

Cezary Mazurek

## Biblioteki cyfrowe – narzędzia i usługi dla efektywnego zarządzania i udostępniania treści cyfrowych w Internecie



Fot. Peter Skadberg

**Od ponad pięć lat  
Poznańskie Centrum  
Suprekomputerowo-  
Sieciorowe prowadzi prace  
badawczo-rozwojowe oraz  
wdrożeńowe w obszarze  
bibliotek cyfrowych.**

Rozwój nowoczesnych usług, istotnych dla przyszłego kształtu społeczeństw informacyjnych, zależy od poziomu zaawansowania trzech zasadniczych czynników: **technologii sieci komputerowych, usług i narzędzi pośredniczących (middleware) oraz treści (zawartości) cyfrowej.** Spośród tych czynników, tworzenie zasobów cyfrowych związanych z kulturą i światowym dziedzictwem kulturowym, nauką, edukacją, jak również wydawnictwami elektronicznymi oraz sektorem audio-wizualnym, ma szczególne znaczenie. Wynika to z faktu, iż odpowiednia ich ilość i jakość jest niezbędna do powstawania zaawansowanych usług w sieciach nowej generacji. Zapewnienie treściom cyfrowym takich atrybutów jak: trwałość, wiarygodność, łatwy dostęp, wielojęzyczność, efektywne zarządzanie, możliwe jest jedynie poprzez integrację dostępu do treści z dostępem do narzędzi. Narzędzia te powinny wspomagać prezentację, zarządzanie i przetwarzanie tych treści w środowisku sieciowym. Najlepszym przykładem środowisk tego typu są biblioteki cyfrowe.

Jedną z pierwszych na świecie bibliotek cyfrowych był rozpoczęty w 1971 roku Project Gutenberg, udostępniający obecnie ponad 16 000 publikacji elektronicznych zebranych dzięki pracy wolontariuszy z całego świata. **Obecnie coraz więcej instytucji dostrzega potrzebę tworzenia ogólnodostępnych bibliotek cyfrowych.**

Biblioteki takie, poza możliwością umieszczania w nich aktualnie powstających podręczników, publikacji naukowych czy zasobów edukacyjnych, umożliwiają także zachowanie dziedzictwa kulturowego. Cenne starodruki mogą być przetwarzane do postaci cyfrowej i bezpiecznie udostępniane milionom czytelników w Internecie. Przykładem tego typu inicjatyw może być amerykański National Digital Information Infrastructure and Preservation Program realizowany przez Library of Congress.

Systemy wspomagające budowę bibliotek cyfrowych rozwijane są obecnie w wielu projektach na świecie. Przykładami mogą tu być projekty Fedora, DSpace, Greenstone, dLibra i oprogramowanie komercyjne np. DB2 Content Manager firmy IBM.

Projekt Fedora, opracowywany na University of Virginia oraz Cornell University, może być traktowany jako platforma zapewniająca podstawową funkcjonalność biblioteki cyfrowej. Platforma ta umożliwia uruchomienie cyfrowego repozytorium oraz zarządzanie nim. Udostępnia ona zdalne interfejsy służące do zarządzania, dostępu do treści cyfrowej oraz przeszukiwania treści i metadanych. W systemie Fedora brakuje jednak rozbudowanego

interfejsu graficznego, który pozwoliłby czytelnikom dotrzeć do zgromadzonych zasobów. Interfejs taki należy samemu dobudować korzystając z umieszczonej na stronach projektu dokumentacji technicznej. Przykładem takiego rozszerzenia może być komercyjny system VITAL zbudowany na platformie FEDORA przez firmę VTLS.

Rozwijany w laboratoriach MIT system o nazwie DSPACE jest natomiast systemem nastawionym na budowanie uczelnianych repozytoriów dokumentów cyfrowych. **Udostępnia on większość typowej funkcjonalności biblioteki cyfrowej, a dodatkowo rozszerza możliwości związane z pracą w grupach roboczych i umożliwia tworzenie społeczności.** System DSPACE jest systemem wielowarstwowym, do którego dostęp odbywa się poprzez przeglądarkę WWW. Dodatkowo udostępnione są usługi umożliwiające przeglądanie metadanych wszystkich obiektów w systemie (przy pomocy protokołu OAI-PMH – *The Open Archive Initiative Protocol for Metadata Harvesting*) oraz wsadowy import i eksport treści cyfrowych. Jako, że cały system udostępniony jest na licencji BSD, możliwe jest samodzielne dobudowanie innych dodatkowych usług.

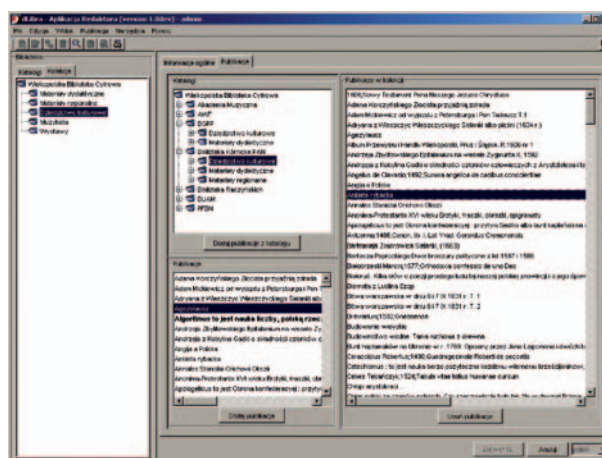
Również w pełni udostępniane przez WWW są biblioteki cyfrowe oparte o nowozelandzki projekt Greenstone. Projekt ten zapewnia podstawową funkcjonalność dostępną w innych systemach bibliotek cyfrowych oraz jest stosunkowo łatwo rozszerzalny, przez co umożliwia dodawanie specjalistycznych cech do bibliotek cyfrowych takich jak np.: wyszukiwanie utworów muzycznych na podstawie próbek dźwięku, czy fragmentów zapisu nutowego. Cechą charakterystyczną wszystkich bibliotek opartych o Greenstone jest to, iż można je nagrać na płyty CD i wykorzystywać bez dostępu do Internetu. Wynika to ze specyficznych zastosowań bibliotek cyfrowych przez jednego z głównych sponsorów systemu Greenstone – organizację UNESCO. Podobnie jak projekt DSPACE, Greenstone również umożliwia przeglądanie bibliotek cyfrowych przy pomocy protokołu OAI-PMH.

Jak wynika z powyższej charakterystyki, każdy z wymienionych systemów bibliotek cyfrowych wyposażony jest w specyficzne dla siebie rozwiązania i mechanizmy, jednakże wszystkie, a także większość pozostałych systemów bibliotecznych, posiadają kilka cech wspólnych. Należą do nich:

- mechanizm zarządzania zawartością: wykorzystywany do przechowywania, zarządzania i udostępniania zawartości publikacji

- mechanizm zarządzania metadanymi: wykorzystywany do opisywania publikacji za pomocą metainformacji zgodnie ze standardami metadanych (np. DublinCore)
- mechanizm organizacji publikacji: wykorzystywany do organizowania publikacji w wirtualne struktury – np. kolekcje
- mechanizm indeksowania i przeszukiwania: wykorzystywany do indeksowania zawartości publikacji oraz metadanych oraz do przeszukiwania na ich podstawie zawartości biblioteki.

Od ponad pięciu lat Poznańskie Centrum Superkomputerowo-Sieciowe prowadzi prace badawczo–rozwojowe oraz wdrożeniowe w obszarze bibliotek cyfrowych. **Powyższy zestaw ogólnych mechanizmów stanowił punkt wyjścia do opracowania i budowy pierwszego polskiego środowiska biblioteki cyfrowej – dLibra.** Obecnie dLibra jest systemem rozproszonym i otwartym, który umożliwia czytelnikom dostęp do zawartości biblioteki poprzez interfejs WWW, oferując równocześnie dostawcom treści i administratorom zaawansowane narzędzia w postaci „Aplikacji Redaktora/Administratora” (Rys. 1). dLibra umożliwia przechowywanie i udostępnianie obiektów cyfrowych dowolnego typu – mogą to być zarówno dokumenty tekstowe w formatach takich jak HTML, PDF czy DjVu, jak i pliki audio czy wideo. Każdy z przechowywanych obiektów może być opisany



Rys. 1. Aplikacja redaktora Systemu Biblioteki Cyfrowej dLibra

metadanymi oraz przypisany do jednej lub wielu zdefiniowanych w danej bibliotece cyfrowej kolekcji. Użytkownicy systemu mają do dyspozycji zaawansowane mechanizmy tworzenia metadanych, takie jak słow-



niki wartości poszczególnych atrybutów czy obsługę formatów MARC i RDF. Czytelnicy mają możliwość przeglądania zawartości biblioteki oraz przeszukiwania metadanych poszczególnych obiektów oraz ich treści (dla określonych formatów). Niezwykle cenna jest również możliwość przeszukiwania zawartości biblioteki cyfrowej przy pomocy zewnętrznych wyszukiwarek internetowych. Przykładem jest jedna z najpopularniejszych wyszukiwarek internetowych – Google. Za jej pośrednictwem można dotrzeć do publikacji przechowywanych w systemie dLibra.

Od 1 października 2002 roku system dLibra jest wdrożony we współpracy z Poznańską Fundacją Bibliotek Naukowych w Wielkopolskiej Bibliotece Cyfrowej (Rys. 2), zawierającej obecnie ponad 6 000 publikacji. Środowisko to posłużyło także w ostatnich latach do budowy kolejnych bibliotek cyfrowych w Polskim Internecie Optycznym PIONIER (Rys. 3). W Polsce działa obecnie kilkanaście bibliotek cyfrowych, gromadzących różnego rodzaju zasoby – głównie edukacyjne, historyczne i naukowe. Ze względu na liczbę posiadanych publikacji **do największych bibliotek cyfrowych w Polsce należą: finansowana przez Ministerstwo Nauki i Informatyzacji Polska Biblioteka Internetowa (PBI) oraz oparte na oprogramowaniu dLibra: Wielkopolska Biblioteka Cyfrowa, Kujawsko – Pomorska Biblioteka Cyfrowa i Biblioteka Cyfrowa Politechniki Wrocławskiej.** Do końca 2005 roku uruchomione zostaną także: Podlaska Biblioteka Cyfrowa, Zielonogórska Biblioteka Cyfrowa, Biblioteka Cyfrowa Politechniki Łódzkiej oraz Biblioteka Cyfrowa Uniwersytetu Wrocławskiego. Wszystkie te biblioteki działać będą przy użyciu najnowszej wersji oprogramowania dLibra. Spośród bibliotek cyfrowych realizowanych w oparciu o własne rozwiązania wymienić należy jeszcze: Bibliotekę Cyfrową Uniwersytetu Gdańskiego oraz Akademicką Bibliotekę Cyfrową AGH w Krakowie. Rosnąca liczba instalacji bibliotek cyfrowych jest powodem do podjęcia dalszych prac w zakresie mechanizmów wzajemnej komunikacji pomiędzy poszczególnymi bibliotekami cyfrowymi. Mechanizmy takie są przedmiotem aktualnie prowadzonych prac w projekcie dLibra. Docelowo umożliwią one rozproszone wyszukiwanie oraz wymianę metadanych pomiędzy biblio-

tekami cyfrowymi uruchomionymi w sieci PIONIER oraz tymi, które implementują standardowe protokoły wymiany danych.

Powstanie i dostępność cyfrowych treści to jeden z podstawowych impulsów rozwoju technologii informacyjnych. Środowiska bibliotek cyfrowych stanowią mogą zaplecze dla rozwoju takich usług jak wirtualne muzea, multimedialne serwisy na żądanie, zdalne nauczanie, elektroniczne wydawnictwa, itp. Te obszary zastosowań powodują, iż w niedalekiej przyszłości możemy się spodziewać powstawania rozproszonych systemów sieciowych bibliotek cyfrowych, które przechowują i udostępniają kolekcje wybrane pod względem znaczeniowym, funkcjonalnym czy też praw własności. Jednakże poza archiwizowaniem i udostępnianiem treści, sprawne budowanie



Rys. 2. Strona główna Wielkopolskiej Biblioteki Cyfrowej

zasobów cyfrowych wymaga zastosowania odpowiedniej organizacji bibliotek cyfrowych i wykorzystania narzędzi do zarządzania ich zawartością w modelu rozproszonym. Uwzględnienie takich mechanizmów (usług) zarządzania pozwoli na budowanie zintegrowanych systemu rozproszonych bibliotek cyfrowych, które zdyktalizują powstawanie nowych usług i aplikacji funkcjonujących w oparciu o dobre i pełne zasoby informacyjne.

**Dr Cezary Mazurek** kieruje obecnie pracami rozwojowymi Poznańskiego Centrum Superkomputerowo Sieciowego związanymi z budową Polskiego Portalu Edukacyjnego, Multimedialnego Informatora Miejskiego o Poznaniu, Wielkopolskiej Biblioteki Cyfrowej oraz innych usług realizowanych na bazie technologii internetowych.