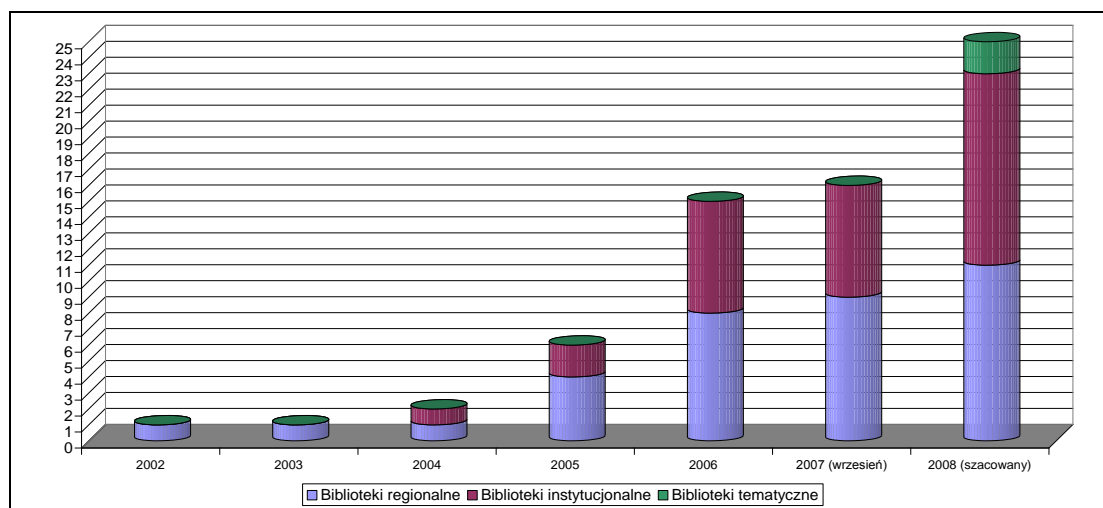


Federacja Bibliotek Cyfrowych w sieci PIONIER – Dostęp do otwartych bibliotek cyfrowych i repozytoriów

Agnieszka Lewandowska, Cezary Mazurek, Marcin Werla
{jagna,mazurek,mwerla}@man.poznan.pl
Poznańskie Centrum Superkomputerowo-Sieciowe
ul. Noskowskiego 12/14, 61-704 Poznań

Wprowadzenie

W roku 2002 oficjalnie udostępniono w Internecie Wielkopolską Bibliotekę Cyfrową (WBC) – pierwszą tego typu inicjatywę w Polsce. Podstawowym celem WBC, współtworzonej przez poznańskie środowisko akademickie oraz biblioteki naukowe i publiczne Poznania i Wielkopolski, było udostępnianie przez Internet cyfrowych publikacji zgromadzonych w czterech kolekcjach tematycznych: materiały dydaktyczne, dziedzictwo kulturowe, materiały regionalne oraz muzykalia. Koordynacją działań związanych z WBC zajęła się Poznańska Fundacja Bibliotek Naukowych, a za stronę techniczną odpowiedzialne było Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS). Instalację WBC oparto na oprogramowaniu dLibra, dzięki czemu możliwe było równoczesne tworzenie biblioteki cyfrowej w sposób rozproszony (tj. przez wiele różnych instytucji, w oparciu o wymianę danych poprzez Internet) oraz spójne i efektywne przechowywanie i udostępnianie gromadzonych zasobów cyfrowych, a także elastyczne zarządzanie nimi [1]. Obecnie WBC zawiera około 43 000 cyfrowych publikacji, co czyni ją największą biblioteką cyfrową w Polsce. Przy obecnym tempie udostępniania nowych zasobów (około 1 500 – 2 000 obiektów miesięcznie) do końca 2007 roku WBC powinna przekroczyć próg 50 000 publikacji.



Rys.1. Liczba publicznie dostępnych bibliotek cyfrowych opartych na oprogramowaniu dLibra (o ile nie zaznaczono inaczej, dane na koniec poszczególnych lat).

Uruchomienie WBC zapoczątkowało powstawanie kolejnych bibliotek cyfrowych. W 2004 roku powstała Biblioteka Cyfrowa Politechniki Wrocławskiej, przekształcona później w Dolnośląską Bibliotekę Cyfrową. Poza WBC największe obecnie biblioteki cyfrowe to Kujawsko-Pomorska Biblioteka Cyfrowa (ok. 15 000 publikacji), Biblioteka Cyfrowa Uniwersytetu Wrocławskiego (ok. 10 000 publikacji) oraz Śląska Biblioteka Cyfrowa, Małopolska Biblioteka Cyfrowa i Cyfrowa Biblioteka Narodowa POLONA (każda

po ok. 5 000 publikacji). Łącznie wszystkie biblioteki cyfrowe udostępniają około 90 000 cyfrowych publikacji. Dynamika wzrostu liczby bibliotek cyfrowych w Polsce przedstawiona jest na rysunku 1.

Obecnie funkcjonuje 16 bibliotek cyfrowych opartych na oprogramowaniu dLibra. Z tego dziewięć to regionalne biblioteki cyfrowe, a siedem to biblioteki instytucjonalne. W roku 2008 prawdopodobnie udostępnione zostaną pierwsze tematyczne biblioteki cyfrowe. Treści do tych bibliotek opracowywane są przez ponad 150 instytucji.

Tak duża liczba bibliotek cyfrowych oraz treści cyfrowych dostępnych on-line stały się podstawą do rozpoczęcia prac nad stworzeniem infrastruktury bibliotek cyfrowych działającej w skali całego kraju. Dzięki pracom badawczo-rozwojowym prowadzonym przez PCSS¹ uruchomione w sieci Polski Internet Optyczny PIONIER biblioteki cyfrowe zostały połączone przy pomocy protokołu OAI-PMH [2]. Ten służący do wymiany metadanych protokół umożliwił zrealizowanie mechanizmu wyszukiwania rozproszonego, pozwalającego na przeszukiwanie opisów zasobów cyfrowych wszystkich bibliotek cyfrowych opartych na oprogramowaniu dLibra z poziomu stron WWW dowolnej z tych bibliotek [3]. Wykorzystanie protokołu OAI-PMH spowodowało również powstanie automatycznie nadawanych, unikalnych w skali światowej identyfikatorów udostępnianych obiektów cyfrowych.

Kolejny element infrastruktury bibliotek cyfrowych w Polsce to wyszukiwarka „Federacja Bibliotek Cyfrowych” (FBC) udostępniona w czerwcu 2007 roku pod adresem <http://fbc.pionier.net.pl/>. Poniżej opisano ideę tego serwisu, zasadę jego działania oraz podstawowe funkcje. Na końcu artykułu przedstawiono również trwające obecnie prace nad rozwojem FBC i sieci polskich bibliotek cyfrowych.

Serwis „Federacja Bibliotek Cyfrowych”

Informacje ogólne

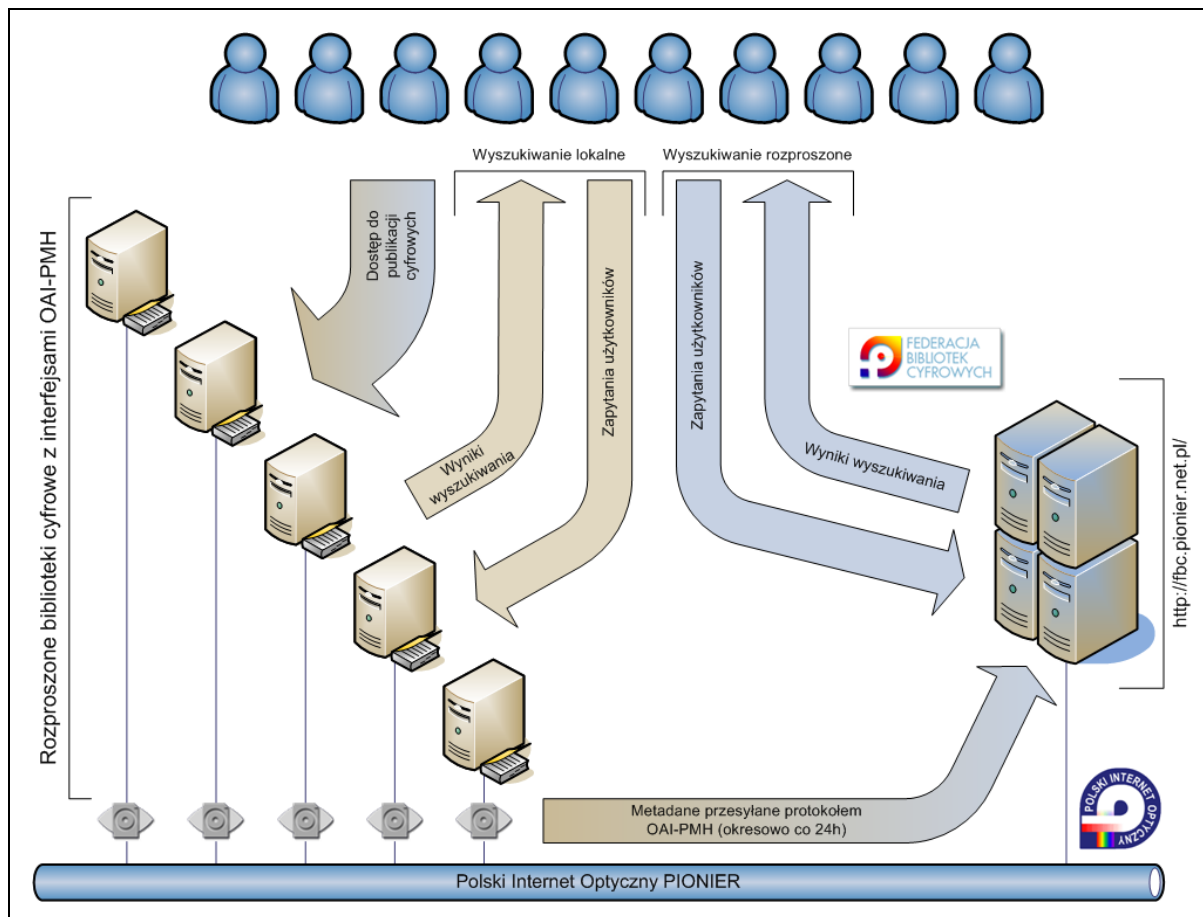
Federacja Bibliotek Cyfrowych ma na celu wirtualne połączenie bibliotek cyfrowych i repozytoriów dostępnych w polskim Internecie oraz udostępnienie nowych zaawansowanych funkcji i usług sieciowych realizowanych w tym środowisku. Ponadto misją serwisu jest ułatwienie wykorzystania zasobów polskich bibliotek cyfrowych i repozytoriów oraz zwiększenie widoczności tych zasobów w światowym Internecie.

Serwis FBC jest utrzymywany i rozwijany przez PCSS. Jednym z podstawowych założeń technicznej koncepcji serwisu jest oparcie komunikacji z bibliotekami cyfrowymi i repozytoriami na otwartych protokołach i formatach danych (obecnie podstawą jest wspomniany wcześniej protokół OAI-PMH oraz schemat Dublin Core). Dzięki temu FBC nie jest związane z żadnym konkretnym dostawcą oprogramowania dla bibliotek i repozytoriów cyfrowych. Jedynym wymaganiem stawianym repozytoriom, które mają być widoczne w FBC jest możliwość komunikacji z nimi przy pomocy protokołu OAI-PMH.

Z formalno-organizacyjnego punktu widzenia współpraca bibliotek cyfrowych i repozytoriów z serwisem FBC jest bardzo prosta. Współpraca taka nie wymaga żadnych opłat ani dodatkowego nakładu pracy ze strony administratorów repozytorium. Nie wymaga ona również przekazywania publikacji cyfrowych na rzecz FBC. Trzeba jedynie podać

¹ Prace prowadzone m.in. w ramach projektu badawczego nr 3 T11C 023 30 Ministerstwa Nauki i Szkolnictwa Wyższego: „Mechanizmy usług atomowych dla rozproszonych bibliotek cyfrowych”.

administratorom FBC adres URL pozwalający na pobieranie danych z repozytorium poprzez protokół OAI-PMH – czyli adres interfejsu OAI-PMH tego repozytorium. Pod tym względem serwis FBC działa tak samo jak inne bazujące na repozytoriach OAI-PMH serwisy takie jak np. OAIster [4]. W ramach współpracy FBC okresowo automatycznie pobiera metadane o udostępnianych obiektach cyfrowych i gromadzi je w lokalnym katalogu. Pobrane z rozproszonych repozytoriów metadane stanowią podstawę działania serwisu FBC. Użytkownicy mogą przeszukiwać lokalne biblioteki cyfrowe, a równocześnie poprzez FBC mają dostęp do wyszukiwania rozproszonego (rysunek 2).



Rys.2. Schemat obrazujący zasadę działania serwisu „Federacja Bibliotek Cyfrowych”.

Wyszukiwanie rozproszone

Aby przeszukać zasoby bibliotek cyfrowych czytelnicy mają teraz dwie możliwości: mogą użyć wyszukiwania lokalnego po kolei w każdej z bibliotek cyfrowych i analizować uzyskane w ten sposób listy wyników lub mogą użyć wyszukiwania rozproszonego w FBC lub innej bibliotece cyfrowej. Z pomocą wyszukiwania rozproszonego otrzymują jedną listę wyników zawierającą zasoby z wielu repozytoriów. Wybranie konkretnego wyniku z tej listy powoduje przejście do właściwej strony biblioteki czy repozytorium, w którym ten obiekt się znajduje. Przykładowe wyniki wyszukiwania przedstawiono na rysunku 3.

Dodatkowo, w celu jak najszerszego promowania polskich zasobów cyfrowych, serwis FBC pozwala na umieszczenie swojego interfejsu wyszukiwawczego na dowolnej innej stronie w postaci mini-wyszukiwarki. Wtedy internauci odwiedzający tę stronę mogą bezpośrednio z niej przeszukać zasoby polskich bibliotek cyfrowych (rysunek 4). Osadzenie na własnej stronie mini-wyszukiwarki jest bardzo łatwe i sprowadza się do nieznacznego

zmodyfikowania kodu HTML tej strony. Ponadto każdy internauta może również dodać funkcję przeszukiwania FBC bezpośrednio do swojej przeglądarki internetowej. Wtedy ma on możliwość wyszukiwania w FBC wprost z pola wyszukiwawczego przeglądarki. Rozszerzenie to jest oparte na formacie OpenSearch [5], który obecnie jest obsługiwany przez przeglądarki Firefox 2 i Internet Explorer 7. Szczegółowe informacje na ten temat są dostępne na stronach FBC w sekcji „Dodatki” (<http://fbc.pionier.net.pl/owoc/text?id=addons>).



Rys.3. Przykładowe wyniki wyszukiwania w serwisie FBC (<http://fbc.pionier.net.pl>, strona odwiedzona 15.09.2007 r.).



Rys.4. Strona katalogów Biblioteki Kórnickiej PAN z osadzoną mini-wyszukiwarką FBC (<http://www.bkpan.poznan.pl/katalogi/index.html>, strona odwiedzona 15.09.2007 r.).

Trwałe referencje do obiektów cyfrowych

Jak wspomniano wcześniej, udostępnienie protokołu OAI-PMH w sieci polskich repozytoriów i bibliotek cyfrowych zaowocowało powstaniem automatycznie nadawanych, unikalnych w skali światowej identyfikatorów udostępnianych obiektów. Mechanizm tych identyfikatorów – OAI Identifier – zawarty jest w standardzie opisującym protokół OAI-PMH. Format identyfikatora wygląda następująco:

`oai:<domena repozytorium>:<identyfikator zasobu w repozytorium>`

a przykładowa jego instancja następująco:

`oai:www.wbc.poznan.pl:8711`

Serwis FBC posiada mechanizm rozpoznawania takich identyfikatorów, który pozwala na uzyskanie informacji oraz aktualnego adresu obiektu cyfrowego na podstawie jego unikalnego identyfikatora. Mechanizm ten może być również wykorzystany do utworzenia trwałej referencji do obiektu cyfrowego np. na potrzeby odwołań w bibliografii. Referencja taka ma następującą postać:

`http://fbc.pionier.net.pl/id/<identyfikator oai>`

a przykładowe odwołanie wygląda następująco:

<http://fbc.pionier.net.pl/id/oai:www.wbc.poznan.pl:8711>

Otwarcie takiego adresu w przeglądarce WWW spowoduje wyświetlenie podstawowych metadanych obiektu cyfrowego o podanym identyfikatorze oraz odnośników do metadanych i/lub treści tego obiektu. Utworzona w ten sposób referencja może być trwałym i poprawnym odwołaniem do cyfrowej publikacji, niezależnie od zmian wprowadzonych w oprogramowaniu biblioteki cyfrowej, która tę publikację udostępnia. Informacje o identyfikatorach poszczególnych obiektów cyfrowych, które można wykorzystać do tworzenia takich referencji zazwyczaj udostępniane są przez poszczególne repozytoria na stronach z opisem obiektów.

Koordinacja digitalizacji

Szczególnie istotną z punktu widzenia twórców polskich zasobów cyfrowych funkcją FBC jest możliwość przeszukiwania zebranych w jednym miejscu planów digitalizacji poszczególnych bibliotek cyfrowych. Ta funkcja FBC bazuje na pobieranych przez protokół OAI-PMH informacjach o planowanych publikacjach udostępnianych przez poszczególne biblioteki cyfrowe. Poza przeszukiwaniem możliwe jest również przeglądanie pełnej listy publikacji planowanych oraz wygenerowanie specjalnej postaci tej listy przeznaczonej do wydruku.

W chwili obecnej system dLibra (wersja 3.0) jest jedynym znanym autorom artykułu oprogramowaniem, które wspiera planowanie digitalizacji i równocześnie udostępnia gromadzone w ten sposób dane poprzez protokół OAI-PMH. Planowanie digitalizacji w obrębie jednej biblioteki cyfrowej dLibra wspiera już od wersji 1.6, wydanej w sierpniu 2005 roku. Cykl życia publikacji elektronicznej w bibliotece cyfrowej opartej na systemie dLibra może rozpoczynać się właśnie od „publikacji planowanej”, która pozwala na przechowanie metadanych na temat jeszcze nie istniejącego obiektu cyfrowego [6]. Publikację taką można później automatycznie przekształcić w pełny obiekt cyfrowy dołączając do niej pliki składowe tego obiektu. Jak wspomniano wcześniej, serwis FBC jest technicznie gotowy do współpracy z innymi systemami do budowy bibliotek cyfrowych, gdy tylko zaczną one gromadzić stosowne informacje i udostępnią je poprzez protokół OAI-PMH.

Liczba publikacji planowanych, o których informacje dostępne są w systemie FBC waha się obecnie w zakresie od 600 do 800 publikacji. Wahania te są naturalnym następstwem działań poszczególnych bibliotek cyfrowych polegających na przekształcaniu obiektów planowanych w pełne obiekty cyfrowe oraz dodawaniu nowych obiektów planowanych.

Dalsze prace

Przedstawiona powyżej funkcjonalność serwisu FBC jest ciągle rozwijana i ulepszana. Najbliższe plany rozwoju serwisu obejmują m.in. mechanizmy związane ze statystykami na temat dostępnych obiektów cyfrowych, uatrakcyjnieniem prezentacji zgromadzonych metadanych (np. poprzez wykorzystanie „chmur tagów”) czy dodatek do FBC umożliwiający integrację z popularnym serwisem iGoogle.

Poza tym trwają prace nad bardziej znaczącymi modyfikacjami serwisu związanymi z dalszą rozbudową infrastruktury sieci bibliotek cyfrowych i repozytoriów w Polsce. Pierwsza z nich polegać będzie na przekształceniu obecnej bazy publikacji planowanych w społecznościową platformę do koordynacji digitalizacji. Docelowo platforma ta ma umożliwiać m.in.

gromadzenie informacji o planach digitalizacji również z tych bibliotek cyfrowych, w których oprogramowanie nie pozwala na zbieranie i współdzielenie takich informacji oraz możliwości konsultacji planów digitalizacji z czytelnikami poprzez głosowanie na poszczególne planowane obiekty oraz proponowanie własnych publikacji. Dyskusja na temat ostatecznego kształtu tej platformy toczy się na forum Biblioteka 2.0 (w wątku dostępnym pod adresem <http://forum.biblioteka20.pl/viewtopic.php?p=896>).

Drugą znaczącą modyfikacją serwisu ma być włączenie go do budowanej w ramach sieci PIONIER rozproszonej platformy uwierzytelniania i autoryzacji dla bibliotek cyfrowych. Platforma ta pozwoli przede wszystkim na wykorzystanie jednego profilu użytkownika w wielu bibliotekach cyfrowych (i w serwisie FBC). Dzięki temu czytelnik będzie mógł np. na swojej wirtualnej półce z ulubionymi książkami umieszczać obok siebie publikacje pochodzące z różnych bibliotek cyfrowych, a półkę tę będzie miał dostępną w każdej bibliotece cyfrowej, do której się zaloguje. Rozszerzenie obejmie też wykorzystanie wielu różnych zewnętrznych baz użytkowników (np. baz studentów wyższych uczelni) przy autoryzacji dostępu do zasobów cyfrowych (np. skryptów akademickich). Podobne systemy budowane są również w innych krajach na świecie (np. system ATHENS w Wielkiej Brytanii – <http://www.athensams.net/>) jednak zazwyczaj skupiają się one przede wszystkim na kwestiach autoryzacji dostępu do określonych zasobów (np. licencjonowanych czasopism elektronicznych) i nie dają funkcji sieciowego profilu czytelnika.

Zakończenie

Opisana w niniejszym artykule „Federacja Bibliotek Cyfrowych” jest najnowszym elementem infrastruktury bibliotek cyfrowych w Polsce. Serwis ten rozwijany i utrzymywany jest przez Poznańskie Centrum Superkomputerowo-Sieciowe. Dostępne w systemie funkcje i dodatki pozwalają na efektywne wyszukiwanie w polskich bibliotekach cyfrowych, na przeszukiwanie i przeglądanie planów digitalizacji tych bibliotek oraz na łatwe identyfikowanie i odwoływanie się do poszczególnych obiektów cyfrowych. Twórcy serwisu planują rozwijać go tak, aby był on nowoczesnym narzędziem przydatnym zarówno dla użytkowników zasobów polskich bibliotek cyfrowych jak i dla twórców tych zasobów. Kierunki rozwoju serwisu FBC oraz całej sieci bibliotek cyfrowych w Polsce zgodne będą również z kierunkami podobnych prac realizowanych w innych projektach na świecie. Działania takie pozwolą na funkcjonowanie jednej spójnej przestrzeni informacyjnej, której polskie biblioteki cyfrowe będą ważnym i nieodzownym elementem.

Bibliografia

- [1] Mazurek C., Stroiński M., Werla M.: „*Wdrażanie regionalnych bibliotek cyfrowych w sieci PIONIER w oparciu o środowisko dLibra*”. Konferencja „INFOBAZY 2005 – Bazy Danych dla Nauki”, 25-27.09.2005, Gdańsk. Materiały konferencyjne. str. 58 - 64. ISBN 83-908112-3-5.
<http://dlibra.psnc.pl/biblioteka/publication/16>
- [2] *Specyfikacja protokołu OAI-PMH w wersji 2.0*.
<http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [3] Mazurek C., Stroiński M., Werla M., Węglarz J.: „*Metadata harvesting in regional digital libraries in PIONIER Network*”. Terena Networking Conference TNC 2006, 15-18.05.2006, Catania, Italy. Conference Proceedings: Campus-Wide Information Systems, Vol. 23, No. 4, str. 241 - 253, Emerald Group Publishing Limited 2006. ISSN 1065-0741. ISBN 1-84663-184-X.
<http://emeraldinsight.com/10.1108/10650740610704117>

- [4] Serwis *OAIster*
<http://www.oaister.org/>
- [5] *Specyfikacja protokołu OpenSearch w wersji 1.1.*
<http://www.opensearch.org/Specifications/OpenSearch/1.1>
- [6] Mazurek C., Werla M.: „*Digital Object Lifecycle in dLibra Digital Library Framework*”. 9th International Workshop of the DELOS Network of Excellence on Digital Libraries on Interoperability and Common Service, 11-13.05.2005, Heraklion, Greece. Workshop Proceedings.
[http:// delos-wp5.ukoln.ac.uk/dissemination/pdfs/werla.pdf](http://delos-wp5.ukoln.ac.uk/dissemination/pdfs/werla.pdf)