

Małgorzata Kowalska
Instytut Informacji Naukowej i Bibliologii UMK

Zasoby czasopiśmiennicze w polskich bibliotekach cyfrowych i bibliograficznych bazach danych - koegzystencja, konkurencja czy kooperacja?

Abstrakt

Bibliograficzne bazy danych, a szczególnie dziedzinowe bazy danych, stanowią nieodzowny element dla nauki i badań. Rejestracja specjalistycznych czasopism czy niskonakładowych materiałów konferencyjnych czyni je kompleksowym i aktualnym źródłem informacji na temat tendencji rozwojowych w poszczególnych dziedzinach i dyscyplinach naukowych. Na skutek gwałtownego rozwoju bibliotek i repozytoriów cyfrowych, a zwłaszcza postępującej w ostatnich latach digitalizacji współczesnych zasobów czasopiśmienniczych, bibliograficzne bazy danych nie mogą już jednak pretendować do miana jedyne źródła informacji o istnieniu danej publikacji. Coraz częściej pojawia się zatem pytanie o zasadność i celowość ich tworzenia. W artykule podjęto próbę analizy obecności czasopism znajdujących się w wybranych bibliotekach cyfrowych, a rejestrowanych także przez dziedzinowe bazy danych. Autorka stara się wskazać obszary i tendencje wzajemnego przenikania obydwu źródeł informacji, odpowiedzieć na pytanie o stopień wykorzystania zasobów czasopiśmienniczych z nich pochodzących oraz określić rozbieżności interesów i polityk w zakresie ich tworzenia, tak by uniknąć ich w przyszłości.

Słowa kluczowe

bibliograficzne bazy danych, pełnotekstowe bazy danych, biblioteki cyfrowe, repozytoria cyfrowe, czasopisma dziedzinowe, koegzystencja zasobów informacyjnych

Abstract

Journal resources in polish digital libraries and bibliographic databases – coexistence, competition and/or cooperation? The bibliographic databases, especially for the domain ones, determine indispensable element of science and research. Registration of subject journals or the small edition conference materials make them the complex and actual source of information about the present tendencies in particular domain and field of study as well. As a result of the library and digital repository rapid development, in particular digitalization of journals and magazines which develops lately, the bibliographical databases can not pretend to be the only source of information about the existence of specific publication any longer. More and more frequently emerge questions about the validity and usefulness of their creation and maintenance. In this paper author has made an effort to analyze the presence of journals preserved in digital libraries and also registered by the domains databases at the same time. Author tried to indicate the fields and tendencies of both sources mutual permeability then is trying to answer questions to what degree the holding of magazines are used and finally is defining the conflict of interests and database creation policy that ought to be avoid in the future.

Keywords

bibliographic databases, full-text databases, digital libraries, digital repositories, subject journals, coexistence of information resources

Zasoby cyfrowe jako element infrastruktury nauki

Od zarania dziejów integralnym elementem każdego procesu komunikacji był przekaz informacji. O ile do końca XIX wieku możliwości jego szybkiej realizacji były dość ograniczone, bo uwarunkowane zasięgiem głosu i różnorodnych systemów sygnalizacyjnych, pojemnością i zawodnością pamięci czy kapryсами aury, to wraz z wiekiem XX rozpoczęły się tak daleko idące zmiany, że dotychczas stosowane formy obiegu informacji okazały się całkowicie bezużyteczne. Największe przeobrażenia w krajobrazie informacyjnym wywołały technologie informacyjno-komunikacyjne, dzięki którym zniknął problem fizycznego oddalenia odbiorcy informacji od jej źródła, a samo dostarczanie informacji stało się łatwe i szybkie. Te nowe cechy przekazu spowodowały, że nowoczesne technologie znalazły swoje zastosowanie we wszystkich dziedzinach aktywności człowieka, w tym także w komunikacji naukowej. Z czasem przedstawiciele

środowiska naukowego zaczęli się porozumiewać ze sobą za pomocą poczty elektronicznej, uczestniczyć w telekonferencjach i publikować wyniki własnych badań na łamach źródeł elektronicznych. Wkrótce nieodłącznym atrybutem komunikowania w świecie nauki stały się także bazy danych, rozproszone zasoby obliczeniowe, wirtualne laboratoria i repozytoria cyfrowe.

Dziś, gdy nikogo nie dziwi już powszechna dostępność infrastruktury sieciowej, a przydatności techniki cyfrowej nikt nie kwestionuje, pojawiają się inne problemy. Przedmiotem dyskusji staje się choćby zjawisko współlistnienia zasobów oferujących środowisku akademickiemu podobne, a niejednokrotnie te same źródła informacji naukowej. W związku z tym coraz częściej stawia się pytania o sens podejmowania nowych inicjatyw cyfrowych czy celowość dublowania prac w ramach już istniejących. Pojawiają się także apele o „wolną naukę działającą w trybie on-line”, a co za tym idzie szerokie i wielopoziomowe rozpowszechnianie treści naukowych. Z uwagi na brak neutralności wielu narzędzi informatycznych (różne systemy filtrujące i mechanizmy rankingu stron WWW ograniczają w pewnym sensie dostęp do informacji i arbitralnie rozstrzygają, co jest informacją złą, a co dobrą[1]), argumenty zwolenników ruchu wolnego dostępu wydają się po części racjonalne. Swoje uzasadnienie znajdują jednak także obawy wynikające z obserwacji gwałtownego przyrostu zasobów cyfrowych, powielania ich zawartości czy braku wzajemnych powiązań między nimi. Ponieważ zjawiska te nie wpływają na podniesienie jakości informacji sieciowej, a wielokrotnie przyczyniają się nawet do powstawania zasobów informacji niezamawianej, użytkownicy Internetu, zwłaszcza „pokolenia Google”, borykają się nie tylko z problemem wyszukiwania informacji, ale także braku umiejętności jej krytycznej oceny. W tym kontekście rodzą się więc pytania o rzeczywisty sens tworzenia poszczególnych archiwów danych, merytoryczny poziom ich zawartości, stopień wykorzystania czy wreszcie stan dostosowania do potrzeb odbiorców. Tego typu rozważaniom w odniesieniu do polskich bibliotek cyfrowych i bibliograficznych baz danych poświęcony jest niniejszy artykuł.

Baza danych, repozytorium, biblioteka cyfrowa – natura zasobów

Czasopisma dziedzinowe, teksty artykułów naukowych, pomoce dydaktyczne, dokumenty źródłowe i ikonograficzne oraz wiele innych rodzajów materiałów cyfrowych to rzeczywistość współczesnej nauki. Choć wszystkie z nich postrzega się jako źródła niezwykle przydatne i pomocne, to niejednokrotnie ich podobna struktura i zawartość wprawiają potencjalnego użytkownika w zakłopotanie, nie jest on bowiem w stanie stwierdzić, z jakim typem źródła, zasobu czy narzędzia[2] ma do czynienia. Dotyczy to także baz danych, repozytoriów i bibliotek cyfrowych.

Pierwsze problemy pojawiają się już przy próbie zdefiniowania ich istoty. Samo traktowanie zasobów projektu „Gutenberg” przez użytkowników jako najstarszego repozytorium tekstów, bazy publikacji z domeny publicznej, a wreszcie pierwszej biblioteki cyfrowej dowodzi, że literatura zawodowa jeszcze nie do końca wchłonęła te nowe zjawiska. Zrozumiałe wydają się więc nieustannie powracające wątpliwości, czy baza danych to wyłącznie zbiór rekordów bibliograficznych uzupełnionych abstraktami, czy może repozytorium różnego typu materiałów udostępnianych drogą elektroniczną; czy repozytorium to tylko magazyn informacji, czy może baza danych rejestrująca metadane o publikacjach i ich zawartości; czy biblioteka cyfrowa to jedynie kolekcja materiałów zdigitalizowanych, czy może system rozproszonych baz danych? By odpowiedzieć na tego typu pytania, warto przyjrzeć się podstawowym własnościom wymienionych zbiorów informacji. Ich wybór prezentuje tab. 1.

Własności	Baza danych	Repozytorium	Biblioteka cyfrowa
Definicje	<ul style="list-style-type: none"> zbiór wzajemnie powiązanych danych opisujących pewien 	<ul style="list-style-type: none"> centralny „magazyn informacji” przeszłych danych, 	<ul style="list-style-type: none"> zasoby przeszukiwalne za pomocą techniki

	<p>wybrany fragment rzeczywistości, wyposażony w oprogramowanie umożliwiające ich definiowanie, wykorzystywanie i modyfikowanie[3];</p> <ul style="list-style-type: none"> • uporządkowany zbiór informacji (opisowych i content units) z określonej dziedziny lub tematyki, ze wspólnym interfejsem użytkownika i oprogramowaniem przeznaczonym do wyszukiwania i przetwarzania danych[4]. 	<p>nieulegających już zmianie[5];</p> <ul style="list-style-type: none"> • „magazyn” wszelkiego typu dokumentów: elektronicznych obrazów zeskanowanych dokumentów papierowych, dokumentów elektronicznych, danych z systemów informatycznych, plików dźwiękowych i wideo[6]; • miejsce uporządkowanego przechowywania dokumentów, z których wszystkie są przeznaczone do udostępniania[7]; • magazyn najrozmaitszych zasobów cyfrowych[8]; • system udostępniający metadane na temat obiektów cyfrowych[9]; • baza danych, w której są przechowywane metadane o publikacjach (m.in. rekordy bibliograficzne) oraz zawartości publikacji[10], • baza danych wymagań systemu[11]. 	<p>komputerowej;</p> <ul style="list-style-type: none"> • zasoby udostępniane lokalnie i zdalnie; • punkt dostępu do wielu źródeł informacji naukowej wraz z usługami instruktażowymi; • kolekcja materiałów zdigitalizowanych lub kodowanych cyfrowo; • biblioteka posiadająca dostęp do zasobów Internetu i zbiory CD-ROM; • agencja, która jednoczy instytucje dla zapewnienia dostępu do informacji elektronicznej; • biblioteka skanująca swoje materiały[12]; • archiwum cyfrowe[13]; • repozytorium wiedzy[14]; • baza danych, chroniona raczej prawem sui generis niż autorskim[15].
<p>Cechy charakterystyczne</p>	<ul style="list-style-type: none"> • trwałość danych; • rozmiar danych; • złożoność danych; • zarządzanie danymi; • archiwizacja danych; • udostępnianie danych[16]. 	<ul style="list-style-type: none"> • trwałość danych; • rozmiar danych; • złożoność danych, • zarządzanie zasobami; • archiwizacja zasobów; • udostępnianie zasobów; • wykorzystanie usług 	<ul style="list-style-type: none"> • trwałość danych; • rozmiar danych; • złożoność danych, • cyfrowe gromadzenie; opracowanie, udostępnianie; • architektura silnie rozproszona; • otwarty standard

		wspólnych – sieciowy profil użytkownika[17].	komunikacji; <ul style="list-style-type: none"> • skoordynowane zarządzanie zasobami; • wykorzystanie usług wspólnych – sieciowy profil użytkownika[18].
Wymagania/funkcje	<ul style="list-style-type: none"> • gwarantowanie spójności danych; • zapewnienie bezpieczeństwa i autentyczności danych; • zapewnienie efektywnego przetwarzania danych; • właściwe odzwierciedlanie obiektów i zależności między nimi; • ochrona przed nieautoryzowanym dostępem do danych; • przestrzeganie prawa autorskiego; • zapewnienie jednoczesnego dostępu do danych wielu użytkownikom; • zapewnienie dostępu do źródeł informacji kwalifikowanej; • udostępnianie metadanych; • zachowanie standardów międzynarodowych; • zapewnienie trwałości organizacyjnej i finansowej[19]. 	<p>jak w przypadku baz danych oraz:</p> <ul style="list-style-type: none"> • przyjęcie odpowiedzialności za długookresowe przechowywanie zasobów cyfrowych; • zapewnienie metod ewaluacji systemu; • zapewnienie mierzalności i oceny polityk i praktyk przechowywania informacji; • określenie rodzaju i zakresu przechowywanych danych cyfrowych; • zapewnienie możliwości składowania danych przez różnego rodzaju dostawców; • zapewnienie możliwości wyszukiwania informacji zapisanej w wielu, często odmiennych, językach[20]. 	<p>jak w przypadku baz danych i repozytoriów oraz:</p> <ul style="list-style-type: none"> • zapewnienie możliwości wyszukiwania rozproszonego, umożliwiającego przeszukiwanie zawartości różnych bibliotek cyfrowych za pomocą jednego narzędzia; • zapewnienie możliwości przeszukiwania pełnotekstowego publikacji; • wspieranie procesu e-learningu; • przełamywanie barier dotyczących cyfrowego wykluczenia; • promowanie kultury narodowej; • wspieranie rozwoju infrastruktury społeczeństwa informacyjnego[21].
Typologia	<p>Ze względu na:</p> <ul style="list-style-type: none"> • rodzaj gromadzonej informacji, np. bibliograficzne i faktograficzne; • rodzaj denotacji utrwalonych danych, np. faktograficzne, 	<p>Ze względu na:</p> <ul style="list-style-type: none"> • typ gromadzonych materiałów, np. eprintów, prac badawczych, artykułów z czasopism, born digital, counterparts; 	<p>Ze względu na:</p> <ul style="list-style-type: none"> • podmiot nadzorujący, np. instytucjonalne, konsorcjalne, pojedyncze, współpracujące, scalone;

	<ul style="list-style-type: none"> źródłowe, odsyłające; zakres tematyczny, np. jedno- i wielod dziedzinowe; podstawę znakowania, np. alfa-numeryczne, numeryczne, tekstowe, ikonograficzne; organizację danych, np. logiczne, fizyczne; przyjęty model danych, np. hierarchiczne, sieciowe, relacyjne; zakres realizacji funkcji wyszukiwawczej w charakterystyce wyszukiwawczej, np. pełnotekstowe; przeznaczenie, np. zadaniowe, informatorowe, adresowe, programowe[22]. 	<ul style="list-style-type: none"> zakres tematyczny, np. dziedzinowe, tematyczne; podmiot nadzorujący, np. instytucjonalne, osobowe, wydziałowe; zasięg terytorialny, np. narodowe, regionalne, lokalne. 	<ul style="list-style-type: none"> zastosowane technologie, np. I, II i III generacji; typ gromadzonych materiałów, np. dziedzictwa narodowego, interdyscyplinarne, regionalne.
Dostęp	<ul style="list-style-type: none"> za pośrednictwem sieci komputerowych; w trybie wsadowym (CD-ROM, dyskietki). 	za pośrednictwem sieci komputerowych.	za pośrednictwem sieci komputerowych.
Przykłady	BazTech, Bazy BN, SympoNet, Arton, AGRO, SIGŻ, PEDAGOG, EDUKACJA, BazHum, PBL, ERIC, LISTA, MEDLINE, INSPEC itp.	ePrints, ROAR, E-LIS, OpenDOAR, eScholarship itp.	biblioteki cyfrowe dLibry, Digitalni Knihovna, Digitale Sammlungen, Europeana, American Memory Historical Collections itp.

Źródło: opracowanie własne.

Jak nietrudno zauważyć, określenie natury baz danych, repozytoriów i bibliotek cyfrowych nie jest prostym zadaniem. Jeszcze do połowy lat 80. XX wieku, gdy wśród baz danych dominowały systemy scentralizowane zarządzane przez jeden komputer, w których rejestrowane dane były przede wszystkim opisami obiektów rzeczywistych, można było dokonywać klarownej kategoryzacji narzędzi i źródeł elektronicznych. Obecnie staje się to coraz większym problemem, dochodzi bowiem do zacierania różnic między nimi. Rozwój nowych technologii sprawia, że bazy danych coraz częściej przejmują pewne cechy charakterystyczne dla innych zasobów elektronicznych. By móc z nimi konkurować, a jednocześnie lepiej odzwierciedlać rzeczywisty

świat, wiele baz danych nie ogranicza się już do udostępniania „suchych” opisów obiektów, lecz oferuje sprawne zarządzanie danymi niesformatowanymi, np. tekstami, obrazami, mapami czy sygnałami dźwiękowymi. Tendencja przekształcania się baz danych w bardziej kompleksowe narzędzia informacyjne jest widoczna zwłaszcza w przypadku pełnotekstowych baz danych, które dzięki gromadzeniu kopii oryginalnych wydawnictw zwartych, sprawozdań, artykułów z czasopism, druków muzycznych i dokumentów kartograficznych[23], pretendują raczej do miana repozytoriów tekstów niż zbiorów obiektów fakto- czy bibliograficznych, a tym samym mogą śmiało konkurować z repozytoriami dziedzinowymi czy też bibliotekami cyfrowymi. Przyglądając się repozytoriom i bibliotekom cyfrowym, trudno z kolei nie dostrzec w ich naturze cech charakterystycznych dawniej wyłącznie dla baz danych, takich jak: trwałość i integralność danych, ich złożoność i duży rozmiar wolumenu, umożliwienie jednoczesnego dostępu do danych wielu użytkownikom, istnienie powiązań o charakterze strukturalnym i semantycznym. Mające miejsce w repozytoriach czy bibliotekach cyfrowych przechowywanie danych w różnych, oddalonych od siebie komputerach czy połączenie ich odpowiednimi mechanizmami konsolidującymi pozwala postrzegać je nawet jako swoistego rodzaju systemy rozproszonych baz danych[24].

Niestety, uchwycenia kluczowych różnic pomiędzy wymienionymi zasobami nie ułatwiają także ani misja wypełniana przez nie w społeczeństwie, ani realizacja zadań podstawowych. Wszystkim przypisać można przecież informowanie o dorobku naukowym, dbałość o bezpieczeństwo i autentyczność danych, przestrzeganie prawa autorskiego, czy wreszcie w kontekście globalnym – promowanie polskiej nauki i propagowanie otwartego dostępu do wiedzy. Dokonując kategoryzacji, warto wziąć jednak pod uwagę pewne cechy szczególne, jak choćby kompletność zasobów, stopień ich rozproszenia, poziom sformalizowania (zinstytucjonalizowania), zróżnicowanie pod względem treściowym, zakres realizacji funkcji wyszukiwawczych i usługowych, sposób dostępu.

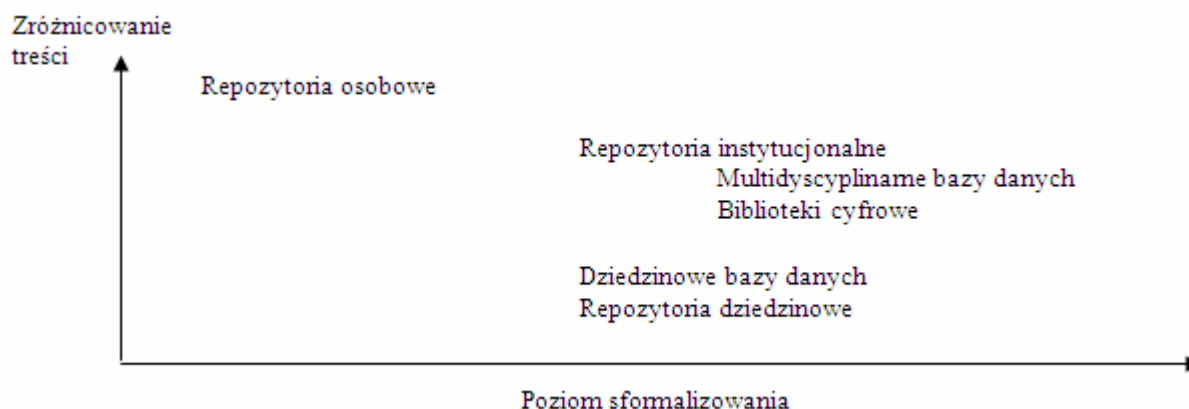
Bez wątpienia w przypadku każdego z omawianych zbiorów informacji zdefiniowanie rodzaju i zakresu przechowywanych danych odgrywa istotną rolę, a sam **proces gromadzenia** materiałów przebiega zgodnie z przepisami obowiązującego prawa. Celem tworzenia większości baz danych jest zgromadzenie w jednym miejscu w miarę kompletnego wykazu piśmiennictwa z wybranych dziedzin wiedzy (np. bazy dziedzinowe) lub rejestracja konkretnych typów wydawnictw (np. bibliografie narodowe). W przypadku wielu dziedzinowych baz danych można zatem mówić o tworzeniu swoistych dziedzinowych przewodników, uporządkowanych według kategorii i pretendujących do roli serwisów tematycznych o kontrolowanej jakości typu subject gateway. Biblioteki cyfrowe trudno postrzegać jako podobne zbiory informacji kwalifikowanej. Mimo że ich twórcom nie można zarzucić braku celowości gromadzenia materiałów, to mało zasadne wydaje się włączanie do ich kolekcji wybranych fragmentów piśmienniczych (np. kilku numerów danego czasopisma, pojedynczych artykułów z wydawnictw zbiorowych, pre- i postprintów) czy dokonywanie digitalizacji pewnych wydawnictw tylko dlatego, że znajdują się one w zasobach cyfrowych innych instytucji.

Biorąc pod uwagę kwestię **rozproszenia zasobów**, nie trudno zauważyć, że w bazach danych zjawisko to nie przybiera takich rozmiarów jak w przypadku bibliotek czy repozytoriów cyfrowych. Nawet jeśli opracowywanie rekordów do baz danych następuje w ujęciu instytucjonalnym, w ośrodkach współpracujących, to ostatecznie zbiór konkretnej bazy danych udostępniany jest z reguły na jednym serwerze centralnym (np. dla BazTechu – ICM UW). W przypadku bibliotek i repozytoriów cyfrowych dane są wprawdzie magazynowane również na lokalnych serwerach, ale jednocześnie stanowią element bardziej rozproszonej struktury sieciowej (np. Federacja Polskich Bibliotek Cyfrowych, dalej: FBC).

Dokonując różnicowania zasobów pod względem **poziomu ich sformalizowania** (czas przeznaczony na opracowanie i obsługę, dostępność sprzętu, finansowanie), można pokusić się o stwierdzenie, że największym stopniem zinstytucjonalizowania charakteryzują się biblioteki

cyfrowe i multidyscyplinarne bazy danych. Ze względu na uczestnictwo w ich tworzeniu wielu instytucji, wielkość i różnorodność gromadzonych danych, wymagają one największych inwestycji o charakterze kapitałowym, technologicznym i organizacyjnym. Mniejszych nakładów wymagają repozytoria instytucjonalne i dziedzinowe oraz dziedzinowe bazy danych, rejestrujące z reguły przygotowane już pod względem merytorycznym i technicznym dane, uporządkowane w sposób zhierarchizowany zgodnie ze strukturą organizacyjną danej uczelni czy systemem klasyfikacyjnym konkretnej dyscypliny naukowej. Najniższym poziomem sformalizowania cechują się z kolei repozytoria osobowe, gromadzące prace danego autora, a zorganizowane i archiwizowane zgodnie z potrzebami właściciela danego zasobu.

Odmienne wnioski można natomiast sformułować w odniesieniu do **zróżnicowania treści** analizowanych źródeł. Najbogatsze pod tym względem wydają się właśnie repozytoria osobowe, rejestrujące różnego typu obiekty (artykuły, fotografie, mapy, sprawozdania), będące przejawem różnorodnej aktywności ich właścicieli. Na drugim miejscu plasują się repozytoria instytucjonalne, gromadzące i przechowujące dorobek intelektualny społeczności naukowej jednego lub wielu uniwersytetów, których zakres tematyczny zależy jednak od profilu działalności uczelni. Na kolejnym miejscu można ułożyć biblioteki cyfrowe i multidyscyplinarne bazy danych, które gromadzą wprawdzie różnorodne materiały pod względem tematycznym, lecz swoim zakresem nie obejmują tak wielu dyscyplin wiedzy, jak w przypadku repozytoriów instytucjonalnych, a jedynie ich wybór. Jeszcze mniejszym zróżnicowaniem pod tym względem cechują się repozytoria dziedzinowe i dziedzinowe bazy danych, ograniczające swój zasób do materiałów pochodzących z jednej lub kilku dyscyplin naukowych. Stopień zróżnicowania zasobów pod względem treściowym i instytucjonalnym obrazuje rys. 1.



Rys. 1. Źródła elektronicznej informacji naukowej

Źródło: opracowanie własne na podstawie: NAHOTKO, M. Cyfrowa nauka-cyfrowe publikacje-cyfrowe biblioteki. *Przegląd Biblioteczny* 2007, z. 1, s. 18.

Analizując możliwości wyszukiwawcze omawianych narzędzi, nietrudno przyznać, że ich największy wybór oferują biblioteki i repozytoria cyfrowe. Poza wyszukiwaniem według kryteriów formalnych i rzeczowych wiele z nich pozwala także na przeszukiwanie treści publikacji (nie chodzi przy tym o opcję „znajdź tekst w dokumencie PDF lub DjVu”). W przypadku baz danych funkcję taką realizują bazy pełnotekstowe, choć także nie wszystkie. Wiele z nich umożliwia zastosowanie opcji „search in all text” lub „poszukaj w tekście” (np. bazy EBSCO), ale istnieją także takie, które dopuszczają przeszukiwanie zasobów wyłącznie z wykorzystaniem elementów pochodzących z rekordów bibliograficznych czy abstraktów artykułów, sprowadzając operacje wykonywane na pełnym tekście publikacji do jej przeglądania. Jeszcze mniejsze możliwości oferują pozostałe typy baz danych. Mimo zastosowania różnorodnych opcji wyszukiwawczych (proste, rozszerzone, eksperckie), dostęp do ich zawartości odbywa się głównie z wykorzystaniem podstawowego zestawu kryteriów formalnych oraz wybranych kryteriów rzeczowych. Niewątpliwą

zaletą bibliotek i repozytoriów cyfrowych jest możliwość wyszukiwania rozproszonego, pozwalającego na przeszukiwanie zawartości wielu zasobów za pomocą jednego narzędzia. W przypadku baz danych taką usługę realizują jedynie systemy informacyjne on-line (np. DIALOG), specjalne platformy bazodanowe (np. EBSCOhost) lub zintegrowane systemy biblioteczne (np. Aleph). Za dodatkową funkcjonalność bibliotek i repozytoriów cyfrowych należy także uznać udostępnianie użytkownikom zaawansowanych usług sieciowych, w tym sieciowego profilu użytkownika, umożliwiającego logowanie do wszystkich powiązanych zasobów i zarządzanie zebranymi informacjami w zindywidualizowany sposób. Choć podobne rozwiązanie stosuje także wiele zagranicznych baz danych (EBSCO, DIALOG), w polskich bazach nie jest ono jeszcze powszechnym zjawiskiem (indywidualne zarządzanie danymi bibliograficznymi umożliwiają obecnie jedynie bazy udostępniane w ramach systemu ALEPH [funkcja „koszyk” lub/i „e-półka”], np. SYMPOnet, EDUKACJA, PEDAGOG, bazy Biblioteki Sejmowej).

Ostatnim, omawianym tu, elementem różnicującym analizowane źródła elektroniczne jest sposób dostępu do nich. Bez wątplenia warunkiem sine qua non istnienia repozytoriów czy bibliotek cyfrowych jest sieciowy dostęp do ich zasobów, realizowany zarówno w ramach wewnętrznej, jak i zewnętrznej infrastruktury. W przypadku baz danych dodatkową formę dostępu stwarzają płyty CD-ROM.

Podsumowując rozważania o naturze wybranych zasobów elektronicznych, warto jeszcze raz podkreślić, że dokonanie pełnej i niebudzącej wątpliwości ich kategoryzacji jest zadaniem trudnym i w dużej mierze subiektywnym. O ile w przypadku bibliotek i repozytoriów cyfrowych można jeszcze mówić o pewnym pokrewieństwie, czego potwierdzeniem jest także synonimiczne stosowanie obu terminów, to już porównanie własności tych zasobów z bazami danych pozwala na uwydatnienie pewnych wyraźnych cech wszystkich omawianych zbiorów informacji.

Zawartość polskich bibliotek cyfrowych

Statystyki FBC wskazują, że na dzień 15 kwietnia 2009 r. łączna liczba publikacji w polskich bibliotekach cyfrowych przekroczyła 211 tysięcy. Najczęściej występującymi typami zasobów były przy tym: zasób edukacyjny (skrypty, podręczniki i monografie naukowe), zasób dziedzictwa kulturowego (wybrane zabytki piśmiennictwa), regionalia (piśmiennictwo dotyczące regionów) i muzykalia (nuty i piśmiennictwo związane z muzyką). W kilku bibliotekach wydzielono także tzw. kolekcje specjalistyczne, w których umieszczono kopie cyfrowe wszystkich typów dokumentów znajdujących się w zbiorach bibliotecznych instytucji współpracujących, a wyselekcjonowane ze względu na ich wartość historyczną, zabytkową lub merytoryczną.

Dane zamieszczone na stronach internetowych poszczególnych bibliotek cyfrowych pozwalają stwierdzić, że pod względem formalnym dominują w nich czasopisma. Ich odsetek obejmuje od 50% do 85% wszystkich udostępnianych publikacji. Na dalszych miejscach pod względem liczebnym znajdują się gazety, książki, grafiki, mapy, fotografie – z różną liczbą pozycji w poszczególnych bibliotekach.

Ogółem we wszystkich bibliotekach należących do FBC pod kategorią „czasopismo” zarejestrowano około 67 tysięcy pozycji (stan na dzień 15 kwietnia 2009 r.). Choć jest to liczba imponująca, warto pamiętać, że nie jest to jednak suma zaindeksowanych tytułów wydawnictw periodycznych, a jedynie lista publikacji spełniających warunek „typ zasobu = czasopismo” (jako odrębną pozycję traktuje się tu pojedynczy numer lub rocznik danego czasopisma). Bez względu na sposób inwentaryzacji pozycji w bibliotekach cyfrowych, dane statystyczne dowodzą dużego zainteresowania wydawnictwami periodycznymi, które należą do pozycji najczęściej wyświetlanych przez użytkowników (17–85%). Ponieważ do sumarycznej liczby pozycji udostępnianych i wyświetlanych dolicza się także usunięte publikacje, stanowiące kolejne wydania

tych samych dzieł czy wejścia do zasobów z komputerów pracowni digitalizacji, statystyki prezentowane na łamach poszczególnych bibliotek cyfrowych nie w pełni odzwierciedlają rzeczywiste wykorzystanie pojedynczych tytułów[25].

Zasoby czasopiśmiennicze w polskich bibliotekach cyfrowych są umieszczane w różnych działach/kolekcjach. Podczas gdy część bibliotek wprowadza dla nich bardzo ogólną kategorię „czasopisma” (np. Biblioteka Cyfrowa „Polona”, Biblioteka Cyfrowa Politechniki Łódzkiej „eBiPol”, Jeleniogórska Biblioteka Cyfrowa, Biblioteka Cyfrowa CODN) lub „czasopisma i gazety” (np. Bałtycka Biblioteka Cyfrowa), inne wydzielają dział „czasopisma”, dokonując jego podziału ze względu na zasięg terytorialny (np. Małopolska Biblioteka Cyfrowa, Wejherowska Biblioteka Cyfrowa) lub zakres tematyczny udostępnianych periodyków (np. Biblioteka Cyfrowa Politechniki Śląskiej, Akademicka Biblioteka Cyfrowa, Elbląska Biblioteka Cyfrowa, Biblioteka Cyfrowa Uniwersytetu Łódzkiego, Pedagogiczna Biblioteka Cyfrowa). Nieliczne biblioteki cyfrowe stosują jeszcze inne rozwiązanie, tj. dla najbardziej reprezentatywnych tytułów wydawnictw periodycznych tworzą odrębne zakładki, nadając im nazwę identyczną z tytułem czasopisma (np. „SilvaRerum” w Bibliologicznej Bibliotece Cyfrowej, „Annales Academiae Paedagogicae Cracoviensis” w Pedagogicznej Bibliotece Cyfrowej, „Czasopismo Techniczne” w Bibliotece Cyfrowej Politechniki Krakowskiej). W dużej grupie bibliotek w ogóle nie wyodrębnia się oddzielnego miejsca dla wydawnictw periodycznych, a umieszcza się je w innych kolekcjach, tj. w kolekcjach specjalnych (np. „Historia nauki i techniki” w Bibliotece Cyfrowej Politechniki Warszawskiej), materiałach dydaktycznych czy dziedzictwie kulturowym (np. biblioteki cyfrowe Uniwersytetów Zielonogórskiego, Warmińsko-Mazurskiego, Mikołaja Kopernika, Wrocławskiego), zasobach regionalnych (np. Sanocka Biblioteka Cyfrowa, Radomska Biblioteka Cyfrowa, Biblioteka Cyfrowa UMCS), wydawnictwach własnych uczelni (Biblioteka Cyfrowa Politechniki Łódzkiej, Biblioteka Cyfrowa UMW), publikacjach współczesnych (np. Biblioteka Cyfrowa Uniwersytetu Wrocławskiego, Świętokrzyska Biblioteka Cyfrowa). W większości przypadków wydawnictwa ciągłe przypisane są do kilku kolekcji jednocześnie (np. do materiałów dydaktycznych, regionalistów i dziedzictwa kulturowego, jak „Pielgrzym” w Kujawsko-Pomorskiej Bibliotece Cyfrowej). Zdarzają się jednak także takie sytuacje, że należą one wyłącznie do jednej z kategorii (np. do dziedzictwa kulturowego, jak „Adecadnik” w Wielkopolskiej Bibliotece Cyfrowej). Niestety, w procesie przyporządkowywania czasopism do poszczególnych kolekcji trudno dostrzec stałe i konsekwentnie stosowane zasady. Podczas gdy w jednych bibliotekach cyfrowych wydawnictwa seryjne uczelni są rejestrowane jako publikacje własne biblioteki (np. Biblioteka Cyfrowa Politechniki Łódzkiej), w innych są ujmowane jako typowe czasopisma naukowe (np. biblioteki cyfrowe Politechniki Śląskiej i Lubelskiej). Zdarza się także, że kategorią „czasopisma” obejmuje się biuletyny informacyjne uczelni, a jej zeszyty naukowe już nie (np. Biblioteka Cyfrowa Politechniki Łódzkiej). Choćby z tego powodu przytoczone powyżej statystyki odnoszące się do liczby zarejestrowanych w bibliotekach cyfrowych wydawnictw periodycznych należy traktować z pewną ostrożnością i dystansem.

Poza problemem lokalizacji wydawnictw periodycznych w strukturach bibliotek cyfrowych odrębnego rozpatrzenia wymaga także kwestia doboru ich tytułów. Ponieważ celem tworzenia bibliotek cyfrowych jest prezentacja w Internecie kulturowego dziedzictwa kraju, publikowanie naukowego dorobku regionów oraz wspieranie działalności dydaktycznej i naukowej, nie dziwi fakt, że do zasobów bibliotek włącza się także czasopisma i artykuły z czasopism. O ile digitalizacja cennych czasopism regionalnych czy też publikowanych przez instytucje będące uczestnikami projektów budowy bibliotek cyfrowych nie budzi większych zastrzeżeń, to pewnym niepokojem napawają zasady, jakimi kierują się twórcy bibliotek cyfrowych podczas wyboru i umieszczania w ich zasobach czasopism branżowych. Potencjalnemu użytkownikowi właściwie trudno zorientować się, które czasopisma dziedzinowe i dlaczego akurat te rejestruje się w poszczególnych bibliotekach cyfrowych. Z pomocą przychodzi oczywiście interfejs wyszukiwawczy FBC, pozwalający na jednoczesne przeszukiwanie wszystkich bibliotek

cyfrowych, ale i on nie rozwiązuje do końca problemu. Poszukując np. zdigitalizowanych czasopism bibliotekarskich, użytkownik może wpisać do wyszukiwarki słowo kluczowe „bibliotekarstwo” i wyświetlić sobie listę dostępnych tytułów wraz z informacją o ich lokalizacji w obrębie jednej z bibliotek cyfrowych. Niestety, lista wyników jest uporządkowana zgodnie z datą dodania lub modyfikacji danej publikacji, co sprawia, że poszczególne roczniki czy numery czasopism nie następują po sobie w układzie chronologicznym, ale są rozrzucone w obrębie całej listy wyników. O ile przy niewielkim zbiorze wyszukiwawczym lokalizacja interesującego użytkownika numeru czy rocznika czasopisma nie stanowi jeszcze problemu, to przy dużej liczbie rezultatów istnieje ryzyko niedostrzeżenia czy pominięcia danego tytułu. Inną wymagającą wyjaśnienia kwestią jest pytanie o zasady wyboru do digitalizacji poszczególnych tytułów i numerów czasopism. Zastanawiające jest, co sprawia, że część czasopism dziedzinowych rejestrowana jest w jednej, część w drugiej, a jeszcze inne w kolejnej bibliotece cyfrowej (np. „Bibliotekarz” w Kujawsko-Pomorskiej Bibliotece Cyfrowej, „Bibliotheca Nostra” w Śląskiej Bibliotece Cyfrowej, a „Notes Biblioteczny” w Małopolskiej Bibliotece Cyfrowej)? Dlaczego w niektórych bibliotekach cyfrowych umieszcza się jedynie wybrane roczniki bądź numery danego czasopisma (np. Wielkopolska Biblioteka Cyfrowa, czasopismo „Chemik” – całość rocznika 2000 oraz numery 1–9 z rocznika 2001; Podlaska Biblioteka Cyfrowa, czasopismo „Acta Mechanica et Automatica” – rocznik 2007, nr 1 i 2 oraz rocznik 2008, nr 1–4; Biblioteka Cyfrowa Politechniki Lubelskiej, czasopismo „Eksploracja i Niezawodność”, całość roczników 2001, 2004, 2006, 2007, 2008, po trzy numery z roczników 2002, 2003, 2005)? O ile odpowiedzią na pierwsze pytanie mogą być obowiązujące w naszym kraju regulacje prawne, wymagające zgody wydawcy na publikowanie materiałów w postaci cyfrowej, a co za tym idzie zmuszające digitalizujące biblioteki do intensyfikacji zabiegów na rzecz pozyskiwania tych praw, to w drugim przypadku racjonalnymi przesłankami do takiej wybiórczej digitalizacji mogą być: albo zły stan zachowania obiektów i konieczność ich natychmiastowej cyfryzacji, albo braki w księgozbiorach bibliotecznych, albo chęć udostępniania w sieci choćby fragmentów posiadanych kolekcji. Mało prawdopodobne wydaje się, by zainteresowanie czytelników ograniczało się do jednego wybranego numeru czasopisma i by biblioteki podejmowały kosztowny proces digitalizacji, biorąc pod uwagę tylko to kryterium. Niestety, pewne niekonsekwencje w działaniu twórców bibliotek cyfrowych każą domniemywać o zgubnej rywalizacji o pierwszeństwo pod względem liczby oferowanych publikacji.

Dodatkowych dowodów na potwierdzenie tej tezy dostarczają statystyki dotyczące artykułów z czasopism. Ich łączna liczba we wszystkich bibliotekach cyfrowych dLibry przekracza 3200 pozycji (stan na dzień 15 kwietnia 2009 r.). W tej grupie można odnotować zarówno pełne teksty artykułów z czasopism, notatki i ogłoszenia prasowe, listy gratulacyjne, jak i spisy treści poszczególnych numerów, roczników i tomów wydawnictw periodycznych, a co zaskakujące, także dokumenty życia społecznego (np. „Raport z Winobrania” – wycinek z czasopisma „Nadodrze” [19 X 1985, nr 20, s. 6–7] w Zielonogórskiej Bibliotece Cyfrowej), a niekiedy nawet całe numery wydawnictw ciągłych (np. „Wiadomości Polskie” nr 16 i 26 z 1915 r. w Śląskiej Bibliotece Cyfrowej). Jak dotąd, największą liczbę artykułów z czasopism w swoich kolekcjach posiada Wejherowska Biblioteka Cyfrowa (ponad 2600 pozycji), która digitalizacji poddaje zarówno artykuły pozbawione ochrony prawnoautorskiej, jak i te nadal nią objęte (np. artykuły z „Gazety Wyborczej”, „Dziennika. Polska, Europa, Świat”, „Expressu Gdyńskiego”, „Dziennika Bałtyckiego” itp.). W kontekście tego typu zabiegów po raz kolejny nasuwa się pytanie, dlaczego akurat te, a nie inne artykuły z czasopism znajdują się w zasobach bibliotek cyfrowych bądź dlaczego z danego numeru czasopisma są skanowane wyłącznie wybrane teksty (np. dwa artykuły z „Zagadnień Informacji Naukowej” R. 2001, nr 2 w Pedagogicznej Bibliotece Cyfrowej). Przeglądając poszczególne artykuły pod względem tematycznym, nie sposób nie zauważyć, że wiele z nich – choć nie wszystkie – włącza się do bibliotek cyfrowych w celu powiększenia ich zasobów. Nawet jeśli niektóre z artykułów rzeczywiście dokumentują pewne okolicznościowe wydarzenia, a poprzez to stanowią część dziedzictwa regionalnego, to istnieje tu też spora liczba takich, które nie mają wartości stricte naukowej czy historycznej, lecz popularyzatorsko-

sprawozdawczą (np. „Młoda, piękna i bogata” – wycinek prasowy dotyczący wizyty Grażyny Wolszczak w Wejherowskim Centrum Kultury opublikowany w Wejherowskiej Bibliotece Cyfrowej).

O ile w przypadku zdigitalizowanych całości czasopism można jeszcze mówić o konsekwencji w zakresie tworzenia powiązań między poszczególnymi ich numerami, tomami czy rocznikami, to w odniesieniu do artykułów z czasopism czasami trudno ustalić, na łamach jakich wydawnictw periodycznych zostały one opublikowane. Wprawdzie duża część bibliotek cyfrowych wprowadza wyraźne powiązania między artykułami i tytułami czasopism macierzystych, ale istnieją także takie, które informacje te całkowicie pomijają lub podają nieregularnie (np. Pedagogiczna Biblioteka Cyfrowa, Wejherowska Biblioteka Cyfrowa).

Analogiczny wniosek można sformułować także w odniesieniu do charakterystyk treściowych poszczególnych wydawnictw periodycznych oraz artykułów z nich pochodzących. W większości bibliotek cyfrowych są podawane podstawowe informacje dotyczące częstotliwości ukazywania się danego wydawnictwa, jego przeznaczenia, formatu i wydawcy. Nieco inaczej wygląda kwestia opisu zawartości pojedynczych artykułów z czasopism. Ich najbardziej wnikliwą charakterystykę w postaci rozbudowanego streszczenia zamieszczają: Małopolska Biblioteka Cyfrowa i Biblioteka Cyfrowa Politechniki Krakowskiej. Pozostałe biblioteki ograniczają się w zasadzie do wskazania tematu publikacji i identyfikujących ją słów kluczowych.

Żadna z bibliotek cyfrowych nie umieszcza w swoich zasobach odnośników do polskich baz danych, rejestrujących przecież niekiedy te same tytuły wydawnictw periodycznych i artykułów. Wszystkie z kolei informują o trybie dostępu do poszczególnych publikacji. Z uwagi na obowiązujące przepisy o prawie autorskim, część z nich jest powszechnie dostępna w sieci Internet, część jednak wyłącznie za pośrednictwem terminali znajdujących się na terenie danej placówki (biblioteki, uczelni, wydziału itp.).

Niestety, obecnie w procesie budowy zasobów cyfrowych nie udaje się uniknąć pewnych problemów. Poza wspomnianą już kwestią doboru tytułów do digitalizacji, przypisywaniem ich do odpowiednich kolekcji czy łamaniem prawa autorskiego, rozwiązaną wymaga także problem duplikowania wydawnictw zamieszczanych na łamach różnych bibliotek cyfrowych. Choć obecnie zjawisko to nie ma jeszcze charakteru epidemicznego (liczba potencjalnych duplikatów w bibliotekach dLibry na dzień 15 kwietnia 2009 r. wyniosła zaledwie 235 pozycji na ponad 211 tysięcy[26]), to w przyszłości może eskalować, zwłaszcza w przypadku czasopism. Zachodzi bowiem niebezpieczeństwo, że na skutek przyrostu zasobów cyfrowych i rywalizacji w zakresie liczby udostępnianych publikacji cyfrowych, biblioteki posiadające kompletne roczniki danego wydawnictwa periodycznego nie będą w stanie sprawdzać, czy inna biblioteka nie poddała już digitalizacji jego pojedynczych numerów. Z kolei odnalezienie w Internecie wybranych numerów posiadanego przez bibliotekę kompletu czasopisma może ją zniechęcać do „szatkowania” materiału i podejmowania kosztownych prac digitalizacyjnych wyłącznie dla potrzeb umieszczenia w sieci brakujących elementów kolekcji.

Wydaje się, że wypracowania nowych rozwiązań będzie wymagała także digitalizacja czasopism regionalnych. Coraz większa liczba bibliotek cyfrowych wymusi, prędzej czy później, na lokalnych bibliotekach cyfrowych podzielenie się zakresem kompetencji, tak by wszystkie nie digitalizowały tych samych tytułów czasopism lokalnych. W tym kontekście z należytą uwagą trzeba już dziś przyglądać się projektom uruchamiania kolejnych bibliotek cyfrowych, zwłaszcza tym, których twórcy w sposób nieuprawniony zawłaszczają dzieła cyfrowe z innych repozytoriów (jak np. Biblioteka Cyfrowa Archiwum Diecezjalnego w Tarnowie[27]) lub nagminnie łamią prawo z chęci powiększania swoich zasobów.

Najpoważniejszym zadaniem, z jakim będą musiały się zmierzyć polskie biblioteki cyfrowe, będzie jednak kwestia pełnotekstowego przeszukiwania dokumentów i ich indeksowania w wyszukiwarkach. Obecnie ustalenie chociaż procentowych szacunków dotyczących liczby zasobów cyfrowych przeszukiwanych pełnotekstowo jest niezwykle trudne i zależy od trzech czynników:

1. Formatu publikacji – czy wspiera przechowywanie informacji tekstowych (np. PDF czy DjVu vs. JPG);
2. Zawartości tekstowej publikacji – czy dokument zawiera jakąś informację tekstową (czy jest to dokument cyfrowy prymarny [*born-digital*] czy zeskanowany z zastosowaniem techniki OCR);
3. Mechanizmów indeksowania – czy serwis, z poziomu którego zasób jest przeszukiwany, indeksuje tekst z obiektów danego formatu.

Większość obiektów bibliotek cyfrowych dLibry jest zapisana w formacie, który wspiera przechowywanie informacji tekstowych, czyli DjVu, PDF, HTML. 8% publikacji posiada format HTML, co pozwala z dużym prawdopodobieństwem założyć, że posiadają one tekstową zawartość. Kolejne 8% to pliki PDF, w tym zarówno dokumenty cyfrowe prymarne, jak i zeskanowane. 82% to obiekty zapisane w formacie DjVu – w zdecydowanej większości skany. Chcąc odpowiedzieć na pytanie o procentowy udział zasobów przeszukiwalnych pełnotekstowo, konieczne byłoby więc przeprowadzenie pogłębionych badań dotyczących liczby skanów poddawanych procesowi OCR i dokładności tego procesu. Niestety, w Polsce dotychczas nie prowadzono tego typu badań, stąd trudności w precyzyjnej ocenie realizacji funkcji wyszukiwania pełnotekstowego w bibliotekach cyfrowych. Oczywiście, podejmując próby wyszukiwawcze z poziomu konkretnych bibliotek cyfrowych i mając do dyspozycji zamieszczoną w większości z nich opcję wyszukiwania w treści publikacji, potencjalny użytkownik może przypuszczać, że jest ona w nich realizowana. Niestety, praktyka pokazuje, że założenia te są błędne. Spora część bibliotek – mimo sygnalizowania usługi wyszukiwania pełnotekstowego – nie oferuje jej. Również serwis Federacji Bibliotek Cyfrowych nie stwarza takiej możliwości, gdyż indeksowane są w nim same opisy publikacji, nie zaś ich treść. Wreszcie pełnej możliwości przeszukiwania treści publikacji polskich bibliotek cyfrowych nie zapewnia także wyszukiwarka Google, o ile bowiem rejestruje ona jeszcze pliki HTML, PDF, doc czy ppt, to, niestety, nie indeksuje formatu DjVu, w którym zapisana jest większość obiektów cyfrowych dLibry[28].

Problemów z indeksowaniem publikacji poszczególnych bibliotek cyfrowych przez Google dowodzi prosty test. Wystarczy wpisać do pola wyszukiwawczego Google zapytanie dotyczące metadanych publikacji z dowolnej biblioteki cyfrowej (w postaci *site:adres biblioteki cyfrowej/docmetadata*), by przekonać się, że nie wszystkie dokumenty cyfrowe są przez wyszukiwarkę indeksowane (przykładowo dla Wielkopolskiej Biblioteki Cyfrowej, Google wyświetla około 61 900 wyników, podczas gdy biblioteka zawiera 80 tys. publikacji; dla Kujawsko-Pomorskiej Biblioteki Cyfrowej – 98 900 wyników???, gdy ta uwzględnia 27 tys. publikacji [na dzień 15 kwietnia 2009 r.]). Gdyby więc założyć, że jeden wynik Google to jedna publikacja, to można by przyjąć, że około 75% opisów Wielkopolskiej Biblioteki Cyfrowej i aż 366%???, Kujawsko-Pomorskiej Biblioteki Cyfrowej jest indeksowanych przez wyszukiwarkę. Rezultaty testu wyraźnie wskazują, że indeksacja Google jest daleka od precyzji, a wyświetlane wyniki to niejednokrotnie opisy tych samych publikacji. Podobnie wypada test odnoszący się do elementów treściowych (*site:adres biblioteki cyfrowej/Content*). W odpowiedzi na pytanie o zawartość treściową publikacji z wybranych bibliotek cyfrowych, Google wyświetla 2250 wyników pochodzących z Wielkopolskiej Biblioteki Cyfrowej oraz 432 wyniki – z Kujawsko-Pomorskiej Biblioteki Cyfrowej (na dzień 15 kwietnia 2009 r.). Rezultat taki każe sądzić, że w Google indeksowana jest treść jedynie 2,5% publikacji z pierwszej i 1,4% publikacji z drugiej biblioteki. Po raz kolejny dane te okazują się jednak mało precyzyjne, gdyż w wynikach są wyłącznie uwzględniane publikacje zapisane w formatach przyjaznych dla wyszukiwarki, a więc PDF i

text/html, a całkowicie pomijane pliki DjVu. W kontekście rezultatów tych prostych testów, zdania o możliwościach pełnotekstowego przeszukiwania w polskich bibliotekach cyfrowych oraz o pełnej indeksacji zasobów polskich bibliotek cyfrowych w Internecie wydają się wielce przesadzone. By móc wiarygodnie określić zarówno procentowy udział zasobów bibliotek cyfrowych przeszukiwanych pełnotekstowo, jak i ich widoczność w sieci, konieczne byłoby więc przygotowanie specjalnych narzędzi pomiarowych oraz opracowanie jednoznacznej metodologii badań zasobów cyfrowych (co liczyć: strony, utwory, pozycje, znaki, skany, tytuły?).

Zawartość polskich baz danych

Po dokonaniu analizy zawartości polskich bibliotek cyfrowych nasuwa się wniosek, że wszystko to, co obecnie wchodzi w ich zakres, było – w mniejszym lub większym stopniu – rejestrowane od lat w bazach danych. Wiele z nich już od początku swego istnienia oferowało opisy bibliograficzne różnego typu wydawnictw: książek, czasopism, artykułów z czasopism, poloników, utworów dźwiękowych i kartograficznych. Z czasem zaczęto rejestrować także dokumenty elektroniczne i specjalne materiały czytelnicze, a obok baz stricte bibliograficznych pojawiły się bazy bibliograficzno-abstraktowe i pełnotekstowe. Od momentu przygotowania pierwszych polskich baz danych do dziś zawartość wielu z nich właściwie nie uległa zmianie, nadal są bowiem tworzone bazy składające się z „suchych” opisów bibliograficznych wybranych form wydawniczych czy piśmienniczych (np. bazy bibliograficzne Biblioteki Narodowej), które podając podstawowe informacje identyfikujące dany tytuł, pełnią przede wszystkim rolę wykazów czy rejestratorów produkcji wydawniczej danego kraju. Obok nich coraz liczniej powstają jednak także bazy bibliograficzno-abstraktowe, prezentujące w sposób ciągły zawartość czasopism i innych wydawnictw dotyczących danej dyscypliny lub obszaru geograficznego czy dorobek określonych środowisk naukowych.

Ponieważ charakterystyka zawartości wszystkich polskich bibliograficznych i bibliograficzno-abstraktowych baz danych przerosłaby ramy niniejszej publikacji, ograniczono się w niej wyłącznie do omówienia tendencji charakterystycznych dla baz reprezentowanych na konferencji „Bibliograficzne bazy danych: kierunki rozwoju i możliwości współpracy”, dla potrzeb której niniejszy tekst został przygotowany.

Analizując zawartość polskich baz danych, można stwierdzić, że pod względem formalnym w ich zbiorach dominują opisy bibliograficzne artykułów z czasopism i fragmentów piśmienniczych z wydawnictw zbiorowych (90%). Choć trudno dokonać precyzyjnych szacunków co do sumarycznej liczby wszystkich rekordów, można pokusić się o stwierdzenie, że kilkakrotnie przekracza ona łączną liczbę wszystkich pozycji udostępnianych w polskich bibliotekach cyfrowych (dla przykładu sam BazTech rejestruje ponad 150 tys., a baza AGRO ponad 300 tys. rekordów [stan na kwiecień 2009 r.]). Dokonując tego typu porównań, należy jednak pamiętać, że podobnie jak w przypadku bibliotek cyfrowych, również w bazach danych można zaobserwować zjawisko dublowania się części opisów bibliograficznych. Jest to znamienne zwłaszcza dla baz tworzonych w pokrewnych dyscyplinach naukowych (np. PEDAGOG i EDUKACJA, BazTech i BIMET, SIGŻ i AGRO, „Polska Bibliografia Wojskowa” i BazHum[29]), ale zdarza się także w przypadku baz własnych bibliotek i bibliografii narodowych, co jest bezpośrednim wynikiem przejmowania opisów bibliograficznych z jednych do drugich (np. „BIBLIOgrafia Prac Pracowników Politechniki Warszawskiej” i BZCz). Poza opisami artykułów z czasopism, w rekordach wielu polskich baz danych są także gromadzone informacje dotyczące innego typu wydawnictw, tj. książek, wydawnictw seryjnych, aktów prawnych i różnych form recepcji literatury naukowej (np. audycji telewizyjnych i radiowych, niepublikowanych prac doktorskich). Podczas ewidencji materiału bibliograficznego uwzględnia się publikacje wydawane w kraju w języku polskim i językach obcych oraz wydawane poza jego granicami w języku polskim, a także obcojęzyczne, jeśli dotyczą pisarzy polskich lub gdy ich autorem jest Polak (np. „Polska Bibliografia Literacka”).

Prezentowane w zbiorach bazodanowych materiały czasopiśmiennicze, podobnie jak w bibliotekach cyfrowych, są wynikiem opracowywania różnego typu wydawnictw ciągłych: periodyków ogólnopolskich, czasopism branżowych i dziedzinowych, wydawnictw seryjnych i zeszytów naukowych publikowanych przez placówki naukowe, wybranych tytułów prasy emigracyjnej oraz zagranicznych czasopism naukowych. Liczba tytułów wydawnictw periodycznych rejestrowanych w poszczególnych bazach danych waha się od kilkunastu (jak np. w bazie Arton – 26) do kilkuset (jak np. w bazie AGRO – 845), choć oczywiście w przypadku bibliografii narodowych osiąga liczbę kilku lub kilkunastu tysięcy (np. „Czasopisma Polskie 1985–2000” – 10 246 dokumentów).

Ponieważ popularność dziedzinowych baz danych w środowisku akademickim sukcesywnie wzrasta, a ich twórcy dążą do tego, by stawały się one specjalistycznymi informatorami i przewodnikami po literaturze fachowej, materiały w nich rejestrowane nie pochodzą ze wszystkich ukazujących się w kraju branżowych wydawnictw periodycznych, ale z tych najważniejszych i uznanych za najbardziej reprezentatywne w danej dyscyplinie czy obszarze badawczym. Inaczej niż w bibliotekach cyfrowych, w bazach danych na decyzję o rejestrowaniu materiałów z selektywnych periodyków bezpośrednio wpływają m.in. popularność danego tytułu w środowisku naukowym, jego poziom merytoryczny, obecność na liście czasopism punktowanych czy wreszcie efektywność współpracy z wydawcami czasopism. Choć w przypadku baz danych, podobnie jak w odniesieniu do bibliotek cyfrowych, można mówić o pewnej selekcji tytułów czasopism, a także – analogicznie jak w bibliotekach cyfrowych – o różnym stopniu kompletności danych, to wydaje się, że udostępnienie użytkownikom zhierarchizowanych list zarejestrowanych w bazach roczników czasopism, a w ich obrębie numerów i artykułów z nich pochodzących (np. BazTech, BazHum, AGRO, „Polska Bibliografia Literacka”), jest rozwiązaniem o wiele bardziej przydatnym niż wyświetlanie zasobu według dat dodania/modyfikacji publikacji, jak ma to miejsce w niektórych bibliotekach cyfrowych. Innym, właściwym z punktu widzenia potrzeb informacyjnych użytkownika, udogodnieniem jest umieszczanie w bazach danych podstawowych informacji dotyczących poszczególnych wydawnictw periodycznych (częstotliwość, wydawca, zmiany tytułu, impact factor, punktacja MNiSW) oraz dołączanie do artykułów ich streszczeń (lub adnotacji o zawartości) w języku polskim lub/i angielskim. Niestety, tej ostatniej zasady – podobnie jak w bibliotekach cyfrowych – przestrzega się z różną konsekwencją (np. część artykułów w BazTechu i bazie AGRO posiada abstrakty, część nie). Twórcy niektórych baz danych (np. BazTech) zastrzegają jednak, że sytuacja taka może mieć miejsce i wynika bezpośrednio ze sposobu opracowywania zawartości rekordów. Ponieważ większość z nich pochodzi z oryginalnych tekstów publikacji, stąd jeśli w opracowywanej pozycji pewne dane nie występują (np. słowa kluczowe, streszczenie, informacje o autorze), to nie pojawiają się one także w rekordach bazy danych. Biblioteki cyfrowe nie informują swoich użytkowników o tego typu anomaliach.

Przyglądając się zawartości rekordów wielu zagranicznych baz danych (np. EBSCO), uwagę zwracają dwa istotne elementy, tj. obecność bibliografii załącznikowej przy artykułach z czasopism i istnienie odnośników do pełnych tekstów wydawnictw ciągłych. To swoiste skupienie kilku źródeł w jednym miejscu czyni z baz danych kompleksowe narzędzia informacji, a użytkownikom pozwala na rozszerzenie obszaru poszukiwań literatury przedmiotu oraz doskonalenie umiejętności poruszania się po różnych źródłach informacji. Niestety, niewiele polskich baz danych stwarza taką możliwość: bibliografie załącznikowe w rekordach artykułów z czasopism zamieszcza jedynie BazTech^[30] (przy tym także nie dla wszystkich, lecz włączanych do bazy od 2006 r.), zaś hiperłącza do pełnych tekstów publikacji – ten sam BazTech oraz dwie bibliografie: „Bibliografia Publikacji Pracowników Politechniki Krakowskiej” i „Bibliografia Publikacji Pracowników Uniwersytetu Mikołaja Kopernika w Toruniu”. O ile „Bibliografia Publikacji Pracowników Politechniki Krakowskiej” kieruje użytkownika wyłącznie do zasobu biblioteki cyfrowej uczelni, to BazTech odsyła zarówno do stron internetowych wydawców czasopism, na których udostępnia się pełne teksty artykułów z nich pochodzących, jak i do polskich bibliotek cyfrowych. W największą liczbę powiązań wyposażona jest „Bibliografia Publikacji Pracowników Uniwersytetu Mikołaja

Kopernika w Toruniu”, która stosuje połączenia hipertekstowe do Kujawsko-Pomorskiej Biblioteki Cyfrowej, własnych wydawnictw uczelnianych, „Biuletynu EBIB”, stron WWW wydawców czasopism oraz kilku innych zasobów zewnętrznych (np. portali bibliotek akademickich, zagranicznych baz danych, Biblioteki Wirtualnej Nauki). Choć w dobie rozwoju społeczeństwa informacyjnego mogłoby się wydawać, że jedną z naturalnych konsekwencji tworzenia instytucjonalnych bibliotek cyfrowych będzie umieszczanie w katalogach i bazach własnych bibliotek specjalnych odnośników do tworzonych przez siebie zasobów cyfrowych, polska praktyka pokazuje, że jest to błędne przekonanie. Do rekordów większości elektronicznych bibliografii, rejestrujących dorobek naukowy pracowników poszczególnych uczelni, nie wprowadza się odnośników do zdigitalizowanych kopii dokumentów, znajdujących się w tworzonych przez te uczelnie bibliotekach cyfrowych (np. bibliografie pracowników AGH, Politechniki Lubelskiej, Łódzkiej, Warszawskiej, Wrocławskiej, Śląskiej, uniwersytetów Marii Curie-Skłodowskiej, Wrocławskiego, Białostockiego, Rzeszowskiego, Zielonogórskiego). Również w wielu katalogach OPAC nie umieszcza się stosownych hiperłączy, w tych zaś, w których one występują (np. katalog OPAC Akademii Górniczo-Hutniczej, Uniwersytetu Ekonomicznego w Krakowie, Politechnik: Warszawskiej, Lubelskiej, Łódzkiej, UMCS, UMW, UMK), rozwiązanie to stosowane jest niekonsekwentnie (często wprowadza się połączenia dla wydawnictw zwartych, a pomija dla wydawnictw periodycznych [np. Katalog OPAC Biblioteki Uniwersyteckiej w Toruniu nie odsyła użytkownika do czasopism zdigitalizowanych w ramach Kujawsko-Pomorskiej Biblioteki Cyfrowej]).

Z całą pewnością wielość, a niejednokrotnie brak odpowiednich rozwiązań nie ułatwia użytkownikowi pracy z zasobami informacji naukowej. W tej sytuacji jak najbardziej uzasadnione wydają się postulaty nawiązania ścisłej kooperacji pomiędzy twórcami różnorodnych narzędzi informacyjnych oraz wypracowania pewnych mechanizmów koordynacji działań. Może warto przy tym wziąć pod uwagę istniejące już modele postępowania, przyjęte np. w Bibliotece Politechniki Krakowskiej, która zarówno w swoim katalogu głównym, jak i tworzonej przez siebie bazie danych oraz bibliotece cyfrowej wprowadza wzajemnie krzyżujące się odsyłacze do zasobów. Przykłady ich umieszczenia w elementach opisu bibliograficznego obrazują rys. 2a, b, c.

Książka Egzemplarze Zamówienie

Tytuł: Hydraulika : zestawienie pojęć i wzorów stosowanych w budownictwie : podręcznik dla studentów wyższych szkół technicznych / Krzysztof Wojciech Książczyński # [wyd.2]

Egzemplarze:
Do wypożyczenia: 9 (Filia DS-4)
Wypożyczone: 33
Dostępne w Czytelnii Głównej: 1
Pełny tekst on-line: dostępny w zasobach **Biblioteki Cyfrowej PK**

Autor: Książczyński, Krzysztof Wojciech (1951-)
Wydawca: Wydaw. Politech. Krakowskiej
Dane wydawnicze: Wyd. 2 popr. i uzup.; Kraków; 2002
ISBN: 83-7242-230-3
Opis fizyczny: 122 s. : rys., wykr. ; 24 cm
Nośnik informacji: Dokumenty drukowane
Typ publikacji: Podręczniki
Język dokumentu: pol
Symbol klasyfikacji: 532(075.8)
Uwagi: Bibliogr. s. 121
Hasło tezaursusa: Mechanika cieczy. Hydromechanika. Hydraulika.

Rys. 2a. Rekord z katalogu on-line Biblioteki Politechniki Krakowskiej

Źródło: *Katalog Biblioteki Politechniki Krakowskiej: Książki* [on-line]. [Dostęp 15 kwietnia 2009].
Dostępny w World Wide Web: <http://twonk.biblos.pk.edu.pl/BC/>.

Książka	Egzemplarze	Zamówienie
----------------	--------------------	-------------------

Tytuł: Hydraulika : zestawienie pojęć i wzorów stosowanych w budownictwie : podręcznik dla studentów wyższych szkół technicznych / Krzysztof Wojciech Książczyński #[wyd.2]

Egzemplarze:
 Do wypożyczenia: 9 (*Filia DS-4*)
 Wypożyczone: 33
 Dostępne w Czytelnicy Głównej: 1
 Pełny tekst on-line: dostępny w zasobach **Biblioteki Cyfrowej PK**

Autor: Książczyński, Krzysztof Wojciech (1951-)

Wydawca: Wydaw. Politech. Krakowskiej

Dane wydawnicze: Wyd. 2 popr. i uzup.; Kraków; 2002

ISBN: 83-7242-230-3

Opis fizyczny: 122 s. : rys., wykr. ; 24 cm

Nośnik informacji: Dokumenty drukowane

Typ publikacji: Podręczniki

Język dokumentu: pol

Symbol klasyfikacji: 532(075.8)

Uwagi: Bibliogr. s. 121

Hasło tezaursusa: Mechanika cieczy. Hydromechanika. Hydraulika.

Rys. 2b. Rekord z Bibliografii Publikacji Pracowników Politechniki Krakowskiej
 Źródło: *Bibliografia Publikacji Pracowników Politechniki Krakowskiej* [on-line]. [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://twonk.biblos.pk.edu.pl/BC/uep.shtml>.

Książka	Egzemplarze	Zamówienie
----------------	--------------------	-------------------

Tytuł: Hydraulika : zestawienie pojęć i wzorów stosowanych w budownictwie : podręcznik dla studentów wyższych szkół technicznych / Krzysztof Wojciech Książczyński #[wyd.2]

Egzemplarze:
 Do wypożyczenia: 9 (*Filia DS-4*)
 Wypożyczone: 33
 Dostępne w Czytelnicy Głównej: 1
 Pełny tekst on-line: dostępny w zasobach **Biblioteki Cyfrowej PK**

Autor: Książczyński, Krzysztof Wojciech (1951-)

Wydawca: Wydaw. Politech. Krakowskiej

Dane wydawnicze: Wyd. 2 popr. i uzup.; Kraków; 2002

ISBN: 83-7242-230-3

Opis fizyczny: 122 s. : rys., wykr. ; 24 cm

Nośnik informacji: Dokumenty drukowane

Typ publikacji: Podręczniki

Język dokumentu: pol

Symbol klasyfikacji: 532(075.8)

Uwagi: Bibliogr. s. 121

Hasło tezaursusa: Mechanika cieczy. Hydromechanika. Hydraulika.

Rys. 2c. Rekord z Biblioteki Cyfrowej Politechniki Krakowskiej
 Źródło: *Biblioteka Cyfrowa Politechniki Krakowskiej* [on-line]. [Dostęp 15 kwietnia 2009].
 Dostępny w World Wide Web: http://www.biblos.pk.edu.pl/bc_zasoby&operation=details&id=645.

Wydaje się, że rozwiązanie przyjęte w zasobach Biblioteki Cyfrowej Politechniki Krakowskiej nie tylko sprawdziłoby się w innych repozytoriach i archiwach danych, ale znacznie ułatwiłoby docieranie do literatury naukowej. Ponieważ poszukując piśmiennictwa na określony temat, pracownik akademicki rozpoczyna najczęściej od najbardziej kompletnej w danej dziedzinie bazy bibliograficznej, a następnie jest zainteresowany szybkim i prostym dostępem do tekstu źródłowego, baza taka powinna być wyposażona w odpowiednie opcje umożliwiające swobodne przemieszczanie się pomiędzy jej rekordami a zewnętrznymi zasobami. Osobliwa wydaje się sytuacja, gdy duża część czasopism fachowych ukazujących się w danej dziedzinie poddana została digitalizacji, a podstawowa baza bibliograficzna rejestrująca ich tytuły nie informuje o tym

użytkownika. W następstwie tego zamiast zdalnie korzystać z istniejących kopii cyfrowych dokumentów, udaje się on do czytelnicy, by kopie te wykonać. Częściowo problem rozwiązałoby więc stosowanie odpowiednich odsyłaczy (bazy danych EDUKACJA i PEDAGOG mogłyby dla przykładu kierować użytkowników do Pedagogicznej Biblioteki Cyfrowej lub Biblioteki Cyfrowej CODN, a Polska Bibliografia Bibliologiczna np. do Kujawsko-Pomorskiej Biblioteki Cyfrowej, gdzie pełnotekstowo udostępniono pewną liczbę czasopism z zakresu szeroko pojętej nauki o książce). Wprowadzenie hiperłączy do bibliotek cyfrowych wydaje się szczególnie sensowne zwłaszcza w przypadku baz pełniących rolę centralnych rejestratorów (np. bibliografie BN, SYMPOnet), które przechowują opisy bibliograficzne wielu zdigitalizowanych już dokumentów. Tę potrzebę doskonale zrozumieli twórcy wymienionych już wcześniej baz danych: BazTech, „Bibliografii Publikacji Pracowników Politechniki Krakowskiej” (choć tu można byłoby jeszcze umieścić odpowiedni odnośnik do BazTech-u) i „Bibliografii Publikacji Pracowników Uniwersytetu Mikołaja Kopernika w Toruniu”, stwarzając użytkownikowi możliwość dostępu do publikacji pełnotekstowych. By pomysł ten mógł być realizowany na szerszą skalę, konieczne byłoby nawiązanie ścisłej współpracy twórców baz danych i bibliotek cyfrowych, tak by ci pierwsi nie byli zmuszeni do penetrowania istniejących zasobów cyfrowych w poszukiwaniu dokumentów, których opisy zostały zarejestrowane w stworzonych przez nich bazach.

Pobieżne przeglądy statystyk publikowanych przez Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego czy twórców BazTech-u wskazują, że liczba sesji czy pobrań dokumentów z poszczególnych baz danych jest proporcjonalna do wielkości baz, liczby jej użytkowników czy wreszcie liczby członków poszczególnych instytucji uczestniczących. Podobnie jak w przypadku bibliotek cyfrowych ocena wykorzystania poszczególnych baz danych jest niezwykle trudna, gdyż w większości przypadków ich twórcy nie udostępniają publicznie statystyk ich użytkowania. Trudności sprawia także udzielenie odpowiedzi na pytanie o obecność obiektów poszczególnych baz danych w Internecie. Przeprowadzenie analogicznego testu, jak w przypadku bibliotek cyfrowych, dostarcza równie nieprecyzyjnych wyników. Wprowadzenie do wyszukiwarki Google zapytań w postaci *artykuł site:adres bazy danych* lub *czasopismo site:adres bazy danych* prowadzi do wniosku, że zawartość większości polskich baz danych nie jest indeksowana przez tę wyszukiwarkę, choć istnieją w sieci. Wyjątek na tym tle stanowią BazTech, AGRO i Polska Bibliografia Literacka, których opisy są rejestrowane przez Google. I w tym przypadku – podobnie jak w teście przeprowadzonym dla bibliotek cyfrowych – rezultaty wyszukiwania dowodzą nieprecyzyjnych algorytmów indeksowania dokumentów: w odpowiedzi na zapytanie dotyczące liczby artykułów pochodzących z BazTechu, Google wyświetla bowiem 131 000 wyników (na ogólną liczbę 152 770 rekordów w bazie), w odpowiedzi zaś na zapytanie o liczbę czasopism – 115 000 wyników, podczas gdy baza rejestruje 497 tytułów. Analogicznie wypadają testy dla pozostałych dwóch baz. Niestety, rekordy z pozostałych polskich baz danych nie są widoczne dla użytkowników standardowej sieci. Dzieje się tak dlatego, że o ile robot Google jest w stanie odnaleźć stronę główną konkretnej bazy danych, to nie potrafi już zainicjować jej działania czy też samodzielnie wygenerować jej rekordów jako typowych dla siebie stron HTML, wymaga to bowiem zadania bazie określonego zapytania. Akcję taką może uruchomić użytkownik, który stosując własne kryteria czy preferencje, generuje stronę o odpowiedniej zawartości. Ponieważ taka strona posiada zazwyczaj bardzo długi adres, który w swej strukturze może zawierać różne znaki specjalne (np. &, %, ?,=), nie każdy robot potrafi ją odnaleźć i zaindeksować. Wiele zależy zatem od samej struktury bazy danych i technologii, w jakiej jest tworzona. Bez ich znajomości żadna wyszukiwarka nie będzie w stanie ani prawidłowo czytać danych, ani powiązać ich ze sobą. By zasoby polskich baz danych stały się widoczne dla standardowych wyszukiwarek, konieczna jest ścisła współpraca twórców obu narzędzi i przynajmniej na razie konwersja tych pierwszych do regularnych stron HTML[31].

Bazy danych w bibliotekach cyfrowych czy biblioteki cyfrowe w bazach danych?

Istnienie wielu różnorodnych narzędzi i zasobów informacji naukowej sprawia, że współczesny użytkownik, by dotrzeć do poszukiwanej literatury, musi poruszać się w obrębie wielu interfejsów i technik nawigacji. Korzystając przy tym z systemów zakwalifikowanych jako pełnotekstowe, wielokrotnie dowiadyuje się, że droga do treści publikacji jest realizowana w różny sposób – niekiedy jest to tylko dostęp do strony internetowej czasopisma lub samej informacji o zasadach udostępniania wersji pełnotekstowej, innym razem jest to konieczność przejścia przez wiele ekranów i odsyłaczy. Zdarza się i tak, że po pomyślnym zakończeniu całej procedury dostępu pełny tekst wcale się nie otwiera, ponieważ użytkownik nie ma do niego uprawnień, albo dlatego, że serwis nie do końca jest pełnotekstowy[32].

Z uwagi na to, że rozpoznanie struktur poszczególnych platform i faktycznej zawartości wielu serwisów oraz nawiązanie dialogu z różnymi systemami wymaga dużych sprawności, obecnie obserwuje się tendencję do tworzenia systemów, które byłyby w stanie wchłonąć wszelką zawartość i zastąpić wszystkie inne. Próbuje się więc budować „nadsystemy”, które w sposób niewidoczny dla użytkownika i w jego interesie scalałyby istniejące źródła[33]. Do takiej roli po części pretendują biblioteki i repozytoria cyfrowe. Obecność w wielu z nich zasobów czasopiśmienniczych, rejestrowanych także w bibliograficznych bazach danych, sprawia, że pojawiają się postulaty scalenia zasobów. Jednym z argumentów podawanych przez zwolenników takiego rozwiązania jest możliwość zapewniania przez biblioteki cyfrowe dostępu do pełnych tekstów publikacji oraz ich przeszukiwania na poziomie treści. Z całą pewnością zaletą bibliotek cyfrowych jest operowanie różnego typu danymi, niejednokrotnie zapisanymi w odmiennych formatach oraz udostępnianie pełnych tekstów publikacji. Jak dowiedziono w niniejszym opracowaniu, nie wszystkie biblioteki cyfrowe oferują jednak przeszukiwanie swoich zasobów poprzez treść publikacji i nie wszystkie zapewniają zdalny dostęp do całego zasobu zdigitalizowanych pozycji. Pewnych wątpliwości dostarcza także proces doboru materiałów do bibliotek cyfrowych, który w wielu przypadkach jest mało konsekwentnie realizowany. Dołączenie do tych zasobów w miarę kompletnych, lecz posiadających mimo wszystko inną strukturę baz danych, wydaje się więc mało uzasadnione.

Rozważając propozycję połączenia obu źródeł, warto wziąć pod uwagę także pewne specyficzne zadania baz danych – odmienne od tych realizowanych przez biblioteki cyfrowe. Choć większość polskich bibliograficznych baz danych nie oferuje dostępu do pełnego tekstu publikacji, są one wykorzystywane przez wiele bibliotek przy układaniu bibliografii poszczególnych autorów, wykonywaniu indeksów do zawartości czasopism na przestrzeni wielu roczników czy ustalaniu listy czasopism rankingowych. Zasobów bibliotek cyfrowych nie można wykorzystać w ten sam sposób, gdyż rejestrują one tylko pewien wybór publikacji i nie dążą do kompletności. Dzięki rejestracji czasopism recenzowanych oraz obecności licznych odsyłaczy, kierujących użytkowników od jednej publikacji do innych, bazy danych stanowią bardzo cenne źródło informacji, dając przegląd merytorycznie wartościowego piśmiennictwa naukowego. Ich stała aktualizacja dodatkowo podnosi ich wartość w tym zakresie. Bez wątplenia korzystanie z takiego źródła pozwala na ścisłe ukierunkowanie i efektywne wyszukiwanie literatury przedmiotu. Ze względu na ograniczenia prawa autorskiego, w przeważającej większości bibliotek cyfrowych udostępnia się stare publikacje. Choćby z tego powodu zasoby te nie mogą stanowić dobrego źródła informacji dokumentującego rozwój współczesnych tendencji naukowych czy ich przepływ w obrębie poszczególnych dyscyplin.

Oczywiście również w przypadku tworzenia baz danych nie udaje się uniknąć pewnych niekonsekwencji. Część baz podaje informacje o wartości indeksowanych czasopism (impact factor lub punktacja MNIŚW), część dane te pomija. Te same elementy opisu bibliograficznego przybierają niekiedy różną formę i są zamieszczane w różnych miejscach (np. informacje o

bibliografii załącznikowej bywają zapisywane w jednej bazie raz w polu „cytata”, innym razem w polu „uwagi”; część baz danych dla wskazania rodzaju indeksowanej publikacji stosuje pole „typ dokumentu”, inne „wartość zawartości”). Dodatkowo w rekordach większości baz danych nie podaje się ani szczegółowych informacji na temat istniejącej bibliografii załącznikowej, ani informacji o istniejących w innych zasobach kopiach cyfrowych dokumentów. Z całą pewnością stosowanie różnych systemów baz danych nie ułatwia pracy potencjalnemu użytkownikowi. Być może uzasadnione byłoby więc stworzenie zintegrowanej platformy bazodanowej (np. na wzór serwisu EBSCO) i udostępnianie z jej poziomu przynajmniej zawartości wszystkich dziedzinowych baz danych. By urzeczywistnić taką ideę, konieczne byłoby jednak ujednoczenie elementów rekordów bibliograficznych i stworzenie odpowiedniego zaplecza technicznego (być może jego podstawy już istnieją w postaci systemu YADDA i Domeny Internetowych Repozytoriów Wiedzy?).

Dokonując oceny kondycji polskich baz danych, warto zauważyć, że w chwili obecnej biblioteki cyfrowe nie stanowią dla nich konkurencji, ponieważ mimo deklarowanej możliwości przeszukiwania pełnotekstowego większość z nich nie oferuje tej funkcji na poziomie artykułów. Zbyt wysokie koszty korekty i niedoskonałości techniki OCR sprawiają, że wielu polskich bibliotek cyfrowych nie stać jeszcze na digitalizację z wykorzystaniem tej metody. Jednak w przyszłości, gdy programy optycznego rozpoznawania pisma zostaną udoskonalone, sytuacja ulegnie zmianie. By móc konkurować z bibliotekami cyfrowymi, a jednocześnie stanowić kompleksowe narzędzie informacyjne, bibliograficzne bazy danych już dziś powinny rozwijać się w kierunku baz pełnotekstowych. Wydaje się, że większość z nich jest do tego technicznie przygotowana. Stosownych regulacji będą wymagać oczywiście kwestie zdobywania pełnych tekstów od wydawców czasopism i pozyskiwania praw autorskich oraz rywalizacji o nie z twórcami bibliotek cyfrowych. By nie konkurować w tym względzie, bazy danych powinny skupić się na udostępnianiu artykułów z czasopism dziedzinowych, a tym samym pełnieniu roli repozytoriów tekstów specjalistycznych, biblioteki cyfrowe zaś na digitalizacji wydawnictw zwartych i dokumentów regionalnych, w tym także czasopism lokalnych.

Pomysł prezentacji w bibliotekach cyfrowych dziedzictwa kulturowego regionu wydaje się uzasadniony z kilku powodów. Po pierwsze, już dziś zgodny jest z misją powołania wielu bibliotek cyfrowych. Po drugie, bibliotekom będzie znacznie łatwiej pozyskiwać fundusze na digitalizację, gdy będzie ona miała wymiar regionalny i związana będzie z promocją środowiska lokalnego. Po trzecie, potrzeby informacyjne użytkowników zasobów cyfrowych będą stale rosły, a nie wszystkie polskie biblioteki cyfrowe będą w stanie je w pełni zaspokoić. Już obecni użytkownicy bibliotek cyfrowych odczuwają pewien niedosyt, gdyż w ich zasobach przeważają publikacje pozbawione praw autorskich, a więc choć wartościowe, to jednak stare. Ponieważ w przyszłości nadal trudno będzie konkurować z gigantami światowymi (jak Google Books czy Library of Congress), konieczne stanie się znalezienie pewnej niszy dla polskich zasobów cyfrowych. Może ją stanowić właśnie tematyka regionalna, z jednej strony słabo reprezentowana w wielkich bibliotekach świata, a z drugiej – systematycznie gromadzona, opracowana i często unikalna w zbiorach polskich bibliotek. By móc wypełnić nową rolę, twórcy bibliotek cyfrowych będą musieli jednak zmienić swoje nastawienie i digitalizować materiały nie tylko z myślą o środowiskach naukowych, ale także dla potrzeb przeżywających swój renesans społeczności lokalnych[34].

Dziś, kiedy perspektywiczne rozwiązania są jeszcze w fazie przemyśleń lub planowania, warto przynajmniej powiązać ze sobą już istniejące źródła i zasoby. Umieszczenie wzajemnych odsyłaczy (np. choćby w formie minimalnych odnośników: *Czasopismo indeksowane w „NAZWA BAZY DANYCH”* [w bibliotekach cyfrowych] czy *Pełny tekst publikacji dostępny pod adresem: ADRES BIBLIOTEKI CYFROWEJ* [w bazach danych]) przyniesie użytkownikom wiele korzyści. Z jednej strony pozwoli bowiem na lokalizację i zapoznanie się z zawartością kopii cyfrowych publikacji zarejestrowanych w bazach danych (procedura wyszukiwania: baza danych -> biblioteka cyfrowa),

z drugiej natomiast umożliwi gromadzenie merytorycznie wartościowej literatury przedmiotu czy weryfikację danych wymaganych do oceny parametrycznej poszczególnych badaczy i jednostek naukowych (procedura odwrotna: biblioteka cyfrowa -> baza danych). Zastanawiając się nad sensem tego typu rozwiązania, warto uświadomić sobie fakt, że o ile dla pracownika akademickiego punktem wyjścia w poszukiwaniach piśmiennictwa będą przede wszystkim bibliograficzne bazy danych, to potencjalny przedstawiciel „pokolenia Google” rozpocznie swoje zmagania z informacją od skorzystania z wyszukiwarki. Brak odpowiednich powiązań między zasobami skieruje pierwszego z użytkowników do tradycyjnych zasobów bibliotek, drugiego zaś do wybiórczego zestawu opisów bibliograficznych i pełnych tekstów publikacji cyfrowych, których wartości ani przydatności nie będzie umiał krytycznie ocenić. W obu przypadkach potwierdzi się więc teza, że sukces w świecie wyszukiwania informacji determinują jej odpowiedniość, kompletność i widoczność.

Przypisy

[1] NAISBITT, J. *Megatrendy. Dziesięć nowych kierunków zmieniających nasze życie*. Poznań: Zysk i S-ka, 1997, s. 250. ISBN 978-83-715-007-18.

[2] Terminy *źródło, zasób, zbiór informacji i narzędzie informacyjne* będą w niniejszym artykule traktowane jako synonimiczne i używane zamiennie.

[3] PŁOSKI, Z. *Słownik Encyklopedyczny „Informatyka”*. Warszawa: Europa, 1999, s. 46–47. ISBN 83-87977-16-0.

[4] BOJAR, B. (oprac.) *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych*. Warszawa: Stowarzyszenie Bibliotekarzy Polskich, 2002, s. 27. ISBN 83-87629-84-7 oraz ISO 2789:2006(E) *Information and documentation – International library statistics*, Fourth edition: 2006-09-15, ISO 2006, s. 6.

[5] DUCH, W. *Notatki do wykładów wstępnych o komputerach i oprogramowaniu* (wykład 10) [on-line]. Wydział Fizyki, Astronomii i Matematyki Stosowanej UMK [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://www.fizyka.umk.pl/~duch/Wyklady/komput/spdypl.html>.

[6] *ArchiDoc – Elektroniczne repozytorium dokumentów* [on-line]. ArchiDoc [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://www.archidoc.pl/index.php?option=com_content&task=view&id=140&Itemid=39.

[7] Repozytorium. W: *Wikipedia. Wolna encyklopedia* [on-line]. [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Repozytorium>.

[8] Tamże.

[9] WERLA, M. *Biblioteka cyfrowa jako repozytorium OAI-PMH* [on-line]. dLibra. Poznańskie Centrum Superkomputerowo-Sieciowe [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://dlibra.psnk.pl/community/display/KB/Biblioteka+cyfrowa+jako+repozytorium+OAI-PMH>.

[10] *Konfiguracja DeskLight z bazą danych Postgres* [on-line]. YADDA [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://yaddainfo.icm.edu.pl/DeskLight/download/docs/DeskLight_I_Postgres.pdf.

[11] WESOŁOWSKI, T., KUDŁACIK, P., GOLA, M. *Repozytorium systemu wspomaganie lingwistycznego analityka obiektowego LIANA* [on-line]. Politechnika Śląska. Instytut Informatyki. Zakład Oprogramowania [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://sun.aei.polsl.pl/~pkudlac/resources/articles/rep_syst_LIANA.pdf.

- [12] NAHOTKO, M. Cyfrowa nauka-cyfrowe publikacje-cyfrowe biblioteki. *Przegląd Biblioteczny* 2007, z. 1, s. 19-22. ISSN 0033-202X.
- [13] Biblioteka cyfrowa. W: Wikipedia. Wolna encyklopedia [on-line]. [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://pl.wikipedia.org/wiki/Biblioteka_cyfrowa.
- [14] Tamże.
- [15] SZCZEPAŃSKA, B. *Otoczenie prawne biblioteki cyfrowej* [on-line]. Biblioteka Publiczna m. st. Warszawy [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://www.koszykowa.pl/biblioteka/files/11%20-%20Szczepanska%20Barbara%20-%20Otoczenie%20prawne%20biblioteki%20cyfrowej.pdf>.
- [16] WREMBEL, R. *Wykład 1: wprowadzenie do problematyki baz danych* [on-line]. Uczelnia on-line. Wydział Matematyki, Informatyki i Mechaniki UW [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://wazniak.mimuw.edu.pl/images/7/74/BD-2st-1.2-w01.tresc-1.1.pdf>.
- [17] POTĘGA, J. *Narodowe repozytorium dokumentów elektronicznych* [on-line]. Śląska Biblioteka Cyfrowa [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://www.sbc.org.pl/Content/10978/15_Potega_Narodowe_repozytorium.pdf.
- [18] MAZUREK, C. [i in.] *Infrastruktura bibliotek cyfrowych w sieci PIONIER* [on-line]. dLibra. Poznańskie Centrum Superkomputerowo-Sieciowe [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://dl.psnc.pl/biblioteka/dlibra/doccontent?id=155>.
- [19] WREMBEL, R. dz. cyt.
- [20] POTĘGA, J. dz. cyt. Por. też KARWASIŃSKA, E., RYCHLIK, M. Repozytorium instytucjonalne jako czynnik wspomagający rozwój nauki w środowisku akademickim. *Biblioteka* 2007, nr 11, s. 153–167. ISSN 1506-3615.
- [21] *Biblioteka Cyfrowa – Zachodniopomorski System Informacji N@ukowej i Region@lnej* [on-line]. Książnica Pomorska [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://www.ksiaznica.szczecin.pl/kp/index.php?s=6&ai=407&ak=2>.
- [22] BOJAR. B. (oprac.), dz. cyt., s. 27–28.
- [23] *ISO 2789:2006 Information and documentation...*, s. 7.
- [24] *System rozproszonych baz danych* jest zbiorem współpracujących z sobą baz danych, z których każda znajduje się na innym serwerze. Z punktu widzenia użytkownika wszystkie te bazy logicznie stanowią jedną rozproszoną bazę danych. Więcej: WREMBEL, R. dz. cyt.
- [25] SKUBAŁA, E., KAZAN, A. Dobór publikacji w polityce zarządzania biblioteką cyfrową. Analiza wykorzystania zasobów bibliotek cyfrowych wspólnie przeszukiwanych w ramach FBC [on-line]. Repozytorium Zespołu Bibliotek Cyfrowych PCSS [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://dl.psnc.pl/biblioteka/dlibra/publication/150/content>.
- [26] *Zestawienie liczby duplikatów* [on-line]. Federacja Bibliotek Cyfrowych [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://fbc.pionier.net.pl/owoc/duplicates-matrix-report>.
- [27] Zob. JUCHA, D. Biblioteczne piractwo. *TEMI. Galicyjski Tygodnik Informacyjny*, 28.01.2009, s. 23. ISSN 0208-7006.

[28] WERLA, M. Re: *Przeszukiwanie pełnotekstowe* [on-line]. Do: M. Kowalska. 14 kwietnia 2009, 10:44 [Dostęp 15 kwietnia 2009]. Korespondencja osobista.

[29] Przykłady tytułów wydawnictw ciągłych rejestrowanych przez bazy tworzone dla pokrewnych dyscyplin naukowych: 1. bazy EDUKACJA i PEDAGOG: „Edukacja i Dialog”, „Nowa Szkoła”, „Problemy Opiekuńczo-Wychowawcze”, „Ruch Pedagogiczny”, „Szkoła Specjalna”, „Wychowanie na co Dzień”, „Wspólne Tematy”; 2. BazTech i BIMET: „Automatyka Elektroenergetyczna”, „Biuletyn Instytutu Odlewnictwa”, „Materiały Budowlane”, „Nafta-Gaz”, „Pneumatyka”, „Postępy Techniki Jądrowej”, „Rudy i Metale Nieżelazne”, „Wiertnictwo, Nafta, Gaz”; 3. AGRO i SIGŻ: „Acta Agrophysica”, „Acta Biochimica Polonica”, „Acta Scientiarum Polonorum. Piscaria”, „Agrotechnika”, „Herba Polonica”, „Kwiaty”, „Przegląd Piekarski i Cukierniczy”, „Polish Journal of Ecology”; 4. Polska Bibliografia Wojskowa i BAZHUM: „Arcana”, „Archeion”, „Barok”, „Biuletyn Instytutu Pamięci Narodowej”, „Czasopismo Prawno-Historyczne”, „Karta”, „Kwartalnik Historyczny”, „Meander”, „Nasza Przeszłość” itd.

[30] Wprawdzie wiele baz danych podaje informację o istnieniu bibliografii załącznikowej (np. AGRO, SIGŻ, DOROBK, PEDAGOG, Bibliografia Etnografii Polskiej), a nawet wskazuje liczbę uwzględnionych w niej pozycji (np. BIMET, SPORT, BASPAW, Polska Bibliografia Lekarska) czy numery stron, na których ją umieszczono (np. „Bibliografia Geografii Polskiej”, „Bibliografia Publikacji Pracowników AGH”, „Bibliografia publikacji pracowników PK”), lecz nie przybiera to – jak w BazTechu – postaci pełnego wykazu publikacji.

[31] DERFERT-WOLF, L. Odkrywanie niewidzialnych zasobów sieci. W: *E-prints in Library and Information Science* [on-line]. [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://eprints.rclis.org/8560/1/derfert_CPI.pdf.

[32] HOLLENDER, H. Biblioteki cyfrowe w Polsce i BWN: w poszukiwaniu drogi racjonalnego rozwoju. W: *Konferencja na temat: Polskie biblioteki akademickie w Unii Europejskiej, Łódź, 23–25 czerwca 2004 r.: materiały konferencyjne* [on-line]. Łódź: Wydawnictwo Politechniki Łódzkiej 2004 [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: http://bg.p.lodz.pl/konferencja2004/pelne_teksty/hollender.doc.

[33] Tamże.

[34] Por. komentarze do wpisu TARKOWSKI, A. Konferencja „Polskie biblioteki cyfrowe”: dwie ważne statystyki o polskich zasobach. W: *Kultura 2.0* [on-line]. 26.11.2008 16:22 [Dostęp 15 kwietnia 2009]. Dostępny w World Wide Web: <http://kultura20.blog.polityka.pl/?p=682>.

Zasoby czasopiśmiennicze w polskich bibliotekach cyfrowych i bibliograficznych bazach danych - koegzystencja, konkurencja czy kooperacja? / Małgorzata Kowalska, // 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/kowalska.php>