

Michał T. Szczepański, Tomasz Chmielak

Muzeum Historii Polski

BazHum: baza zawartości czasopism humanistycznych. Dokonania i perspektywy

Abstrakt

Celem projektu BazHum jest utworzenie bazy zawartości polskich czasopism humanistycznych. Zamierzamy zarejestrować opisy artykułów z możliwie szerokiej liczby czasopism. Prowadzimy globalną kwerendę tytułów. W kręgu naszych zainteresowań są periodyki ukazujące się od połowy XIX wieku, opracowujemy je od pierwszego numeru. Obecnie istnieje możliwość prostego przeszukiwania bazy za pomocą słów z tytułów oraz nazwisk i nazw autorów. Pracujemy nad rozbudową możliwości wyszukiwania, np. poprzez słowa kluczowe, po tekstach abstraktów, za pomocą operatorów logicznych. Zespół projektowy ma strukturę modułową: obsługa projektu, bibliotekarze i katalogerzy. Od wielu miesięcy współpracujemy z kilkoma instytucjami zewnętrznymi, przede wszystkim z bibliotekami PAN. Otrzymujemy wiele sygnałów, szczególnie ze strony pracowników naukowych, na temat użyteczności bazy BazHum. Zaczyna być ona wykorzystywana przez biblioteki przy układaniu bibliografii poszczególnych autorów lub wykonywaniu indeksów do zawartości czasopism na przestrzeni wielu roczników. Opracowaliśmy zestaw narzędzi do gromadzenia i edycji danych. Zaadaptowaliśmy standardy opisu obowiązujące w nowoczesnych bazach bibliograficznych. Dostępny jest i przeszukiwalny on-line zbiór ok. 50 tys. rekordów bibliograficznych (w wersji uproszczonej) w ramach kolekcji na platformie YADDA.

Słowa kluczowe

baza bibliograficzna, baza pełnotekstowa, bibliografia zawartości czasopism, polskie czasopisma naukowe, zawartość czasopism on-line, nauki humanistyczne i społeczne, elektroniczne repozytorium danych, artykuły

Abstract

The goal of the project Bazhum is to create a database of contents of Polish humanities journals. We plan to upload the list of contents from a wide spectrum of journals. We are currently researching the titles. In the field of our interest are all journals since the mid-19th century, they are being worked on from the first issue. It is already possible to do a basic search of the database by the names of authors and titles of articles. We are currently developing the search options to make it possible to do an advanced search by keywords, fragments of abstracts, Boolean operators etc. The project team has a modular structure: service, librarians and cataloguers. For many months we have been working with external institutions, like the libraries of Polish Academy of Science (PAN) and others. We receive a lot of positive feedback from the world of academia, regarding the usefulness of the database Bazhum. It starts even to be used by libraries in creating bibliographies of individual authors and in creating the indexes of contents for journals over a span of many years. We have developed a group of tools for editing and data collection. We have adopted the standards of bibliographic description in modern bibliographic databases. A collection of approx. 50.000 simplified bibliographic records is accessible on the YADDA platform.

Keywords

bibliographical database, full-text database, contents of academic journals, polish academic journals, periodical content online, humanities and social sciences, digital repository, articles

Koncepcja i założenia projektowe

Projekt BazHum ma na celu zbudowanie bibliograficzno-abstraktowej, a następnie pełnotekstowej bazy zawartości polskich czasopism naukowych z zakresu nauk humanistycznych i społecznych. Zamierzamy umieścić w niej dane dotyczące wszelkiego rodzaju publikacji ogłoszonych na łamach czasopism, a więc artykułów, artykułów recenzyjnych, recenzji, zapisek, miscellaneów, nekrologów itp. Zawartość periodyków jest opracowywana w ujęciu retrospektywnym, począwszy od pierwszych opublikowanych numerów aż po zeszyty najnowsze i bieżące.

BazHum powstaje jako narzędzie służące pracy naukowej i dydaktycznej. Sądzymy, że nowoczesne opracowanie i udostępnienie osiągnięć polskiej humanistyki przyczyni się do zdynamizowania badań i postępów w przekazywaniu wiedzy. Korzyścią odczuwaną najbardziej bezpośrednio będzie oczywiste przyspieszenie procesu poszukiwań bibliograficznych. Samo pojęcie i związana z tym praktyka kwerendy ulegnie przekształceniu. BazHum umożliwi poświęcenie większej uwagi czynności definiowania i stawiania problemów badawczych, wzmocni intelektualny aspekt poszukiwania, tworzenia i przyswajania wiedzy.

W ramach BazHum-u zostanie zarejestrowana, opracowana i udostępniona ogromna część historycznego i współczesnego dorobku polskiej nauki. Opracowanie tego materiału ma wartość samą w sobie. Kolekcja przechowująca w sposób zorganizowany i strukturalnie uporządkowany źródła wiedzy i refleksji naukowej jest nie tylko ich nośnikiem, staje się ich wtórnym źródłem o znaczeniu zarówno w kontekście archiwizacyjnym, jak i naukowym. BazHum będzie bazą informacji i wiedzy. Dopiero zbudowanie takiego zasobu umożliwi podejmowanie wysiłków na rzecz wzmocnienia społecznej świadomości jej dorobku, jak również jego popularyzację i promocję.

BazHum jest przedsięwzięciem badawczym, opracowanym i realizowanym przez Muzeum Historii Polski. Prace pilotażowe przeprowadzono w 2006 r. W następnym roku podjęto współpracę z ICM, co pozwoliło na wykorzystanie doświadczeń i narzędzi powstałych w ramach projektu BazTech. W pierwotnym zamierzeniu BazHum był przeniesieniem konceptu bazy bibliograficzno-abstraktowej czasopism technicznych na dziedzinę nauk humanistycznych i społecznych. Opracowanie specyficznego zasobu i odmienna organizacja pracy spowodowały odejście od tego modelu, jakkolwiek w dalszym ciągu BazHum zachowuje wyraźne pokrewieństwo do swego wzoru.

BazHum jest w całości realizowany ze środków muzeum przeznaczonych na działania merytoryczno-projektowe, a więc ze środków pochodzących z budżetu publicznego. W celu utrzymania planowanej dynamiki rozwoju projektu obecnie podejmujemy starania o zróżnicowanie i rozszerzenie bazy finansowej. Dywersyfikacja źródeł finansowania jest nieodzowna, nie tylko ze względu na globalną redukcję wydatków z kasy publicznej, lecz także z uwagi na inwestycje związane z budową gmachu muzeum.

Projekt powstaje we współpracy z Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, które utrzymuje platformę YADDA i jest odpowiedzialne za zbudowanie narzędzi do gromadzenia danych. Naszymi partnerami merytorycznymi są biblioteki Instytutu Sztuki PAN oraz Instytutu Archeologii i Etnologii PAN. Baza jest udostępniana za pomocą platformy YADDA pod adresem

<http://bazhum.icm.edu.pl>[1]. Zawartość można przeszukiwać również w ramach kolekcji MHP pod adresem <http://yadda.icm.edu.pl>.

W projekcie BazHum przewiduje się opracowanie możliwie szerokiego wyboru czasopism naukowych z dziedziny humanistycznej i społecznej. Termin „naukowy” definiujemy w zgodzie ze współczesnym publikacji, niekiedy już być może nieaktualnym sensem tego pojęcia. Dziedzinę nauk humanistycznych i społecznych wyznaczają obecnie akceptowane znaczenia tych pojęć. Można też powiedzieć, że jest to obszar wiedzy i badań, który wyłączają ze swych zainteresowań nauki o precyzyjnie określonym polu badawczym (np. nauki przyrodnicze, matematyczne, stosowane itp.).

Do kolekcji zamierzamy włączyć periodyki ukazujące się na ziemiach polskich lub w granicach państwa polskiego bądź też wytworzonych przez polskie instytucje za granicą w okresie od połowy XIX wieku do czasów najnowszych. W ramy klasyfikacji włączamy tytuły publikowane w obecnych lub dawnych granicach kraju przez instytucje nie-polskie lub nieposiadające wyraźnej identyfikacji państwowej i narodowej. Jako ziemie polskie traktujemy obszary wchodzące w skład dawnej Rzeczypospolitej, które zostały wcielone do sąsiednich państw w wyniku rozbiorów i później wchodziły lub nie w skład różnych państw i ciał administracyjnych powstających na jej terytorium. Należy dodać, że przyjmujemy dwa stopnie identyfikacji docelowego zasobu BazHum-u. Ogólne ramy klasyfikacji, o których była mowa, służą do opracowania globalnego rejestru czasopism, który następnie zawężamy do tytułów przeznaczonych do bieżącego opracowania. W tym celu posługujemy się dodatkowymi kryteriami selekcji, jak np. dostępność czasopisma w zasobach publicznych i ogólnie dostępnych, jego powaga i naukowa użyteczność, szczególne zainteresowanie opracowaniem zawartości ze strony instytucji kooperującej w projekcie.

Jednym z istotnych priorytetów kształtujących koncepcję bazy BazHum była jej praktyczna użyteczność do pracy badawczej i dydaktycznej. Uprzywilejowanym punktem widzenia są zatem potrzeby użytkownika końcowego, a jego praktyczne doświadczenie jest ostatecznym kryterium oceny dla poprawności przyjętych założeń, zwłaszcza jeśli chodzi o zawartość kolekcji, szczegółowość informacji bibliograficznych, możliwości wyszukiwania danych itp. W znacznej mierze są to kryteria możliwe do zrealizowania dopiero po dłuższym okresie pracy, które w obecnej chwili traktujemy jako wytyczne.

Drugim priorytetem dla BazHum-u było tworzenie takich struktur danych, które utrzymają zdolność do wymiany i ewentualnej konwersji do innych zewnętrznych baz. W tej sytuacji uznano konieczność odwołania się do powszechnie przyjętych standardów opisu bibliograficznego, określiły one wyjściowy model formatu danych. W ramach prac nad projektem podjęto próbę zharmonizowania obu punktów widzenia, w momentach niepewności skłaniając się ku preferowaniu priorytetu dotyczącego standardu metadanych jako bardziej obiektywnego.

Kierunki działania

BazHum jest obecnie bazą bibliograficzną, która w przyszłości zostanie poszerzona o teksty artykułów. W ograniczony sposób planuje się natomiast publikować w niej teksty abstraktów. Streszczenie abstraktowe lub innego typu jest wciąż jeszcze na tyle wyjątkowym elementem publikacji z dziedziny humanistycznej i społecznej, że trudno uczynić je przedmiotem systematycznego opracowania. Ponadto praca nad materiałem retrospektywnym, którego elektroniczną postacią nie dysponujemy w pierwszej fazie opracowania, wymagałaby

wykonania specjalnych czynności na tekście, co niekorzystnie zwiększa proporcje nakładów sił i środków do uzyskanych efektów.

Realizacja projektu rozwija się w trzech kierunkach: 1) opracowanie danych bibliograficznych, 2) opracowanie i publikacja zawartości pełnotekstowej oraz 3) rozwój interfejsu i wewnętrznych narzędzi bazy służących efektywnej pracy użytkownika.

1. Pierwszy kierunek polega w szczególności na opisaniu i zgromadzeniu danych bibliograficznych z wybranego zestawu czasopism, przy czym zależy nam na rejestracji wszystkich jednostek tekstowych w czasopiśmie, przeprowadzeniu globalnej kwerendy polskich czasopism humanistycznych i społecznych, sklasyfikowaniu czasopism do opracowania w węższych ramach konkretnego planu pracy rocznej, przeprowadzeniu kwerendy zawartości czasopism już opracowanych (metadane odpowiadające normom opisu bibliograficznego zgromadzone w kolekcji posiadającej teoretyczną możliwość transferowania lub przyłączenia do bazy BazHum) lub udostępnionych w wersji pełnotekstowej (teoretyczna możliwość transferu lub zbudowania systemu odniesień z lub do BazHum-u).
2. Praca nad pełnotekstową zawartością bazy polega na wypracowaniu zasad i procedur dotyczących skanowania, a następnie zastosowania techniki OCR wobec pozyskanego materiału, skonstruowaniu odpowiadającym im procesów technologicznych w ramach ściśle określonych warunków, przy wykorzystaniu zewnętrznych zasobów bibliotekarskich, sprzętowych i do pewnego stopnia również technicznych, określeniu zasad publikowania materiału elektronicznego w wersji dostępnej oraz częściowo lub całkowicie zabezpieczonej, przeprowadzeniu konsultacji z wydawcami dotyczących warunków udostępniania i zastosowania odpowiedniego okresu karencji.
3. Kierunek działań dotyczący rozwoju narzędzi służących zwiększeniu wydajności kwerend bibliograficznych i innej pracy z bazą nie jest obecnie realizowany. Uruchomienie działań planujemy w czwartym kwartale tego roku. Będą się one skupiały na opracowaniu efektywnych zasad poruszania się po interfejsie, opracowaniu konsekwentnych logicznie metod wyszukiwania danych w bazie na poziomie użytkownika zewnętrznego, rozbudowaniu funkcji wyszukiwania m.in. adaptacji operatorów logicznych itp.

W najbardziej zaawansowanym stadium realizacji jest obecnie opracowanie danych bibliograficznych, a zwłaszcza ta jego część, która dotyczy pozyskiwania i gromadzenia opisów bibliograficznych zawartości czasopism. Działania są zorganizowane na zasadzie pracy grupowej, w wysokim stopniu oparte na systemie procedur. Składają się z trzech powiązanych sekwencji czynności: 1) rejestracja i gromadzenie danych bibliograficznych, 2) kontrola poprawności i korekta, 3) scalanie danych i ich publikacja.

Standard danych

Opracowuje się wszystkie artykuły z czasopisma, także artykuły odredakcyjne, recenzyjne, kroniki, nekrologi itp. Zdecydowaliśmy się na zbieranie prawie wszystkich dostępnych danych. Są to: tytuł główny; tytuły równoległe (tzn. również z obcojęzycznych spisów treści lub obcojęzycznych streszczeń – jeżeli czasopismo zdecydowało się przetłumaczyć tytuł główny); tytuły działów, w których został umieszczony artykuł; autorzy i inni twórcy odpowiedzialni za artykuł (np. tłumacze, osoby opracowujące artykuły); język tekstu oraz języki ewentualnych streszczeń; opis fizyczny: strony liczbowane i nieliczbowane, materiał ilustracyjny, bibliografie załącznikowe oraz informacje o streszczeniach.

Program edycyjny posiada więcej pól, np. na tekst abstraktu, dane autora. Obecnie pola te są pomijane w trakcie opracowywania rekordów. Dane ułożone są hierarchicznie. Wymogi programu spowodowały, że podstawą tej hierarchii jest wydawnictwo, ale doraźnie rozwiązaliśmy ten problem, zastępując nazwy wydawnictw nazwą naszej kolekcji. Następnie dane ułożone są już standardowo: tytuł czasopisma, rok, tom, numery i same artykuły, ułożone obecnie alfabetycznie. Zgłosiliśmy postulat do twórców oprogramowania o wprowadzenie zasady ułożenia artykułów według numeracji stron.

Tab. 1. Pola opisu rekordów bazy danych BazHum tworzonych w programie DeskLight

Pole DeskLight	Uwagi
Tytuł	Tytuł zapisywany według brzmienia „karty tytułowej” artykułu. Przykład: O granicach Mazowsza : w odpowiedzi na recenzję / J. Kowalski ; przeł. z łac. B. Zieliński.
Nowy tytuł	Tytuł lub tytuły z obcojęzycznych spisów treści lub obcojęzycznych streszczeń. Przykład: De Mazovia limitates.
Nowy tytuł	Tytuł działu, w którym umieszczony jest artykuł. Przykład: Polemiki.
Języki dokumentu	Pierwszym symbolem oznacza się język tekstu, następnymi języki ewentualnych streszczeń. Przykład: PL, LA.
Twórca	W miarę możliwości podaje się autorów w pełnym brzmieniu imion i nazwisk. Przykład: Jan Kowalski (Autor), B. Zieliński (Tłumacz).
Opis bibliograficzny	Opis fizyczny. Przykład: s. 10–20, 45, [1] k. tabl. złoż. : il. – Bibliogr. s. 19–20; Streszcz. łac.

Źródło: opracowanie własne.

Największym jednak problemem, na jaki natknęliśmy się, są tytuły w alfabetach niełacińskich. Zdecydowaliśmy się wprowadzać je w oryginalnym zapisie graficznym z dwóch powodów: po pierwsze, większość katalogerów nie zna tych języków (chodzi głównie o rosyjski, ukraiński i grecki), więc nie są w stanie ich przetransliterować, po wtóre, projekt musiałby przyjąć jedną z norm transliteracyjnych. Zdecydowaliśmy się więc na rozwiązanie informatyczne tego problemu, tzn. ustawienie narzędzi automatycznie, aby transliterowały według potrzebnych nam norm. Daje nam to jeszcze jedną dodatkową korzyść, przy zmianach zapisu transliteracyjnego będziemy w stanie automatycznie zmienić wszystkie dane, w tym zapis tytułów.

Innym ważnym problemem, na który dotychczas nie znaleźliśmy odpowiedzi, jest stworzenie dodatkowej wyszukiwarki, która nie byłaby zależna od słów zawartych w tytułach, które ostatnimi czasy stają się coraz bardziej abstrakcyjne. Oczywiście można by użyć stosowanych metod, np. haseł przedmiotowych, ale wymagałoby to dużych nakładów pracy i środków, i oczywiście zatrudnienia wyłącznie bibliotekarzy z dużym doświadczeniem. Wyjściem może być stosowanie słownika słów kluczowych, który wydaje się przyjaźniejszy zwykłemu użytkownikowi i jest łatwiejszy do opanowania.

Chcielibyśmy również, aby użytkownicy poprzez formularz dostępny bezpośrednio przy artykułach zgłaszali własne propozycje słów kluczowych. Innym rozważanym wyjściem jest metoda zastosowana przy opracowywaniu Biblioteki Wirtualnej Nauki. Polega ona na

zeskanowaniu potrzebnych tekstów i poddaniu ich procesowi OCR. Wyszukiwarka opracowanych w ten sposób publikacji jest w stanie znaleźć i podać w kontekście każde poszukiwane słowo.

Baza, jak na razie, nie posiada kartoteki dla katalogu autorskiego. Z jednej strony ustalenie szczegółowych informacji dla autora artykułu jest o wiele trudniejsze niż w przypadku autorów wydawnictw zwartych. Z drugiej wydaje nam się sensowniejsze nawiązanie współpracy z innymi podobnymi przedsięwzięciami niż tworzenie kolejnej bazy danych. Większość naszych autorów napisała również książki i ich kartoteki już funkcjonują w ogólnodostępnych informatorach.

Praktyka

Projekt BazHum jest realizowany przez zespół składający się z pracowników muzeum i bibliotek warszawskich instytutów naukowych oraz przez katalogerów, rekrutowanych głównie spośród studentów i doktorantów wydziałów humanistycznych. Organizacja pracy ma charakter modułowy: pierwszą grupę stanowią katalogerzy, którzy są rozliczani z liczby przesłanych rekordów bez narzuconych limitów i terminów. Ich zadaniem jest wprowadzenie jak największej liczby rekordów ściśle według zasad przygotowanej dla nich „Instrukcji”.

Katalogerzy rozpoczynają prace od szkolenia, na którym omówiona zostaje instrukcja, ze szczególnym uwypukleniem elementów nieoczywistych dla osób bez praktyki bibliograficznej (kropki, spacje, opis stron spoza paginacji itp.). Instrukcja zawiera także przykłady praktycznych zapisów omawianych danych. Katalogerzy pogrupowani są w zespoły (grupy katalogerskie) kierowane przez osoby z doświadczeniem w pracach bibliograficznych, tak zwanych kontrolerów czy po prostu bibliotekarzy.

Bibliotekarze weryfikują otrzymane dane i przekazują informacje o liczbie zgromadzonych opisów osobom zajmującym się scaleniem danych i obliczeniem należnego wynagrodzenia. Mają prawo odmowy zaakceptowania przesłanych im danych bibliograficznych, jeśli poziom błędów jest wysoki. Odpowiedzialność za merytoryczną i formalną poprawność zaakceptowanego rekordu spoczywa na bibliotekarzu. Z przydzielonej im puli czasopism konkretne tytuły są przydzielane imiennie poszczególnym katalogerom. Pominąwszy kontrolę danych bibliograficznych, najbardziej pracochłonnym zadaniem bibliotekarzy jest doszkalanie katalogerów, rozwiązywanie przypadków wątpliwych lub niedających ująć się w karby „Instrukcji”.

Zalecane jest, aby dyskusja nad takimi przypadkami odbywała się w ogólnodostępnej części projektowego forum. Dzięki temu inni katalogerzy, natrafiwszy na podobne problemy, mają podane gotowe rozwiązanie, a grono kontrolerów może zgłosić zapis odpowiedniej poprawki do „Instrukcji”. Plik z danymi z opracowanego czasopisma przesyłany jest na serwer projektu, gdzie jest poddawany informatycznej obróbce i umieszczany w bazie internetowej.

O ile to możliwe, czynności są określane przez zespół reguł i procedur, spisanych i podanych do wiadomości wszystkim zainteresowanym współpracownikom. Utrzymanie systemu w działaniu, choć na początkowym etapie projektu wymagało wysiłku i wciąż potrzebuje okresowych aktualizacji, pozwoliło na stosunkowo precyzyjną regulację tego obszaru projektu, który potrzebuje stabilnych i przejrzystych zasad. To z kolei umożliwiło kontrolowanie wysiłku wkładanego w wyznaczone zadania (do pewnego stopnia) oraz wymierną ocenę osiągniętych efektów. W momencie zaangażowania sporego zespołu

współpracowników, działającego w rozproszeniu, były to momenty priorytetowe z organizacyjnego punktu widzenia.

Dzięki naciskowi na proceduralizację pewnej części projektu można było wyodrębnić strefy potrzebujące istotnych i kreatywnych decyzji. Zachowano również bezpośrednią kontrolę nad intensywnością prac projektowych. Gdy okazało się, że z powodów finansowych, trzeba będzie zwolnić tempo działań, w ciągu trzech tygodni byliśmy w stanie przebudować zasady funkcjonowania przedsięwzięcia, wdrożyć zmiany i kontynuować pracę.

Proces opracowania danych bibliograficznych rozdzielono na dwa etapy. Pierwszy z nich – wpisywanie danych do bazy – nie wymaga nakładu sił profesjonalnego bibliotekarza, niezbędne tu umiejętności jesteśmy w stanie przekazać w ramach szkolenia oraz poprzez wyposażenie w stosowne instrukcje. Dla drugiego etapu – kontroli danych – nieodzowna jest praca i doświadczenie bibliotekarza. Zgodnie z oczekiwaniami zastosowany podział zadań znacząco przyspieszył tempo realizacji tej części projektu. Poprzez metodę pracy w grupach katalogerskich starano się stworzyć system lepszego zarządzania wiedzą i umiejętnościami w projekcie.

Nakład wysoko wykwalifikowanej pracy został skupiony na węższym sektorze działania, przez co natychmiast wzrosła jego efektywność mierzona w ilości opracowanego materiału. Było to możliwe dzięki rekrutacji i wdrożeniu do pracy szerokiej grupy nieprofesjonalnych współpracowników, w dostatecznym stopniu przeszkolonych i wdrożonych do przygotowywania danych dla bibliotekarzy. Wykonywali oni fragment procesu niezbędny do realizacji zadania, lecz najmniej wydajny z punktu widzenia relacji pomiędzy nakładem czasu pracy i umiejętnościami a osiągniętymi efektami.

Inną konsekwencją metody zespołowego opracowania danych bibliograficznych była naturalna kontrola zgromadzonych opisów, włączona już w sam system pracy. Przy wzroście tempa opracowania dawało to gwarancje utrzymania wysokiej jakości opracowania, eliminowania możliwych błędów w samych opisach lub niespójności w strukturze danych już na poziomie opracowania. Ponadto pozwalało nam na włączanie do projektu w razie potrzeby dodatkowych ekip wspomagających pracę bibliotekarzy. W organizacji pracy zespołów katalogerskich utrzymaliśmy założenie co do zupełnie swobodnego określania intensywności pracy i stopnia zaangażowania ze strony rekrutowanych katalogerów.

Narzędzia

Przedsięwzięcie realizowane jest przy użyciu kilku narzędzi informatycznych. Projekt posiada swoją witrynę roboczą <http://hist.pl/bazhum>, gdzie zamieszczane są wszystkie ogłoszenia, instrukcje i różne warianty dokumentacji. Poza tym używamy forum <http://hist.pl/forum> dostępnego dla zarejestrowanych użytkowników, na którym odbywają się dyskusje dotyczące rozwiązań bibliograficznych, systemowych oraz organizacyjnych.

Narzędziem dostępnym tylko dla współpracowników projektu jest portal Mantis (<http://hist.pl/bugs>), poprzez który osoby pracujące nad projektem mogą zgłaszać różnorakie problemy lub pomysły rozwiązań informatycznych. Jest to o tyle ważne, że – jak pokazała praktyka – narzędzie edycyjne, które służy do gromadzenia danych, przystosowane jest do pracy z różnymi systemami operacyjnymi (Windows, Macintosh, Linux), ale z każdym z nich pracuje w trochę odmienny sposób, w związku z czym pojawiają się odmienne problemy.

Program DeskLight pozwala nie tylko na wprowadzanie danych przez katalogerów, ale także na ich dodawanie i poprawianie przez kontrolerów. Twórcami tego narzędzia są pracownicy ICM. Program instaluje się na bardzo podstawowym poziomie i pracuje w oparciu o program Java. Do codziennej pracy z DL nie jest potrzebny Internet. Wprowadzane dane eksportuje się w postaci plików (paczek), które można wymieniać także za pomocą przenośnych nośników pamięci.

Projekt posiada zaplecze serwerowe i informatyczne. Dyponujemy własnymi kontami pocztowymi dla współpracowników, serwerami FTP do gromadzenia tymczasowych efektów pracy oraz w pełni opracowanych czasopism. Po ukończeniu opracowania czasopisma, pliki danych są transferowane na platformę YADDA, obecnie na zasadzie eksportu i importu paczek danych, w przyszłości metodą synchronizacji serwerowej wersji DeskLighta.

Platformą YADDA zarządza zespół ICM. Prezentowane są na niej czasopisma z ułożonymi hierarchicznie rekordami. Podstawą hierarchii jest tytuł czasopisma, następnie rok, tom, numer. Ostatnim ogniwem są czasopisma – obecnie ułożone alfabetycznie według pierwszych liter tytułu. Zdecydowaliśmy się stopnie hierarchii opisywać samymi cyframi, bez dodawania określników typu: R., T., nr.

Stan zaawansowania i perspektywy

Zasady wynagrodzenia współpracowników projektu zostały oparte na stawce za opracowanie jednego rekordu danych, przez który rozumiemy sekwencję metadanych opisujących jednostkę tekstową opublikowaną w danym czasopiśmie. Zachowaliśmy różnice między stawkami dla różnych kategorii współpracowników, ponadto niewielkie zróżnicowanie ze względu na trudność opracowania danych (alfabety niełacińskie) i premie za wysoką liczbę opracowanych danych. Aktywność musiała być permanentna i do pewnego stopnia uzależniona od rytmu pracy przypisanej im grupy katalogerskiej, dlatego część ich wynagrodzenia była ustalona sztywno.

Obecnie baza BazHum w wersji dostępnej publicznie gromadzi opisy bibliograficzne do zawartości 54 czasopism, w tym przede wszystkim dotyczących nauk historycznych, co przekłada się na zasób około 50 000 rekordów. Udostępniana treść jest jednak niepełna i nie rejestruje m.in. recenzji i drobnych tekstów. Ich dane są stopniowo uzupełniane i będą opublikowane w kolejnej aktualizacji bazy.

W materiałach kompletnie opracowanych, ale jeszcze nieopublikowanych, spoczywa dalsze 50 000 rekordów (przeważnie historia i historia sztuki), w toku bieżących prac natomiast przygotowujemy pełny rejestr kolejnych 50 000 rekordów danych (głównie historia sztuki i archeologia). Planujemy, aby do końca 2009 r. przekazać do otwartego użytku zasób 150 czasopism, w przybliżeniu 150 000 rekordów danych.

Ze względu na priorytety rozwoju przedsięwzięcia w najbliższym okresie działalności do najbardziej istotnych perspektyw i planów należą: zbudowanie konsorcjum instytucji współpracujących przy realizacji projektu, wymiana doświadczeń i rozważenie koordynacji działań z podobnymi projektami realizowanymi w kraju, a zwłaszcza z BazTech-em oraz uruchomienie opracowania komponentu pełnotekstowego.

Przypis

[1] Odesłania do stron internetowych przedstawiają wersję aktualną w dniu 10 maja 2009 r.

BazHum: baza zawartości czasopism humanistycznych. Dokonania i perspektywy / Michał T. Szczepański, , Tomasz Chmielak // W: Bibliograficzne bazy danych : kierunki rozwoju i możliwości współpracy. Bydgoszcz, 27-29 maja 2009. - [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich, K[omisja] W[ydawnictw] E[lektronicznych], Redakcja "Elektronicznej Biblioteki", 2009. - (EBIB Materiały konferencyjne nr 19). - ISBN: 83-921757-5-1. -Tryb dostępu : http://www.ebib.info/publikacje/matkonf/mat19/szczepanski_chmielak.php