

Cezary Mazurek, Tomasz Parkoła, Marcin Werla

Poznańskie Centrum Superkomputerowo-Sieciowe

Autor do korespondencji: Cezary Mazurek, mazurek@man.poznan.pl

**ROLA BIBLIOTEK CYFROWYCH
W ZACHOWANIU REGIONALNEJ KULTURY
I TRADYCJI**

Streszczenie: W niniejszym artykule zaprezentowano etapy rozwoju infrastruktury polskich bibliotek cyfrowych oraz znaczenie regionalnych i lokalnych zasobów cyfrowych w kontekście krajowym oraz europejskim. We wstępie przybliżono tematykę popularyzacji dziedzictwa kulturowego w internecie. W rozdziale pierwszym omówiono etapy rozwoju polskich bibliotek cyfrowych. Rozdział drugi przedstawia ideę oraz misję Federacji Bibliotek Cyfrowych. W rozdziale trzecim przedstawiono polskie biblioteki cyfrowe w kontekście europejskim, a w szczególności umiejscowienie lokalnych i regionalnych cyfrowych zasobów dziedzictwa kulturowego w infrastrukturze europejskiej biblioteki cyfrowej Europeana.

Słowa kluczowe: polskie biblioteki cyfrowe, rozwój bibliotek cyfrowych, infrastruktura dla nauki, dLibra, Federacja Bibliotek Cyfrowych, Europeana.

Klasyfikacja JEL: O33.

Wstęp

Popularyzacja dziedzictwa kulturowego z wykorzystaniem nowoczesnych technologii cyfrowych staje się coraz ważniejszym elementem działalności takich instytucji, jak: biblioteki, muzea i archiwa. Wskazuje na to wiele inicjatyw i projektów realizowanych przez największe instytucje kultury i nauki w Europie i na świecie. Przykładami mogą być projekty Europeana (<http://www.europeana.eu>), cyfrowe kolekcje Biblioteki Kongresu Stanów Zjednoczonych (<http://www.loc.gov/library/libarch-digital.html>) czy Internet Archive (<http://www.archive.org/>). Kierunki prac oraz zalecenia dotyczące działań w zakresie popularyzacji i udostępniania online dokumentów dziedzictwa kulturowego są również opracowywane przez Komisję Europejską. W raporcie „The New Renaissance” [Comité des Sages, 2011] grupy Comité des Sages działającej przy

Komisji Europejskiej znajduje się sześć kluczowych zaleceń w kontekście umieszczania europejskiego dziedzictwa kulturowego w internecie:

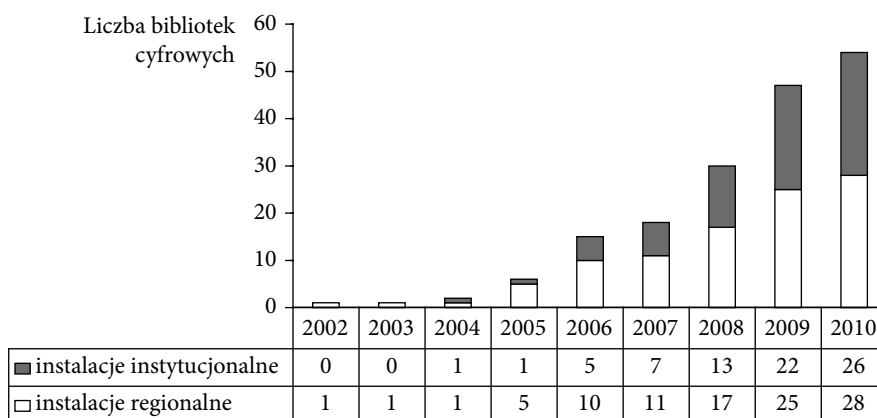
1. Zapewnienie szerokiego dostępu i wykorzystania dokumentów będących własnością publiczną. Warto podkreślić, że zdaniem autorów raportu, proces digitalizacji nie powinien prowadzić do powstania nowych praw autorskich, a opisujące obiekty dziedzictwa kulturowego, metadane, powinny być ogólnodostępne w celu łatwego ponownego wykorzystania.
2. Rozwiązanie problemów utworów osieroconych oraz udostępnienie zasobów o wyczerpanym nakładzie. Kluczowymi elementami w tym kontekście są: wsparcie KE w rozmowach nad dostosowaniem prawa międzynarodowego oraz umożliwienie instytucjom kultury udostępnienia utworów, których nakłady zostały wyczerpane, a właściciele praw i podmioty komercyjne nie podejmują tej inicjatywy.
3. Promocja i wzmocnienie roli portalu Europeana jako podstawowego źródła informacji o zabytkach i dokumentach dziedzictwa kulturowego Europy. Przykładem tego typu działań będzie rozpoczynający się w styczniu 2012 roku projekt *EuropeanaAwareness*, którego polskimi partnerami są Narodowy Instytut Audiowizualny oraz Poznańskie Centrum Superkomputerowo-Sieciowe.
4. Zapewnienie trwałości zasobów cyfrowych w celu umożliwienia dostępu do przechowywanych materiałów w perspektywie długoterminowej.
5. Finansowanie digitalizacji i Europeany w sposób stabilny, przy wykorzystaniu środków krajowych oraz europejskich.
6. Racjonalizacja partnerstwa publiczno-prywatnego w dziedzinie digitalizacji poprzez ustalenie zestawu reguł odnoszących się do inwestycji w digitalizację prywatnych przedsiębiorstw.

W Polsce aktywne działania w tym zakresie podejmuje Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS), będące operatorem krajowej naukowej sieci PIONIER – Polski Internet Optyczny, które od wielu lat rozwija ogólnopolską infrastrukturę bibliotek cyfrowych [Mazurek i in. 2008, s. 9–14; Mazurek, Stroiński i Werla 2005, s. 58–64]. Prace w tym obszarze rozpoczęły się kilkanaście lat temu, wraz z opracowaniem pierwszego polskiego oprogramowania do budowy bibliotek cyfrowych – dLibra [Mazurek 2001; Mazurek, Parkoła i Werla 2010, s. 479–484]. W 2002 roku, we współpracy z Poznańską Fundacją Bibliotek Naukowych oraz Biblioteką Kórnicką PAN, uruchomiona została pierwsza regionalna biblioteka cyfrowa w Polsce – Wielkopolska Biblioteka Cyfrowa (<http://www.wbc.poznan.pl/>). Jej zasoby współtworzone są przez wiele instytucji kultury i nauki z regionu, w tym biblioteki, archiwa i muzea [Górny i in. 2003, s. 109–117]. Szybko wzrastająca liczba bibliotek cyfrowych w Polsce zaowocowała uruchomieniem portalu Federacji Bibliotek Cyfrowych (FBC), który jest nie tylko punktem dostępu do wszystkich zasobów polskich bibliotek cyfrowych, ale oferuje również zaawansowane funkcje użytkownikom internetu oraz zewnętrznym systemom informatycznym [Lewandowska, Mazurek i Werla 2008, s. 256–259]. FBC odgrywa również istotną rolę

w popularyzacji polskich regionalnych i lokalnych zasobów cyfrowych na arenie międzynarodowej, w tym w Europeanie (<http://www.europeana.eu>). Obecnie w sieci PIONIER znajduje się ponad 80 bibliotek cyfrowych, które dają łącznie dostęp do ponad 800 000 obiektów cyfrowych.

1. Rozwój bibliotek cyfrowych w Polsce

Uruchomiona w październiku 2002 roku pierwsza regionalna biblioteka cyfrowa w Polsce – Wielkopolska Biblioteka Cyfrowa, stała się impulsem do tworzenia kolejnych takich bibliotek. Istotnym elementem rozwoju w tym kontekście jest istnienie najpopularniejszego oprogramowania do budowy bibliotek cyfrowych w Polsce – dLibra (<http://dlibra.psnc.pl/>). Innowacyjny charakter tego oprogramowania oraz jego stały rozwój we współpracy z użytkownikami są gwarantami wysokiej jakości tej platformy oraz jej powszechnej akceptacji przez instytucje zainteresowane udostępnianiem zasobów cyfrowych w internecie. Poniższy rysunek przedstawia liczbę bibliotek cyfrowych w kolejnych latach rozwoju infrastruktury polskich bibliotek cyfrowych. Około 75% instalacji opartych jest na oprogramowaniu dLibra. Podział bibliotek cyfrowych na instytucjonalne i regionalne wynika przede wszystkim z charakteru zasobów oraz liczby współpracujących instytucji. W przypadku bibliotek instytucjonalnych odpowiedzialność za ich utrzymanie spoczywa na jednej instytucji, a tematyka zasobów jest ukierunkowana przede wszystkim na profil instytucji. Biblioteki regionalne współtworzone są najczęściej przez wiele instytucji regionu, a tematyka ich zasobów dotyczy w dużej mierze szerokiego spektrum dziedzin i obejmuje regionalne zabytki dziedzictwa kulturowego. Wsparcie techniczne w przypadku regionalnych bibliotek cyfrowych jest często świadczone przez regionalne centra obliczeniowe.



Rozwój bibliotek cyfrowych w Polsce w latach 2002–2010

2. Federacja Bibliotek Cyfrowych

Dynamiczny rozwój sieci bibliotek cyfrowych w Polsce zaowocował powstaniem w czerwcu 2007 roku portalu FBC. Serwis FBC jest utrzymywany i rozwijany przez PCSS. Działanie tego serwisu bazuje na okresowej komunikacji z poszczególnymi bibliotekami cyfrowymi. Komunikacja ta spełnia otwarte i uznane standardy, dzięki czemu biblioteki cyfrowe włączone w FBC oraz sam serwis FBC nie są w żaden sposób uzależnione od dostawców oprogramowania. Podstawowym protokołem komunikacji FBC z bibliotekami cyfrowymi jest OAI-PMH (<http://www.openarchives.org/OAI/openarchivesprotocol.html>), a formatem wymienianych informacji standard Dublin Core (<http://dublincore.org/>) oraz PLMET (<http://dl.psn.c.pl/community/display/FBCMETGUIDE/>). Zasadniczym celem serwisu FBC jest ułatwienie wykorzystania i zwiększenie widoczności w internecie zasobów polskich bibliotek cyfrowych oraz udostępnienie użytkownikom internetu zaawansowanych usług sieciowych bazujących na sieci rozproszonych bibliotek cyfrowych w Polsce. Cel ten jest osiąganym przede wszystkim przez udostępnianie nowych funkcji bazujących na zasobach polskich bibliotek cyfrowych. Najważniejsze funkcje serwisu FBC to:

- wyszukiwanie – pozwala na wyszukiwanie obiektów we wszystkich polskich bibliotekach cyfrowych przyłączonych do FBC przy pomocy informacji zawartych w metadanych obiektów cyfrowych; mechanizm wyszukiwania może być wykorzystywany w formie prostej i zaawansowanej; forma zaawansowana pozwala na budowanie zapytań wyszukiwawczych przy wykorzystaniu operatorów logicznych do łączenia zapytań prostych;
- przeglądanie – pozwala na zapoznanie się z zbiorami bibliotek cyfrowych osobom, które nie chcą korzystać z funkcji wyszukiwawczych (gdyż np. nie szukają niczego konkretnego), możliwe jest przeglądanie obiektów ostatnio dodanych, rekomendowanych, ostatnio oraz najczęściej wyświetlanych;
- plany digitalizacji – serwis FBC gromadzi plany digitalizacji poszczególnych bibliotek cyfrowych i pozwala na ich przeszukiwanie; mechanizm ten wspiera koordynację prac digitalizacyjnych w polskich bibliotekach cyfrowych, na jego podstawie generowany jest okresowo raport z potencjalnymi duplikatami, które znajdują się w polskich bibliotekach cyfrowych;
- lokalizowanie obiektów cyfrowych – funkcja pozwalająca na lokalizację obiektu cyfrowego na podstawie jego unikalnego identyfikatora – format identyfikatora powinien być zgodny z formatem OAI Identifier, który jest wykorzystywany przez większość bibliotek cyfrowych w Polsce;
- baza danych o bibliotekach cyfrowych – zawiera wiele informacji na temat polskich bibliotek cyfrowych i ich zbiorów, a także bezpłatne kursy e-learningowe przeznaczone dla osób chcących więcej się dowiedzieć na temat tworzenia takich bibliotek, zwłaszcza w mniejszych instytucjach kultury;

- repozytorium OAI-PMH – FBC udostępnia metadane obiektów z polskich bibliotek cyfrowych zewnętrznym systemem informatycznym; są one dostępne są poprzez standardowy protokół OAI-PMH.

Mechanizmy udostępnianie przez portal FBC pozwalają użytkownikom końcowym oraz zewnętrznym systemem informatycznym na korzystanie z dóbr kultury dostarczanych przez polskie biblioteki cyfrowe. FBC jest nie tylko miejscem wyszukiwania informacji o zasobach, ale stanowi również element koordynacji prac digitalizacyjnych pomiędzy poszczególnymi instytucjami. Dzięki funkcjom udostępniania zebranych z bibliotek cyfrowych informacji, jest ważnym elementem popularyzacji polskich dokumentów w takich portalach, jak: Europeana (<http://www.europeana.eu>), DART-Europe (<http://www.dart-europe.eu/>) czy ViFaOst (<http://www.vifaost.de/>).

3. Polskie biblioteki cyfrowe w Europie

Komisja Europejska od wielu lat inicjuje i prowadzi działania mające na celu wsparcie rozwoju gospodarki opartej na wiedzy i rozwoju społeczeństwa informacyjnego. Jednym z efektów tych działań było przygotowanie warunków do rozpoczęcia prac w kierunku umożliwienia szerokiego cyfrowego dostępu do zasobów kulturowych. W ramach inicjatywy Komisji Europejskiej „i2010 Digital Libraries” zapoczątkowano współfinansowanie serii projektów realizujących kolejne fragmenty Europejskiej Cyfrowej Biblioteki, Muzeum i Archiwum – Europeany. Europeana ma ułatwić dostęp do dziedzictwa kulturowego i naukowego z tysięcy instytucji rozproszonych w Europie. Tak szeroko zakrojona współpraca wymaga dedykowanego modelu organizacyjnego, który pozwoli na podział zadań pomiędzy głównymi partnerami dostarczającymi danych do Europeany. Przyjęto wielopoziomowy model pozyskiwania metadanych rozproszonych obiektów cyfrowych, który zakłada utworzenie agregatorów, czyli serwisów odpowiedzialnych za gromadzenie (agregowanie) metadanych z bibliotek cyfrowych i przekazywanie ich agregatorom wyższego poziomu, przy czym najwyższy poziom tej struktury stanowić ma sama Europeana.

W modelu założono, że instytucje o charakterze narodowym (biblioteki, muzea, archiwa) udostępniają metadane swoich obiektów cyfrowych dedykowanym agregatorom (TEL, ATHENA, APEnet). Metadane z pozostałych instytucji są pozyskiwane poprzez agregatory dziedziczne oraz agregatory regionalne lub krajowe, niewyróżniające konkretnej tematyki. W celu wsparcia rozwoju agregatorów przeznaczonych dla lokalnych i regionalnych instytucji kultury i nauki, powołano projekt EuropeanaLocal, którego głównym celem było zwiększenie możliwości użycia metadanych i treści cyfrowych dostępnych w instytucjach lokalnych i regionalnych oraz doprowadzenie do stworzenia sieci regionalnych agregatorów będących w stanie połączyć się z Europeaną. Projekt ten był finansowany w ramach programu eContentPlus przez trzy lata, od czerwca 2008 roku. Polskim koordynatorem prac projektu było PCSS,

a prowadzony przez PCSS portal FBC doskonale wpisnął się w wypracowany przez Europeana model agregacji metadanych [Mazurek i Werla 2009, s. 10–11].

FBC agreguje obecnie metadane z ponad 80 bibliotek cyfrowych i obejmuje swoim zasięgiem kilkadziesiąt instytucji kultury i nauki z całej Polski. Łączna liczba obiektów cyfrowych, które były dostępne poprzez FBC w listopadzie 2011 roku wyniosła ponad 800 000.

Zakończenie

Digitalizacja oraz udostępnianie cyfrowych wersji dokumentów dziedzictwa kulturowego w internecie są obecnie ważnymi elementami działań głównych instytucji kultury i nauki na świecie i w Europie. Szerokie spektrum projektów oraz inicjatyw w tym kontekście wskazuje na istotne znaczenie tych działań, a podejmowane przez KE kroki dowodzą zrozumienia potrzeb finansowych z tym związanych.

W Polsce działania dotyczące digitalizacji oraz udostępniania jej efektów w internecie mają swoje początki w 2002 roku, kiedy to powstała pierwsza w Polsce profesjonalna biblioteka cyfrowa. W ślad za nią powstawały kolejne, a ich dynamiczny rozwój dał podstawy do utworzenia szeregu wspierających usług sieciowych w postaci portalu Federacji Bibliotek Cyfrowych. Kluczową rolę w tym kontekście odgrywa także oprogramowanie dLibra, które oferuje niezbędne funkcje do uruchomienia biblioteki cyfrowej i jest głównym motorem innowacji w tej dziedzinie w Polsce. dLibra jest od początku istnienia bibliotek cyfrowych w Polsce najpopularniejszym oprogramowaniem do budowy bibliotek cyfrowych.

Obecność portalu FBC została zauważona nie tylko w skali kraju, ale również na arenie międzynarodowej. Sztandarowa inicjatywa KE – Europeana – wykorzystuje informacje dostępne w Federacji Bibliotek Cyfrowych i na ich podstawie udostępnia informacje o polskich obiektach cyfrowych. Ponadto z informacji, które agreguje FBC, korzystają portale DART-Europe oraz ViFaOst.

Rozwój oprogramowania dLibra oraz portalu Federacji Bibliotek Cyfrowych to jedne z pierwszych kroków na drodze do automatyzacji procesu digitalizacji oraz udostępniania informacji w internecie. Kolejne inicjatywy, takie jak oprogramowanie dla muzeów cyfrowych dMuseion [Czyż i Romeyko-Hurko 2009, s. 21–29], system do zarządzania i monitorowania prac digitalizacyjnych dLab, stanowią następne, niejako naturalne etapy rozwoju bibliotek cyfrowych w Polsce. Również w kontekście prac badawczych i badawczo-rozwojowych, prowadzonych m.in. przez PCSS w kontekście projektu SYNAT, konstruowane są prototypowe rozwiązania, które pozwolą naukowcom i hobbystom na zaawansowaną eksplorację informacji znajdujących się m.in. w bibliotekach cyfrowych.

Bibliografia

- Comité des Sages, 2011, *The New Rainissance*, http://ec.europa.eu/information_society/activities/digital_libraries/doc/executivesummery/final_renaissance_pl.pdf [dostęp: 25.11.2011].
- Czyż, P.P., Romeyko-Hurko, M., 2009, *dMuseion: od bazy danych do muzeum cyfrowego*, Konferencja „Polskie Biblioteki Cyfrowe”, Poznań, s. 21–29.
- Górny, M., Gruszczyński, P., Mazurek, C., Nikisch, J.A., Stroiński, M., Swędrzyński, A., 2003, *Zastosowanie oprogramowania dLibra do budowy Wielkopolskiej Biblioteki Cyfrowej*. I Krajowa Konferencja Technologie Informacyjne, Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej, Technologie Informacyjne, Gdańsk, s. 109–117.
- Lewandowska, A., Mazurek, C., Werla, M., 2008, *Enrichment of European Digital Resources by Federating Regional Digital Libraries in Poland*, In Research and Advanced Technology for Digital Libraries, 12th European Conference on Digital Libraries, Aarhus, Denmark, 14–19 września, Proceedings. LNCS, vol. 5173, s. 256–259.
- Mazurek, C., 2001, *Biblioteki cyfrowe i elektroniczne wydawnictwa na przykładzie oprogramowania dLibra*, VI Krajowe Forum Informacji Naukowej i Technicznej „Informacja, wiedza, gospodarka”.
- Mazurek, C., Parkoła, T., Werla, M., 2010, *Storage, Management and Online Presentation of Multi-domain Cultural Heritage – dLibra Software Case Study*. Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej, Technologie Informacyjne, t. 18, s. 479–484.
- Mazurek, C., Stroiński, M., Werla, M., 2005, *Wdrażanie regionalnych bibliotek cyfrowych w sieci PIONIER w oparciu o środowisko dLibra*, INFOBAZY 2005 – Bazy Danych dla Nauki, Gdańsk, s. 58–64.
- Mazurek, C., Stroiński, M., Werla, M., Węglarz, J., 2008, *Infrastruktura bibliotek cyfrowych w sieci PIONIER*, Konferencja „Polskie Biblioteki Cyfrowe”, Poznań, s. 9–14.
- Mazurek, C., Werla, M., 2009, *EuropeanaLocal rola projektu w europejskiej infrastrukturze bibliotek cyfrowych*, PIONIER magazine, s. 10–11.

THE ROLE OF DIGITAL LIBRARIES IN PRESERVATION OF REGIONAL CULTURE AND TRADITION

Abstract: In the article we present the stages of building of Polish digital libraries infrastructure and the role of local and regional digital resources at the national and European level. The first chapter presents the most important issues related to online access of cultural heritage. In the second chapter we present crucial stages of the development of the Polish digital libraries network. Third chapter presents the idea and mission of the Digital Libraries Federation. The fourth chapter describes Polish digital libraries in the European context, focusing on matters related to local and regional digital resources and its location in the Europeana information space. The last chapter acts as a summary.

Keywords: Polish digital libraries, advancement in digital libraries, infrastructure for science, dLibra, Digital Libraries Federation, Europeana.

Dawid Grzegorz Węcowski, Tomasz Kaczmarek

Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki
Elektronicznej, Katedra Informatyki Ekonomicznej

Autor do korespondencji: Dawid G. Węcowski, dawid.weckowski@kie.ue.poznan.pl

ZMIENNOŚĆ TREŚCI NA FORACH INTERNETOWYCH

Streszczenie: W niniejszej pracy prezentujemy wyniki eksperymentu przeprowadzonego na próbie ponad 27 900 stron internetowych zebranych z 16 forów w odstępach 2-godzinnych (4256 niezależnych procesów pobierania) w celu zbadania, jak strony te ewoluują w czasie. Rezultaty eksperymentu mogą być podstawą do podejmowania decyzji w procesie projektowania robotów indeksujących działających w sposób przyrostowy (ang. *incremental crawler*), specjalizujących się w pozyskiwaniu dokumentów z forów internetowych w celu utrzymania wysokiego współczynnika aktualności zebranej kolekcji. Jak pokazują przeprowadzone analizy, fora internetowe różnią się od portali ogólnego przeznaczenia, a identyfikacja miejsc w ich strukturze nawigacyjnej, gdzie nowe dokumenty pojawiają się częściej, może pozwolić na podniesienie wydajności robotów indeksujących, jak również na utrzymanie wysokiego współczynnika aktualności lokalnej kolekcji dokumentów.

Słowa kluczowe: fora internetowe, zmienność treści, roboty internetowe.

Klasyfikacja JEL: D83, L86.

Wstęp

Niniejszy artykuł opisuje rezultaty eksperymentu przeprowadzonego na 16 forach internetowych, z których pozyskiwano strony internetowe. Eksperyment został przeprowadzony w celu zebrania danych na temat zmienności stron internetowych. Celem eksperymentu była walidacja wyników otrzymanych wcześniej przez podobne eksperymenty, lecz o nieco innych założeniach (brak ograniczenia do specyficznych stron i mniejsza granulacja czasowa danych). Oprócz tego eksperyment miał dostarczyć wskazówek na temat możliwych schematów dotyczących zmian na stronach internetowych.

Główną motywacją do analizy danych było poprawienie uzyskiwanych dotąd wyników prac nad architekturą robotów internetowych oraz nad szczególnymi przy-