., Roper B., Harvey T., Court Technology Conference 2009, Sept 23, 2009, Rooms 702-706, Denver 2009, Colorado.

- Rudman R., Bruwer R. 2016. Defining Web 3.0: opportunities and challenges, *Electronic Library*, vol. 34. Emerald Group Publishing Limited, pp. 132 - 154.
- Sallai G. 2012. Defining Infocommunications and Related Terms, *Acta Polytechnica Hungarica*, Vol. 9, No. 6. ss. 5–15.
- Sartor G. 2011. Chapter 2. Legislative Information and the Web [w:] *Legislative Xml For The Semantic Web. Principles, Models, Standards for Document Management*, Sartor G., Palmirani M., Francesconi E., Biasiotti M. A., (Ed.), Springer - Law, Governance and Technology Series, Volume 4. ss. 16–20.
- Semantyczna Sieć: Boty uczą się kojarzyć. <https://www.chip.pl> [11.2017].
- Sieci Semantyczne – WWW następnej generacji, <https://4programmers.net> [11.2017].
- Silverstone R. et al. 1991. Listening to a long conversation: an ethnographic approach to the study of information and communication technologies in the home, *Cultural Studies*, 5(2). ss. 204–227.
- Tracy L., Guevara J. K., Stegman E., *IT Key Metrics Data 2009: Key Applications Measure: Life Cycle Distribution: Current Year*, 15 December 2008, ID G00163849.
- W3C. 2004. *Web Services Glossary*, W3C Working Group Note 11 February 2004.
- Wierczyński G. 2006. 15 lat Systemu Informacji Prawnej LEX, (w:) *Informacja prawna a prawa obywatela. Konferencja z okazji XXXV-lecia informatyki prawniczej w Polsce i XV-lecia Systemu Informacji Prawnej LEX*, Gdańsk, 19-20 czerwca 2006 r., Sopot 2006.
- Wiewiórowski W., Wierczyński G. 2012. *Informatyka prawnicza. Nowoczesne technologie informacyjne w pracy prawników I administracji publicznej*, wyd. 3. Warszawa.