

# Zarządzanie zawartością bibliotek cyfrowych

Paweł Gruszczyński Cezary Mazurek Stanisław Osiński  
Andrzej Swędrzyński

*Poznań Supercomputing and Networking Center,  
ul. Noskowskiego 10, 61-704 Poznań, POLAND*

*phone: +48 61 858 20 30, fax: +48 61 852 59 54  
e-mail: {grucha,mazurek,stachoo,kokosz}@man.poznan.pl*

## Streszczenie

W niniejszym artykule przedstawiamy zarządzanie zawartością bibliotek cyfrowych od strony potrzeb nauki, edukacji i kultury. Omawiamy wymagania tych trzech dziedzin stawiane narzędziom służącym publikacji elektronicznej i omawiamy, jak zostały one spełnione w środowisku biblioteki cyfrowej *dLibra*. Przedstawiamy także zakres prac, które są w toku lub zaplanowane są na najbliższy rok w celu pełnego spełnienia tych wymagań.

## 1. Wstęp

Wraz z rozwojem Internetu gwałtownie wzrosły również możliwości i potrzeby elektronicznego publikowania. Użytkownicy mają do dyspozycji coraz bardziej zaawansowane narzędzia przeznaczone do tworzenia, przeglądania i wyszukiwania elektronicznych dokumentów. Narzędzia te spełniają ważną rolę w globalnej infrastrukturze informacyjnej, mogą również wspierać pracę środowisk naukowych i edukacyjnych.

W kolejnych punktach przedstawiamy rolę biblioteki cyfrowej w nauce, edukacji i kulturze odnosząc się do cech środowiska biblioteki cyfrowej *dLibra*. Środowisko biblioteki cyfrowej *dLibra* jest rozwijane w Poznańskim Centrum SuperkomputerowoSieciowym od 1996 roku [1]. *dLibra* wspiera wszystkie etapy procesu powstawania publikacji elektronicznej poprzez wspomaganie trzech podstawowych typów użytkowników: czytelników, autorów i wydawców. Korzystając z przeglądarki internetowej czytelnik może w prosty sposób przeglądać

wybraną publikację oraz przeszukiwać katalog biblioteki. Wyszukiwarka umożliwia mu zadanie zapytania uwzględniającego różne, wielojęzyczne atrybuty metadanych (np. Dublin Core [2]). Autor za pomocą dedykowanej aplikacji może wprowadzać i pobierać z biblioteki publikacje lub ich części w celu dalszej edycji. Wydawca ma do dyspozycji oprogramowanie narzędziowe wspomagające zarządzanie całą strukturą biblioteki.

## **2. Biblioteka cyfrowa — środowisko zarządzania publikacjami naukowymi**

Od dłuższego czasu na międzynarodowym forum toczy się dyskusja na temat trudności, jakie muszą przezwyciężyć naukowcy w celu skutecznej wymiany informacji [3] [4] [5]. Biblioteka cyfrowa *dLibra* odpowiada na potrzeby środowiska naukowego poprzez udostępnienie użytkownikom zaawansowanych opcji zarządzania zawartością, dostępem i wersjami publikacji. Pakiet ten spełnia naszym zdaniem główne wymagania co do procedury publikacji dokumentów naukowych w formie elektronicznej. Najbardziej wyczerpująca lista takich wymagań została przedstawiona przez J. Kircza w [6]. Poniżej odnosimy się do tych wymagań z punktu widzenia narzędzi środowiska *dLibra*.

### **2.1 Powszechny dostęp do publikacji**

Najprostsze i najbardziej oczywiste rozwiązanie do zapewnienia czytelnikom dostępu za pośrednictwem Internetu i przeglądarki www. To wymaganie spełnia praktycznie każde narzędzie umożliwiające budowę bibliotek cyfrowych i *dLibra* nie jest tu wyjątkiem.

### **2.2 Trwałość**

Raz udostępniona publikacja naukowa nie powinna zmieniać formy (np. formatu pliku) lub lokalizacji (np. adresu w sieci). Rozwiązanie zastosowane w systemie *dLibra* polega na przydzielaniu każdemu tworzonemu dokumentowi osobnego i unikalnego identyfikatora, który pozostaje niezmienny przez cały czas istnienia biblioteki. Adres URL, pod którym dostępna jest publikacja, konstruowany jest tylko i wyłącznie na podstawie tego identyfikatora. Oznacza to, że raz zapamiętany (np. na liście *Ulubionych* w przeglądarce internetowej) zawsze będzie on prowadził do jednego i tego samego dokumentu.

### 2.3 Zarządzanie wersjami

W czasie swojego cyklu życia, dokument może istnieć w wielu wersjach. Do wstępnej wersji pracy wprowadzane są kolejne poprawki, by w końcu, po uzyskaniu akceptacji recenzenta, utworzyć ostateczną postać publikacji. Wiele dokumentów jest rozwijanych w sposób ciągły, czego doskonałym przykładem są przepisy prawne czy specyfikacje oprogramowania.

Problem wersjonowania dokumentów jest mniej widoczny w świecie publikacji papierowej, gdyż w momencie wydrukowania dokument przestaje się zmieniać. Kolejne wydania mają inne numery ISBN, inną datę wydania, często również inną formę. W dziedzinie publikacji elektronicznej sytuacja jest jednak inna, ponieważ możliwości tworzenia i publikowania kolejnych wersji dokumentów są znacznie większe.

Ponieważ publikacja elektroniczna może składać się z wielu powiązanych ze sobą modułów [7] (np. rozdziałów książki), konieczne jest zapewnienie możliwości przechowywania zarówno kolejnych wersji poszczególnych modułów, jak i komponowania kolejnych wydań publikacji ze wszystkich dostępnych w danej chwili wersji jej elementów składowych. Ponadto, raz udostępnione wydanie dokumentu nie powinno być już usuwane z biblioteki, nawet wtedy, gdy dostępne są już jego nowsze wersje.

Środowisko *dLibra* zapewnia wsparcie dla wersjonowania zarówno na poziomie modułów jak i na poziomie publikacji. Dodatkowo, istnieje możliwość zaawansowanego zarządzania wersjami dokumentu, np. tworzenia rozgałęzień w drzewie wersji. Opis szczegółowych rozwiązań zamieszczony jest w [8] i [9].

### 2.4 Zarządzanie cyklem życia publikacji

Bardzo często w proces przygotowywania publikacji zaangażowana jest więcej niż jedna osoba. Oprócz autora dokumentu może pracować nad nim współautor, recenzent lub wydawca. Mechanizmy pracy grupowej zapewniające komunikację pomiędzy wszystkimi twórcami publikacji, w znaczący sposób mogą usprawnić współpracę podczas opracowywania dokumentu.

Środowisko *dLibra*, dzięki rozbudowanemu systemowi zarządzania dostępem, umożliwia wydzielanie grup użytkowników pełniących różne role w trakcie wspólnej pracy nad publikacją. Autorzy dokumentu są uprawnieni do czytania i edycji zawartości wszystkich jego wersji. Recenzent ma możliwość odczytu nieopublikowanych jeszcze wydań publikacji, wydawca zaś prawo do udostępnienia ostatecznej postaci dokumentu szerokiej rzeszy czytelników. Istotną

rolę odgrywają tutaj również omówione w punkcie 2.6 mechanizmy komunikacji z użytkownikami biblioteki.

## **2.5 Wiarygodność**

Środowisko oraz zawartość biblioteki cyfrowej muszą być wiarygodne. Aby korzystać i odwoływać się do zawartych w niej dokumentów, czytelnik musi być pewien co do tożsamości ich autora. Powinien być także przekonany o tym, że osoby niepowołane nie wprowadzały zmian do treści publikacji. Spełnienie powyższych kryteriów leży także w interesie autora publikacji. Mógłby on ucierpieć, gdyby przypisywano mu prace niewykonane przez niego lub modyfikowano jego dokumenty.

Widać zatem, że system biblioteki cyfrowej powinien umożliwiać publikowanie tylko osobom upoważnionym i tylko po weryfikacji ich tożsamości. Co więcej, powinien zapamiętywać który użytkownik opublikował dany dokument, nie pozwalać na zmianę tej informacji oraz dopuszczać możliwość edycji publikacji tylko przez jej autora lub osoby przez niego upoważnione. System *dLibra* takie zabezpieczenia posiada.

## **2.6 Zawiadamianie o nowej publikacji lub nowej wersji publikacji**

Cały proces zarządzania zawartością biblioteki cyfrowej podporządkowany jest jednemu celowi: łatwemu odnalezieniu pewnych i interesujących czytelnika informacji. Zawiadamianie czytelnika o zmianach w zawartości biblioteki doskonale nadaje się do tego celu, gdyż dzięki temu może on śledzić na bieżąco pojawianie się nowych publikacji lub ich nowych wersji. Oczywiście jest jednak, że nie należy wysyłać informacji o wszystkich zmianach do każdego czytelnika.

Ponieważ to czytelnik jest właściwym odbiorcą treści znajdujących się w bibliotece, to on powinien decydować o tym, która część biblioteki go interesuje. Dzięki zachowaniu hierarchicznej struktury działów biblioteki, czytelnik może podać systemowi informacje na ten temat i otrzymywać jedynie zawiadomienia o zmianach w tych częściach biblioteki, które go interesują.

Inną ciekawą możliwością jest zdefiniowanie przez czytelnika zapytań do systemu w taki sposób, jakby przeszukiwał on bibliotekę. Wyniki takiego zapytania nie byłyby jednak pokazywane czytelnikowi bezpośrednio. Zapytanie to wykonywałby system okresowo i — podobnie jak to ma miejsce w przypadku interesujących czytelnika działów — wysyłał tylko listę zmian, które zaszły w bibliotece od czasu wysłania poprzedniego zawiadomienia.

W systemie *dLibra* takie zawiadomienia mogą być wysyłane do czytelnika w postaci listu *e-mail* lub też w formie wewnętrznej wiadomości biblioteki, która jest widoczna po zalogowaniu się do systemu. Dzięki wybraniu tej drugiej metody, czytelnik może uniknąć otrzymywaniu dużej liczby listów do skrzynki i skoncentrować się na czytaniu o zmianach wtedy, kiedy ma on na to ochotę i czas.

### **3. Biblioteka cyfrowa — usługi dla społeczeństwa informacyjnego**

Zastosowania biblioteki cyfrowej nie ograniczają się do publikacji naukowych. Szczególnie pożyteczną rolę może ona pełnić w przechowywaniu dziedzictwa kultury oraz edukacji. Dotyczące tych dziedzin aspekty zarządzania biblioteką cyfrową przedstawiamy poniżej.

#### **3.1 Przechowywanie dziedzictwa kultury w formie cyfrowej**

Jednym z najważniejszych projektów służących zachowaniu dzieł literatury pięknej oraz innych publikacji książkowych jest *Projekt Gutenberg* [10]. W ramach tego projektu stworzona została baza książek w postaci czysto tekstowej, ze względu na prawa autorskie głównie tych napisanych przed rokiem 1923. System *dLibra* jest przygotowany na przechowywanie danych multimedialnych, a więc także obrazu i dźwięku. Trwają także w chwili obecnej prace nad tym, aby materiały umieszczone w bibliotece były chronione przed kopiowaniem. Dzięki temu będzie można w środowisku *dLibra* przechowywać i udostępniać czytelnikom dzieła w postaci cyfrowej, które chronione są prawami autorskimi.

Dzięki rozbudowanym możliwościom opisywania znajdujących się w bibliotece dzieł metadanymi oraz wyszukiwania pełnotekstowego, znalezienie interesującej książki w postaci cyfrowej jest o wiele łatwiejsze, niż jej papierowego odpowiednika.

Digitalizacja dzieł sztuki daje także unikalną możliwość prezentacji szerokiemu gronu odbiorców dokumentów, które w innej formie nie ujrzałyby światła dziennego, np. starodruki. Innym przykładem zastosowania biblioteki cyfrowej jest zarządzanie zasobami dla usługi typu wirtualne muzeum.

#### **3.2 Biblioteka cyfrowa w edukacji**

Rozwój technologii informatycznych przyniósł nauczycielom szerokie możliwości wymiany doświadczeń i pomysłów. Nauczyciele są zainteresowani m. in. wymianą scenariuszy lekcyjnych, zadań i sposobów na uatrakcyjnienie lekcji. Część z nich udostępniła swoje prace w Internecie, jednakże rozproszenie tych materiałów na osobiste strony, brak standardowych formatów dokumentów i informacji na temat tych serwisów znacząco utrudnia korzystanie z — często wysokiej jakości — publikacji. Dodatkowym problemem jest fakt, że duża część nauczycieli, szczególnie przedmiotów humanistycznych, chciałaby mieć dostęp do tych materiałów w formie wydruków.

Oczywiście rozwiązaniem problemu dostępności i trudności ze znalezieniem jest utworzenie centralnego repozytorium materiałów edukacyjnych w postaci biblioteki cyfrowej. Zapewnia ona jasny i logiczny podział publikacji na przedmioty, których dotyczą, umożliwi łatwe przeszukiwanie materiałów oraz bieżące śledzenie napływających dokumentów. Materiały takie zbierane są już w chwili obecnej w bibliotece cyfrowej *Interklasy* [11].

Drugi ze wspomnianych problemów jest nieco trudniejszy. W celu jego rozwiązania prowadzone są prace nad prostym edytorem, który umożliwi edycję treści publikacji o ściśle ustalonej strukturze. Edytor ten generuje plik w formacie XML, który następnie jest konwertowany za pomocą specjalnych arkuszy styli do formatów HTML i PDF. Dzięki temu edytorowi raz stworzony dokument może być prezentowany na WWW, jak i w łatwy sposób wydrukowany do bezpośredniego wykorzystania dla nauczycieli.

Należy również zauważyć rolę, jaką biblioteki cyfrowe pełnić mogą dla uczniów. Widoczne jest to szczególnie w przypadku prostych narzędzi, łagodnie wprowadzających dzieci w świat komputerów. W Polsce rozwinięty został cały pakiet służący do łatwej edycji tekstu i grafiki, wykonywania prostych obliczeń czy przechowywania danych o nieskomplikowanej strukturze [12]. Podobnie planujemy wprowadzić do środowiska *dLibra* prosty edytor publikacji elektronicznej. Za jego pomocą w łatwy sposób dzieci będą mogły zaprojektować swoją stronę www, składając ją z dostępnych komponentów na podstawie szablonów z użyciem własnych tekstów i rysunków.

Zintegrowanie tego narzędzia z środowiskiem biblioteki cyfrowej uprości maksymalnie proces publikacji dokumentu w formie elektronicznej. Umożliwi to uczniom łatwe wyszukiwanie prac rówieśników o podobnych zainteresowaniach, a co za tym idzie, zachęci do nawiązywania kontaktów. Ponieważ publikacje będą znajdowały się w dostępnym łatwo i znanym przez wiele osób miejscu, zachęci to uczniów do twórczego podejścia do korzystania z technologii informatycznych.

## 4. Podsumowanie

Biblioteka cyfrowa może być środowiskiem narzędziowym znacząco wspomagającym rozwój nauki, edukacji i kultury. Szczególnie przydatne w każdej z tych dziedzin są takie cechy, jak scentralizowane repozytorium publikacji, hierarchiczna struktura katalogu, łatwe przeszukiwanie zawartości, zarówno na podstawie metadanych, jak i pełnotekstowe. Ważne są także te możliwości systemu biblioteki, które wspierają użytkownika w samym procesie tworzenia i publikacji dokumentów, jak zarządzanie wersjami i cyklem życia, wyspecjalizowane narzędzia do edycji.

Warto także zwrócić uwagę na to, że w praktycznie każdej dziedzinie życia zbierana jest wiedza, a dostęp do niej w formie elektronicznej jest znacznie łatwiejszy niż przy wykorzystaniu standardowych sposobów publikacji. Znaczenie bibliotek cyfrowych będzie zatem rosnąć, a zarządzanie zawartością może się okazać kluczową kwestią dla ich wdrożenia w różnych dziedzinach życia społeczeństwa informacyjnego.

### Bibliografia

- [1] Mazurek C., Szuber S.: Development of Digital Libraries at Poznań Supercomputing and Networking Center. EUNIS, Poznań, Kwiecień 2000.
- [2] Dublin core metadata element set, version 1.1: Reference description.  
<http://dublincore.org/documents/dces/>.
- [3] International Working Group. Defining and Certifying Electronic Publication in Science. A Proposal to the International Association of STM Publishers,  
<http://associnst.ox.ac.uk/icsuinfo/aaas-stm.htm>.
- [4] Shaw D., Moore H., editors. *Proceedings of the Joint ICSU Press/UNESCO Conference on Electronic Publishing in Science*, Paris, Luty 1996. UNESCO,  
<http://associnst.ox.ac.uk/~icsuinfo/ConfProc.htm>.
- [5] International council for science.  
<http://associnst.ox.ac.uk/~icsuinfo/>.
- [6] Kircz J. G. New practices for electronic publishing: how to maintain quality and guarantee integrity. In Shaw D., editor, *Proceedings of the Second ICSU-UNESCO International Conference on Electronic Publishing in Science*. UNESCO, Paris, Luty 2001.  
<http://associnst.ox.ac.uk/~icsuinfo/proc01fin.htm>.

- [7] Harmsze F. *A modular structure for scientific articles in an electronic environment*. PhD dissertation, University of Amsterdam, 2000.  
<http://www.science.uva.nl/projects/commphys/papers>.
- [8] Gruszczyński P., Mazurek C., Osiński S., Swędryński A., Szuber S. *dLibra Content Maintenance for Digital Libraries*. Euromedia WEBTEC, Modena, Kwiecień 2002.
- [9] Mazurek C. Biblioteki cyfrowe i elektroniczne wydawnictwa na przykładzie oprogramowania *dLibra*. IV Krajowe Forum Informacji Naukowej i Technicznej *Informacja, wiedza, gospodarka*, PTIN, Katowice (Polska), Wrzesień 2001.
- [10] Projekt Gutenberg. <http://promo.net/pg>.
- [11] Projekt Interklasa. <http://www.interklasa.pl>.
- [12] Sysło M. Edukacja na rzecz społeczeństwa informacyjnego — ocena stanu, trendy, perspektywy. SIEDEMNASTE JESIENNE SPOTKANIA PTI, Mrągowo, Listopad 2001.