

MAREK NAHOTKO

Instytut Informacji Naukowej i Bibliotekoznawstwa
Uniwersytet Jagielloński
e-mail: marek.nahotko@uj.edu.pl

INTEGRACJA WYSZUKIWANIA W ZASOBACH INFORMACYJNYCH



Marek Nahotko, adiunkt w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego, specjalizuje się w zagadnieniach metadanych dokumentów elektronicznych i wykorzystania Internetu w komunikacji naukowej i działalności informacyjnej. Jest autorem książek *Komunikacja naukowa w środowisku cyfrowym* (Warszawa 2010), *Naukowe czasopisma elektroniczne* (Warszawa 2007), *Opis dokumentów elektronicznych. Teoretyczny model i możliwości jego aplikacji* (Kraków 2006), *Metadane: sposób na uporządkowanie Internetu*. (Kraków 2004) oraz licznych artykułów publikowanych m.in. na łamach „Przełądu Bibliotecznego”, „Zagadnień Informacji Naukowej” oraz „Biuletynu EBIB”.

SŁOWA KLUCZOWE: Zintegrowane wyszukiwanie informacji. Wyszukiwarki internetowe. Katalogi biblioteczne. WorldCat. Google Books. Google Scholar.

ABSTRAKT: W artykule przedstawiono prace służące ujednoczeniu środowiska wyszukiwania informacji naukowej dostępnej w Internecie i bibliotekach naukowych. Ponieważ wyszukiwanie zintegrowane (federated serach) nie funkcjonuje wystarczająco efektywnie w środowisku sieci globalnych, przedstawiono dwie inne drogi: integracja przez wyszukiwarkę Google Scholar oraz nowe prace nad wyszukiwaniem zintegrowanym w skali Webu, prowadzonym poprzez interfejs biblioteki. Wskazano na zalety i wady obu rozwiązań oraz zaproponowano ich połączenie.

WPROWADZENIE

W swoim artykule zatytułowanym *Metadane biblioteczne w XXI wieku* (Nahotko, 2010b) przedstawiłem zapytania na stan obecny bibliotekarstwa zarówno polskiego, jak i światowego, gdyż są one mocno ze sobą powiązane. Truizmem jest bowiem stwierdzenie, że współczesne bibliotekarstwo, szczególnie naukowe, funkcjonuje obecnie jako system globalny, w czym podobne jest do sieci globalnych, takich jak Internet. Jednym z podstawowych wniosków z lektury wspomnianego artykułu jest stwierdzenie o końcu monopolu bibliotek naukowych na gromadzenie

i udostępnianie informacji. W świecie publikacji drukowanych biblioteki nie miały w tym zakresie konkurencji. We współczesnym świecie, w którym coraz bardziej znaczącą, a w niektórych zakresach główną, rolę odgrywają publikacje elektroniczne, biblioteki tracą tę uprzywilejowaną pozycję.

Jeżeli ekspansję Internetu i innych mediów cyfrowych (choćby telefonii, ale także np. telewizji cyfrowej) jako nowego systemu informacyjnego – użytecznego w życiu codziennym, a niezbędnego w bibliotekach naukowych – przyjmiemy do wiadomości, to pozostaje nam zastanowić się nad strategią dalszego postępowania z punktu widzenia interesu bibliotek. Pierwszą postawą bywa chęć konfrontacji z internetowymi firmami komercyjnymi, stanowiącymi konkurencję dla bibliotek. Z takiej postawy po części wynika np. krytyka Google i jego działań w zakresie digitalizacji. Firma ta krytykowana jest m.in. przez ALA, które, w swojej opinii, wyraziło przypuszczenie, że Google mógłby ustanowić ceny za dostęp do zdigitalizowanych zasobów na poziomie maksymalizującym zyski, będącym poza możliwościami bibliotek (ALA Comments, 2009, p. 6). Można też przeczytać, że jeśli zbiory biblioteczne mogą być zdigitalizowane tylko raz, to lepiej żeby nikt ich nie digitalizował, niż żeby zrobiła to firma Google (Nunberg, 2009; Sroka, 2009, s. 18). Ostatecznie taką postawę można przedstawić jako: niech oni sobie robią, co chcą, my mamy własną tradycję, będziemy robić po swojemu, od czasu do czasu pozywając ich do sądu, co pozwoli na trwanie bez zmian. Podobną postawę Bradford Eden wymownie nazywa modelem wymierania (Eden, 2010, p. 96). Mam wrażenie, że w takim sposobie myślenia niektórzy bibliotekarze utwierdzani są przez część firm komercyjnych, działających na rynku dostawców usług i produktów przeznaczonych dla bibliotek, a wyraźnie konkurujących z takimi firmami, jak Google, czy takimi inicjatywami, jak Open Access.

Inną postawą jest próba współuczestnictwa w stosowaniu nowych, globalnych technologii i wyciągnięcia z niego własnych korzyści, pamiętając, że korzyści muszą odnosić także inni. Przez B. Edena jest ona nazywana zmianą narzędzi. Takie podejście jest trudniejsze, gdyż oznacza konieczność wprowadzenia do bibliotek poważnych zmian, a w dodatku często do końca nie wiadomo jakich, bo sytuacja jest nadzwyczaj dynamiczna. Można powiedzieć, że biblioteka powinna być tak zmienna, jak zmienne jest jej otoczenie, sieci rozległe i jak zmieniają się jej użytkownicy. Można więc tylko próbować nadążać za zmianami, mając świadomość niebezpieczeństwa zabrnienia w ślepią uliczkę.

Ponieważ uważam drugą opcję za lepszą, chociaż może trudniejszą, w dalszej części artykułu zajmę się sposobami jej praktycznej realizacji, przedstawiając propozycje konkretnych działań w jednym, wybranym zakresie: katalogów bibliotek naukowych. Są to nowe zagadnienia, ale częściowo podejmowane już przez niektóre biblioteki, także polskie (Ozóg i Potempa, 2009, s. 102; Hollender, 2011, s. 28).

DIAGNOZA

Wiele badań wskazuje na to, że użytkownicy doceniają jakość zasobów bibliotecznych, ale zdecydowanie chętniej korzystają z zasobów sieciowych,

nawet jeśli uznają je za mniej wiarygodne (por. Law, 2008, p. 12; Gauder, ed., 2010, p. 32; Nahotko, 2010b, p. 171). Według Johna Law studenci wskazują na trzy podstawowe problemy podczas korzystania z bibliotek:

1. Trudno odnaleźć miejsce, w którym należy rozpocząć wyszukiwanie. Jeżeli przyjrzymy się stronom Web bibliotek akademickich, to znajdziemy tam setki różnych obiektów, linków do różnych miejsc, które w większości są interesujące przede wszystkim dla samych bibliotekarzy (np. Czy użytkownika rzeczywiście interesuje, skąd biblioteka bierze pieniądze? Informacje o uzyskanych grantach nie powinny zajmować głównego miejsca). Czasem trzeba stracić wiele czasu i dokonać kilku prób poszukiwań, aby trafić do katalogu biblioteki lub udostępnianych przez nią baz danych, a współczesny użytkownik musi otrzymać to, czego szuka natychmiast albo zrezygnuje z poszukiwań.

2. Problemy z dotarciem do odpowiednich zasobów. Użytkowników nie interesuje, czy zasoby są tworzone w bibliotece, czy przez nią kupowane, czy dostępne bezpłatnie. Takie podziały nie mają dla niego znaczenia, bo on za nic nie płaci. Mało istotny jest też dla nich podział ze względu na formę: zarówno książki, jak artykuły i raporty zamieszczane w Internecie są równie użyteczne.

3. Ogólnie użytkownicy nie wiedzą o zasobach udostępnianych przez sieciowy interfejs biblioteki, tym bardziej, jeżeli do tych zasobów nie skierowują stosowane przez nich narzędzia – wyszukiwarki internetowe.

Problemy te powodują, że wynika to z badań OCLC, opublikowanych pod red. Brada Gaudera (Gauder, ed. 2010) i wcześniejszych (Rosa et al., 2005), 84% użytkowników rozpoczyna wyszukiwanie informacji od wyszukiwarki internetowej, nikt natomiast nie rozpoczyna do strony biblioteki. Te ostatnie są jednak używane przez 33% badanych Amerykanów. Dowodzi to, iż są oni do tych stron kierowani przez wyszukiwarki. Podobna tendencja dotyczy użytkowników polskich bibliotek cyfrowych, którzy w 44% przypadków do serwisu kierowani byli przez wyszukiwarkę Google (Mazurek, 2009). Biblioteka w świadomości większości użytkowników istnieje więc pod warunkiem, że zostaną skierowani do niej poprzez wyszukiwarkę.

Aby zrozumieć kierunki pożądanych zmian w bibliotece, należy uświadomić sobie, że w otaczającym środowisku również zachodzą zmiany, wpływające na potrzeby użytkowników i możliwości, jakimi dysponują. Jak twierdzą Alison Head i Michael Eisenberg, studenci – użytkownicy bibliotek wyszukiwanie w Internecie uważają za wygodne, stosunkowo proste i łatwo dostępne. Z drugiej strony, korzystając z wyszukiwarek internetowych, narzekają na zbyt wielkie zasoby (przeładowanie informacją), wręcz zalewające poszukującego (Head i Eisenberg, 2008, p. 2). Być może jednak rację mają ci, którzy, jak Paul Levinson, twierdzą, że nie przeładowanie, a swoiście rozumiane niedoładowanie informacyjne (ang. information underload) jest prawdziwym problemem (Levinson, 2010, p. 194). Polega ono na tym, że odbiorcom brak wiedzy, by skutecznie i najefektywniej korzystać z bogactwa Sieci. Wykazują oni na przykład brak umiejętności właściwej selekcji materiałów ze względu na ich jakość. Szczególnie problemy stwarza odszukanie potrzebnej informacji, o istnieniu której wiadomo i która potrzebna jest natychmiast. Jak twierdzi Levinson, umysł ludzki jest jednak z natury dostosowany do rozwiązywania takich

problemów – nasze mózgi pozwalają dostrzegać nam sens w oszałamiającym i straszliwym zamęciu świata, w którym żyjemy. Potrzebne są tylko odpowiednie narzędzia i umiejętność posługiwania się nimi.

Jedna z wymienionych cech – natychmiastowy dostęp – jest szczególnie istotna. Współczesny użytkownik biblioteki i Internetu przyzwyczajony jest do otrzymywania żądanej informacji natychmiast. Jeśli informacja dostępna jest ze zwłoką, rezygnuje on z oczekiwania, poszukując kolejnego źródła. W taki sposób użytkownik funkcjonuje w Internecie, gdzie istnieją konkurencyjne źródła informacji i można natychmiast przechodzić od jednego do drugiego; zwłoka w reakcji oznacza niewydajne źródło, które należy zastąpić innym. Studenci rozpoczynają wyszukiwanie dokładnie w czasie, gdy informacje są niezbędne, a więc zwykle w ostatniej chwili przed końcowym terminem wykonania pracy. W efekcie trafiają do źródeł takich, jak Wikipedia, która nie tylko ma przyjazną dla nich formę portalu Webu 2.0, ale także dostępna jest w każdej chwili. Biblioteka ceniona jest za możliwość skorzystania z jej baz danych oraz za pomoc, jakiej udzielają bibliotekarze, wskazując pożyteczne źródła informacji trudno dostępnej (np. dane statystyczne) (Head & Eisenberg, 2008, p. 11).

Korzystanie z bibliotek stwarza problemy niespotykane w Sieci, co powoduje, że stają się one mniej atrakcyjne dla studentów (Budd, 2009, p. 3). Użytkownicy wiedzą, że biblioteka posiada przydatne materiały, ale ich wydobyć nastęrcza zasadnicze problemy. Wiąże się z umiejętnością odpowiedzi na takie pytania, jak: Gdzie wyszukiwać? Które zasoby zawierają właściwe dane? Jak od metainformacji opisowej przejść do metainformacji o dostępności, a stamtąd do informacji? Badania wskazują na to, że biblioteka, pomimo wspomagania swoich użytkowników w wyszukiwaniu, lokalizacji i ocenie zasobów, postrzegana jest jako instytucja niewygodna w korzystaniu i onieśmiałająca (McClure & Clink, 2009, p. 121). Interpretacja wyników wyszukiwania w OPAC także stwarza problemy. Zdarza się, że ponad 30% wyszukiwań, prowadzonych przez niedoświadczonych użytkowników, kończy się pustym wynikiem lub daje bardzo mało odpowiedzi (Burke, 2010, p. 4). Po takim niepowodzeniu użytkownicy wracają do łatwiejszych narzędzi, zwykle dających duże zbiory odpowiedzi, co powoduje, że 93% studentów używa Internetu do wyszukiwania źródeł do swoich prac (Gauder, ed., 2010, p. 53).

Dotąd opisane problemy typowe są dla studentów, jednak pracownicy naukowi także napotykają kłopoty w poszukiwaniu informacji. Według Rossa Housewrighta i Rogera Schonfelda naukowcy postrzegają bibliotekę jako instytucję realizującą trzy zadania: zakupy zbiorów, ich archiwizację i pośrednictwo w ich udostępnianiu. O ile dwie pierwsze funkcje nie wzbudzają zastrzeżeń (jeśli pominiemy stałe braki w finansowaniu), to rola biblioteki jako zasobu i pośrednika w realizacji badań zmniejsza się, szczególnie w naukach ścisłych i przyrodniczych. Uczni coraz częściej zwracają się ku zasobom elektronicznym, co powoduje, że choć doceniają zbiory biblioteki, stają się w swych pracach naukowych i dydaktycznych coraz bardziej od nich niezależni; tylko 8% uczonych zaczyna poszukiwania od odwiedzin budynku bibliotecznego, 25% od OPAC (Housewright i Schonfeld, 2006, p. 9). Nie dziwi to, gdy uwzględnimy badania, wskazujące, że np. 40% brytyjskich uczonych twierdzi, że mają przynajmniej raz w tygodniu problem z dostępem do zasobów licencjonowanych w swoich

bibliotekach. Jako główny problem wskazują brak odpowiedniego rocznika czasopisma w wersji elektronicznej (RIN, 2009, p. 7). Biblioteka więc w swoich użytkownikach wywołuje coraz bardziej ambiwalentne uczucia.

Oceniając te wyniki badań Carol Tenopir, wskazała na narastający, według niej, problem, który określiła jako „lukę wartości”. Kolekcje biblioteczne co prawda wymagają wielkich inwestycji, ale są największym atutem tych instytucji, pozwalającym na konkutowanie na rynku informacji. Jednak złożoność procedur i bariery wyszukiwawcze ograniczają pełne wykorzystanie tego kapitału. Dalsze inwestowanie w zbiory, dzięki czemu następuje ich dostosowywanie do potrzeb lokalnej społeczności użytkowników, nie wpływa na zmniejszenie barier związanych z ich udostępnianiem. Te z kolei powodują spadek możliwości wykorzystania zasobów bibliotecznych. W rezultacie każde zakupy i aktualizacja zbiorów dają mniejsze korzyści, niż należałoby oczekiwać. Nie można jednak przestać inwestować, bo spowoduje to dewaluację posiadanego kapitału (Tenopir, 2009). Efektem jest utrata przez biblioteki znaczącej roli w udostępnianiu piśmiennictwa, w szczególności czasopism. Biblioteki pełnią w tym zakresie głównie rolę kanału transferu środków do wydawców (Górny, 2009, p. 34). Znamienna jest w tym kontekście rezygnacja z prenumeraty w postaci drukowanej na rzecz wersji elektronicznych, zarówno na świecie (Cox & Cox, 2010, p. 14), jak i w Polsce (dane AFBN – zob. <http://ssk2.bu.amu.edu.pl/standaryzacja/>).

Bibliotekarze często nie zauważają, że użytkownicy jednakowo traktują wszystkie obiekty informacyjne. Artykuł odpowiedni dla ich potrzeb ma dla nich taką samą albo wyższą wartość niż cała książka. Jeżeli biblioteka udostępnia bazy danych, oznacza to potrzebę organizacji łatwego dostępu do milionów obiektów. Użytkownicy nie rozróżniają informacji według rodzaju lub formatu. Dla nich znaczenie ma użyteczność i łatwość dostępu. Stąd udostępnianie informacji w odrębnych kolekcjach – osobno książki, osobno artykuły, gdzie indziej fotografie i filmy – jest nieadekwatne do potrzeb użytkowników. Informacja powinna być dostępna w wyniku pojedynczego wyszukiwania, które powinno być proste i szybkie (Burke, 2010, p. 5).

KIERUNKI ZMIAN

Powstaje pytanie, czy problemy opisane w tym i poprzednim moim artykule nieuchronnie muszą prowadzić do degradacji bibliotek w powstającym „społeczeństwie informacyjnym”? Jakie są możliwości zainteresowania kolejnych pokoleń użytkowników oferowanymi usługami? W dalszej części artykułu chcę przedstawić możliwości bardziej atrakcyjnej, z punktu widzenia użytkowników bibliotek naukowych, organizacji wyszukiwania metainformacji i udostępniania informacji. Zmiany rozwijają się obecnie w kilku kierunkach:

- od wyszukiwania zintegrowanego,
- poprzez nowe interfejsy użytkownika, dające nowe możliwości wyszukiwania i pracy w serwisach społecznościowych,
- do wyszukiwania zintegrowanego w skali całego środowiska Web; integracja może następować od strony wyszukiwarek lub bibliotecznego OPAC.

WYSZUKIWANIE ZINTEGROWANE

Korzystanie z ogromu rozproszonych zasobów Webu powoduje znaczne uciążliwości. Jedną z możliwości scalania tych zasobów jest realizacja tzw. wyszukiwania zintegrowanego (ang. federated search), nazywanego również metawyszukiwaniem. Pozwala ono na przygotowanie odpowiedzi na pojedyncze zapytanie użytkownika na podstawie zawartości wielu, wybranych baz danych, dostępnych w Internecie. Dostęp ten może być płatny lub bezpłatny, jak np. w przypadku OPAC bibliotecznych. Podstawą metawyszukiwania jest protokół Z39.50, który stanowi sposób na zapewnienie współdziałania w zakresie wyszukiwania w rozproszonych zasobach Internetu. Tego rodzaju wyszukiwanie jest prowadzone między różnymi systemami baz danych, dzięki temu, że lokalny system informacyjny akceptuje zapytania z zewnętrznych systemów wyszukiwania informacji. W tym celu ujednolicona (a przy okazji uproszczona) jest struktura metadanych, akceptowana przez wszystkie współpracujące systemy. Pierwsze tego typu rozwiązania pojawiły się w 1999 r. i można je nazwać prekursorami dzisiejszych rozwiązań w zakresie integracji źródeł informacji, opisanych dalej.

W obecnej sytuacji, gdy biblioteki naukowe zakupują licencje na dostęp do dziesiątek, a nawet w przypadku większych bibliotek, kilkuset baz danych, jednocześnie tworząc także własne zasoby metadanych na czele z OPAC, użytkownicy czują się zagubieni w tej obfitości. Pojawiające się w bibliotekach naukowych systemy wyszukiwania zintegrowanego są próbą rozwiązania tych problemów. Ponieważ użytkownicy preferują wyszukiwanie za pomocą wyszukiwarek Web, bibliotekarze uznali wyszukiwanie zintegrowane za sposób na wykorzystanie funkcjonalności tych wyszukiwarek w zastosowaniu do wybranych zasobów Internetu, często stanowiących wcześniej część ukrytego Webu.

Wyszukiwanie zintegrowane daje wiele korzyści, ale jednocześnie ma pewne wrodzone ograniczenia. Biblioteki prenumerują bazy danych zawierające różne rodzaje zasobów, wykorzystywanych do różnych celów, co utrudnia formułowanie zapytań. Podczas wyszukiwania zintegrowanego zapytania użytkownika muszą być tłumaczone przez uproszczony interfejs wyszukiwawczy na wiele formalnych języków wyszukiwania, stosowanych w zróżnicowanych zasobach. Scalanie w jeden, sensownie skomponowany zbiór wielu zestawów wyników, uzyskanych w trakcie wyszukiwania w pełnotekstowych bazach danych, indeksach cytowań, OPAC-ach, bazach wyszukiwarek internetowych, repozytoriach i archiwach cyfrowych jest zadaniem zarówno metodologicznie, jak i technicznie bardzo trudnym. Metody określania relewancji wyników, wykorzystywane przez wyszukiwarki takie jak Google lub Yahoo, nie dają się łatwo przenieść do wyszukiwania zintegrowanego. Największe wyszukiwarki Web przechowują informacje we własnych bazach danych i obliczają rankingi według własnych, nieznanych ogólnie kryteriów. Wyszukiwarki zintegrowane natomiast konwertują dane wejściowe użytkownika na strukturę zapytań poszczególnych baz danych, tworzonych przez wielu dostawców. W bazach tych używane są różne struktury danych, dostosowane do odmiennych algorytmów wyszukiwawczych. Z tego powodu systemy wyszukiwania zintegrowanego nie mogą zapewnić tworzenia rankingów wyników w sposób podobny do

tworzenia rankingów przez wyszukiwarki Web. Zrozumiałym efektem stosowania wyszukiwania zintegrowanego jest wzrost częstotliwości wyszukiwania we wszystkich udostępnianych bazach danych. Utrudnia to prowadzenie wiarygodnych analiz wykorzystania tych baz. Z drugiej strony często zdarza się, że użytkownicy, przyzwyczajony do konkretnej bazy danych, nie korzystają z innych, równie przydatnych serwisów. W takim przypadku wyszukiwanie zintegrowane może zwiększyć użyteczność rzadziej wykorzystywanych baz danych, z pożytkiem dla wyników wyszukiwania. W początkowym okresie rozwoju tej technologii, metawyszukiwarki funkcjonowały w sposób akceptowalnie sprawny. Wraz ze wzrostem liczby zasobów niezbędna była realizacja coraz większej liczby połączeń, co znacznie spowalnia pracę, dostosowując jej tempo do źródła najwolniejszego (Breeding, 2005, p. 27). Biblioteki narzekają także na niepoprawny często sposób wyświetlania metadanych po ich ujednoczeniu i wspomniane trudności z ich właściwym rankingowaniem (Burke, 2010, p. 7).

Można doszukać się podobieństw pomiędzy możliwościami metawyszukiwarek a usługami Google Scholar. Obie usługi kierują swoją ofertę do tej samej grupy użytkowników, pozwalając na dostęp do różnego rodzaju zagregowanych zasobów rozproszonych dostawców. Wyszukiwarka Google Scholar umożliwia dostęp do treści stanowiących typowy zasób dla metawyszukiwania, takich jak pełne teksty artykułów, metadane i abstrakty, poprzez jeden indeks i jeden interfejs wyszukiwawczy. Wyszukiwanie zintegrowane jest bardziej przydatne dla użytkowników mających dostęp do zasobów licencjonowanych; Google Scholar daje większe możliwości wyszukiwania i dostępu do zasobów Open Access (Herrera, 2007, p. 51).

INTERFEJSY UŻYTKOWNIKA

Interfejsy użytkownika, jako część oprogramowania bibliotecznego służąca komunikacji systemu z użytkownikiem, stanowią o sposobach pracy użytkownika, oferowanych przez system. Są to lokalnie instalowane aplikacje, od zawsze stanowiące część każdego zintegrowanego systemu bibliotecznego. Coraz częściej oferują one funkcje nieobecne w tradycyjnym OPAC, takie jak tzw. wyszukiwanie według rozszerzonych słów kluczowych (różnych form danego terminu, np. synonimicznych i bliskoznacznych), upodabniające wyszukiwanie do prowadzonego za pomocą języka naturalnego, które jest jednak wspierane przez automatyczną kontrolę relacji semantycznych, zachodzących między terminami wyszukiwawczymi. Wykorzystuje się także szczegółowe dane, zawarte w rekordach MARC, do tworzenia faset z zawartości podpól oraz do stosowania określników jako terminów wyszukiwawczych.

Biblioteka powinna starać się, aby jej użytkownicy chcieli ustawiać jej stronę internetową jako stronę startową w swoich komputerach, tak, jak obecnie bardzo często ustawiają stronę Google. Biblioteka to miejsce do wyszukiwania informacji: po wyszukaniu mogą one być dostarczane użytkownikom lokalnie na tradycyjnych nośnikach i/lub nośnikach elektronicznych oraz zdalnie, na nośnikach cyfrowych. Dlatego też interfejsy biblioteczne powinny maksymalnie ułatwiać wyszukiwanie informacji, na

wzór wyszukiwania za pomocą Google. Jako pierwszy powinien ukazywać się ekran, na którym znajduje się okienko do wpisywania słów – terminów wyszukiwawczych, czyli prowadzenia tzw. wyszukiwania prostego (zob. rys. 1). Narzędzia, takie jak kartoteki haseł wzorcowych (KHW) tworzone centralnie na poziomie krajowym, których zawartość jest przedmiotem wymiany na poziomie międzynarodowym, dodatkowo wspomagają wyszukiwanie. Dzięki nim system może np. odróżniać przypadki występowania nazw osobowych jako podmiotu i przedmiotu dokumentu.



Rys. 1. Interfejs typu „jedno okienko, jedno kliknięcie” w polskiej bibliotece (<http://bib2.utp.edu.pl/ipac20/ipac.jsp?profile=mag#focus>)

Prostota podstawowego interfejsu powinna iść w parze z rozbudowanymi możliwościami manipulacji metadanymi i danymi. Dobrze widziane są obecnie cechy, upodabniające interfejs wyszukiwawczy biblioteki do serwisów Webu 2.0. Jest to pożyteczna moda, która jednak w przyszłości może być zastąpiona przez inną. Dzięki stosowaniu odpowiednich narzędzi, OPAC biblioteki staje się serwisem zwanym społecznym OPAC (Social OPAC, SOPAC). Pozwala on na interaktywną współpracę użytkownika z bibliotekarzem i innymi użytkownikami w środowisku informacyjnym, stworzonym przez bibliotekę. Użytkownicy mogą np. dodawać własne etykiety (tagi) do opisów, tworzyć listy własnych preferencji, umieszczać własne zasoby. Angażowani są do wykonywania czynności zarezerwowanych kiedyś dla bibliotekarza, w tym do opracowywania zbiorów, np. audiowizualnych, przez dodawanie słów kluczowych do fotografii lub klasyfikowanie według prostych kryteriów. W ten sposób w środowisku bibliotecznym urzeczywistnia się idea tzw. *crowdsourcingu*¹ (Holley, 2010). Popularne są także wizualizacje wyników prac, w tym również wyników wyszukiwania (Osińska, 2008, s. 170). Sposób prezentacji powinien być mocno kastomizowany, tzn. ustalany indywidualnie przez użytkownika. Dane bibliograficzne nie mogą być „zamknięte” w żadnym pojedynczym zasobie, należy umożliwić swobodną manipulację nimi, czyli tzw. mashup, podobnie jak ma to już obecnie miejsce (na

¹ *Crowdsourcing* to termin utworzony przez analogię do outsourcing, oznacza poszukiwanie inspiracji i rozwiązań nie u własnych pracowników czy u konkretnych innych osób lub firm, ale w anonimowych nieokreślonych wielkich grupach ludzi na zewnątrz organizacji (Cisek 2009, s. 37).

różnych poziomach) w serwisach dla uczonych typu *zbiblios.net*, *Zotero*, *Mendeley*, *CiteULike*, *Connotea* i wielu innych, stale powstających. Interfejs biblioteczny powinien dawać podobne możliwości.

Interfejs użytkownika i możliwości, które oferuje, są bardzo ważnym elementem systemu, ale nie najważniejszym. Bardziej istotne są zasoby, do których interfejs zapewnia dostęp. Podobnie jak w wyszukiwarkach, opis jest o tyle ważny, o ile pozwala na szybką selekcję, a następnie dostęp do wybranych obiektów w formie pełnotekstowej – drukowanej lub elektronicznej. OPAC powinien być narzędziem wyszukiwawczym zapewniającym szybką selekcję i dostęp do potrzebnych materiałów.

INTEGRACJA WYSZUKIWANIA W SKALI WEBU

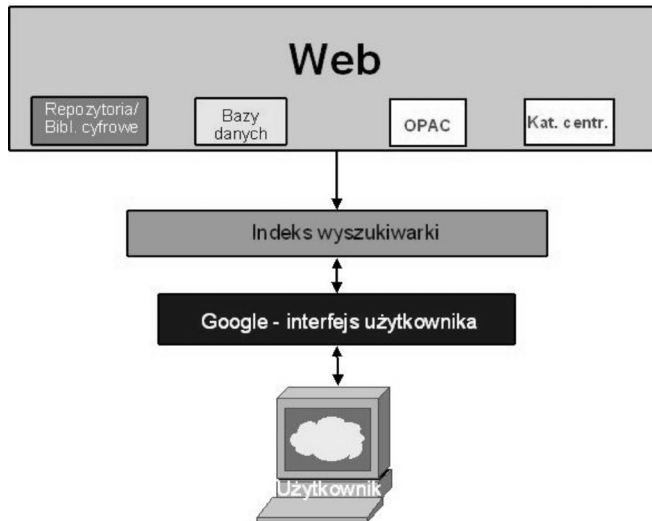
Działalność bibliotek naukowych od dawna podlega globalizacji, początkowo przejawiającej się w globalnym zasięgu przyjmowanych standardów pracy. Komputeryzacja jeszcze przyspieszyła ten proces, umożliwiając łatwą wymianę metadanych. Po upowszechnieniu się sieci globalnych najpierw metadane, a później same zasoby informacji stały się wspólnym zasobem światowym.

Globalizacja zasobów spowodowała potrzebę uniwersalnego dostępu do nich, a więc zapewnienia integracji środowiska wyszukiwania informacji. Ze względów opisanych wcześniej, multiwyszukiwanie nie może pełnić takiej roli. Integrację w skali Webu, przynajmniej z punktu widzenia bibliotekarza, można wyobrazić sobie w dwojaki sposób.

Po pierwsze wyszukiwanie może być zintegrowane wokół interfejsu i indeksu wyszukiwarki. Wyszukiwarka globalna staje się wówczas uniwersalnym narzędziem poszukiwania informacji bez względu na nośnik – fizyczny czy cyfrowy. To jest pierwotny cel realizacji takich projektów, jak *Google Books* – chodziło o masową digitalizację materiałów drukowanych, a więc zapewnienie istnienia ich wersji elektronicznych, co umożliwi wyszukiwanie ich dokładnie w taki sam sposób, jak innych obiektów w Sieci, na przykład stron Web. Być może to stało się też powodem poważnego niedostatku projektu, jakim było zignorowanie znaczenia metadanych o tradycyjnej, bibliotecznej jakości (tzw. ustrukturyzowanych) (Franke, 2011, s. 137). Zmiana tego podejścia nastąpiła po przyłączeniu do projektu kolejnych bibliotek, co spowodowało problemy z identyfikacją digitalizowanych druków, a więc trudności w unikaniu dublowania prac. W efekcie *Google* zaczął współpracować z *OCLC* – otrzymał dostęp do *WorldCat* i rozpoczął tworzenie linków do rekordów tego zasobu. W *WorldCat* pojawiły się odsyłacze do książek obecnych w zasobach *Google Books*.

Po odszukaniu za pomocą wyszukiwarki danych o informacji w wersji elektronicznej istnieje możliwość dotarcia do samej informacji, pod różnymi warunkami organizacyjnymi (np. potrzeba zarejestrowania się) i/lub finansowymi. Po odszukaniu w wyszukiwarce potrzebnych metadanych o dokumencie drukowanym użytkownicy powinni być przeniesieni w odpowiednie miejsce (katalog centralny i/lub biblioteki lokalnej, księgarnia, strona Web), pozwalające na dostęp do wyszukanej książki w każdej istniejącej formie (drukowanej lub cyfrowej).

Takie możliwości daje integracja narzędzi typu Google Scholar, Google Books, metadanych w Amazon i odnośników wspomnianych wcześniej serwisów społecznościowych z bibliotecznymi OPAC (Nahotko, 2010a, s. 177). W efekcie współpracy twórców wyszukiwarek internetowych z właścicielami najpotężniejszych światowych zasobów metadanych (głównie OCLC i wielkimi bibliotekami) oraz zastosowaniu resolverów typu OpenURL powstaje nowy sposób korzystania z katalogów bibliecznych w Internecie, jako jednolitym środowisku informacyjnym (rys. 2). Dzięki integracji wielu zestawów metadanych, pochodzących z różnych miejsc, użytkownik otrzymuje całościowy obraz interesujących go zasobów.



Rys. 2. Integracja wielu różnych zasobów informacyjnych przez wyszukiwarki

Zasoby te gromadzone i opracowywane są lokalnie, jednak informacja o nich dostępna jest globalnie. Na wielu poziomach następuje agregacja treści obiektów cyfrowych i metadanych o tych obiektach, w połączeniu z metadanymi dokumentów tradycyjnych. Jednym z takich interfejsów jest wspomniany katalog WorldCat, który w sposób typowy dla wyszukiwarek i księgarni internetowych przedstawia światowe zasoby bibliotek tradycyjnych. Ponad tym katalogiem stoi jednak wyszukiwarka, która scala metadane wszystkich zasobów w jednym indeksie. Tak więc to wyszukiwarka dostarcza uniwersalny interfejs, za pomocą którego użytkownik uzyskuje dostęp do wszystkich interesujących go zasobów, tradycyjnych (tu pełna metainformacja i dane o warunkach dostępu) i cyfrowych (tu metainformacja i często bezpośredni dostęp do obiektu).

Wyszukiwarką działającą w opisany sposób jest Google Scholar. Za jej pomocą można przejść od opisu znalezionej w Google do katalogu bibliotecznego. Pozwala ona na wybranie katalogów, do których system odsyła w przypadku odnalezienia w nich opisu wyszukanego dokumentu. Dane o dokumentach tradycyjnych pochodzą z WorldCat. Dzięki współpracy polskiego katalogu centralnego z WorldCat, jedną z możliwości wyboru jest NUKAT (Burchard i Kasprzyk, 2007). Z tego katalogu centralnego

można dalej przejść do katalogu biblioteki lokalnej i zamówić wypożyczenie interesującej pozycji². Jednocześnie wyszukiwarka pozwala odszukać informacje na zadany temat również w zasobach internetowych, w wyniku czego użytkownik posiada kompletny obraz interesujących go źródeł. Podobnie działa Google Books, zintegrowany z WorldCat oraz z Amazonem, dzięki czemu w jednym miejscu docelowo dostępne będą: teksty książek z domeny publicznej, możliwość dotarcia do pozostałych książek poprzez wyszukiwanie pełnotekstowe i według metadanych oraz lokalizacja poszukiwanej pozycji w lokalnej bibliotece lub księgarni internetowej (Franke 2011, s. 136).

Jak wynika z cytowanych wcześniej badań, użytkownik nie rozpoczyna wyszukiwania od OPAC ani nawet od portalu bibliotecznego, lecz od wyszukiwarki. Ona pozwala mu na bezpośredni dostęp do poszukiwanego obiektu lub w razie potrzeby kieruje go do odpowiedniego serwisu, na przykład biblioteki – cyfrowej lub tradycyjnej (do jej OPAC) – albo strony wydawcy. Biblioteka, wraz ze swoim systemem bibliotecznym i OPAC, włączona jest w globalne środowisko informacyjne, zamiast stać obok niego, tak jak działa się wcześniej; staje się częścią obszernej platformy Web, gdzie udostępniane dane są podstawą tworzenia nowych usług i serwisów.

Google Scholar stał się konkurencją zarówno dla systemów wyszukiwania informacji w repozytoriach, jak i wyszukiwania zintegrowanego realizowanego przez biblioteki akademickie. Dzięki temu, że stanowi prawie kompletną bazę zawartości czasopism naukowych, stał się także poważną konkurencją dla komercyjnych, bibliograficznych i abstraktowych baz danych, które gwałtownie tracą użytkowników (Chen, 2010b, p. 226). Łączy on indeksowanie i wyszukiwanie pełnotekstowe z wyszukiwaniem na podstawie standardowych metadanych, tworzonych w repozytoriach zgodnie ze standardem OAI. Jednym z obiecujących sposobów wykorzystania metadanych OAI jest wyszukiwanie nie tylko na poziomie poszczególnych obiektów cyfrowych, ale również na poziomie kolekcji (Mischo, 2005), co ułatwia odszukanie właściwych baz danych.

Już obecnie odczuwalny jest wpływ sposobu funkcjonowania takich wyszukiwarek, jak Google i Google Scholar na pracę całego Internetu. Znaczna część ruchu generowana jest właśnie przez te wyszukiwarki (Lubovitz et al., 2009). Skierowują one użytkownika wprost do pełnego tekstu z pominięciem metadanych (zarówno tekst, jak i metadane dostępne są do indeksowania). Powoduje to pewne konsekwencje. Ponieważ użytkownicy stosują głównie Google jako narzędzie wyszukiwawcze, metadane tworzone przez bibliotekarzy nie są bezpośrednio wykorzystywane do wyszukiwania. Oznacza to, że tworzenie metadanych, w szczególności metadanych ustrukturyzowanych, służy temu procesowi w inny sposób niż dotychczas; są one indeksowane, tak jak każdy inny tekst i stają się częścią zasobów indeksowych wyszukiwarki. Tego typu metadane mają coraz większy udział w wyszukiwaniu informacji.

Integracja środowiska wyszukiwania informacji wokół wyszukiwarki może oznaczać dla bibliotek wiele problemów, o których pisałem

² Przeważający hierarchiczny układ katalogów: OPAC – krajowy katalog centralny – światowy katalog centralny, funkcjonuje poprawnie tylko w przypadku szybkiej synchronizacji ich zawartości, najlepiej w czasie rzeczywistym, co nie zawsze ma miejsce.

w poprzednim artykule³. Jeżeli poważna część druków zostanie zdigitalizowana, a produkcja wydawnicza ulegnie cyfryzacji⁴, to zasoby zarówno metadanych, jak również informacji gromadzonych w bibliotekach, stracą swoją dotychczasową wartość dla użytkowników. W zakresie publikacji naukowych sytuacja taka jest łatwa do wyobrażenia. Bibliotekarze staną się wtedy agregatorami zasobów sieciowych i specjalistami od wyszukiwania informacji, włączanymi do zespołów badawczych (Luce 2008, p. 2), a biblioteki w dotychczasowej postaci będą zanikać (Wisner, 2001, p. 68).

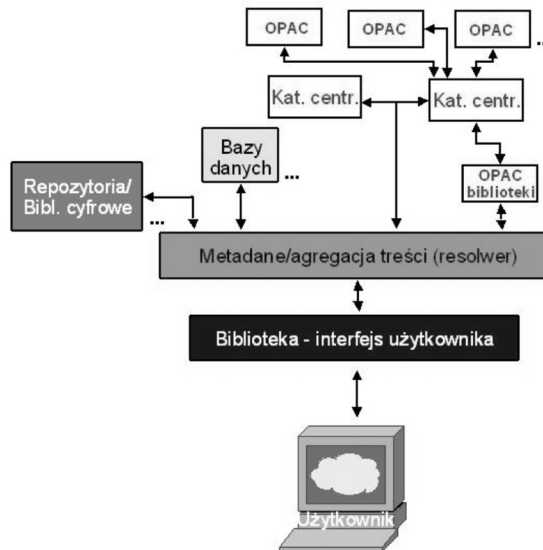
Drugą możliwością, poza integracją wyszukiwania w zróżnicowanych zasobach przez wyszukiwarke, jako „interfejs pierwszego kontaktu”, jest przejęcie tej roli przez rozproszone interfejsy bibliotek, co zaproponował już w 2005 r. Marshall Breeding (Breeding, 2005, p. 28). Prace realizacyjne zostały już rozpoczęte (Wisniewski, 2010). Polegają one na zintegrowanym wyszukiwaniu metadanych o zasobach, które biblioteka chce udostępnić swoim użytkownikom: zarówno o książkach, artykułach, bazach danych, fotografiach, filmach, nagraniach dźwiękowych jak i o innych, tworzonych przez bibliotekę (OPAC) oraz kupowanych od zewnętrznych dostawców. Efektem wyszukiwania jest dostęp do dokumentu na poziomie kompletnych treści (pełnego tekstu dla dokumentów tekstowych), w formie cyfrowej lub drukowanej, a odpowiednią szybkość działania zapewnia stosowanie jednego indeksu zawczasu zgromadzonych metadanych. Wykorzystywana jest więc ideologia raczej OAI-PMH niż Z39.50.

Rozwiązanie to (rys. 3) wymaga zastosowania zupełnie nowego modelu technologicznego i organizacyjnego. Zewnętrzne (z punktu widzenia biblioteki) bazy danych znacznie przewyższają katalogi biblioteczne liczbą indeksowanych obiektów, gdyż zawierają m.in. opisy i teksty setek milionów artykułów z czasopism (Way, 2010, p. 214). Dla wszystkich obiektów tego typu należy gromadzić metadane pobierane od właścicieli baz, na podstawie odpowiednich umów. Metadane te scalane są w jeden, ujednolicony indeks. Musi on być stale aktualizowany, gdyż jego zawartość, tak jak zawartość baz danych, zmienia się bez przerwy. Dostarczane metadane są bardzo zróżnicowane co do struktury, jakości, rozmiarów opisu, jednak muszą być konwertowane na jednolity, obszerny i spójny schemat stanowiący podstawę wyszukiwania i rankingowania wyników. To może powodować utratę specjalnych możliwości wyszukiwawczych, oferowanych przez wyspecjalizowane serwisy. Z technicznego punktu widzenia rozwiązanie to nazywane jest „małym Google”; system pracuje na prekoordynowanym indeksie, traktuje jednakowo wszystkie indeksowane obiekty i wyświetla je po wyszukaniu na jednej liście wyników. Różnica polega na tym, że zintegrowany indeks składa się z ustrukturyzowanych metadanych, tworzonych przez bibliotekarzy i dokumentalistów. Przejście do opisywanych obiektów cyfrowych (pełnych tekstów) odbywa się poprzez OpenURL, dzięki czemu biblioteka może wyznaczyć zasoby dostępne dla

³ Zob. M. Nahotko: *Metadane biblioteczne w XX wieku*. „Przegląd Biblioteczny” 2010, z. 2, s. 165-178 [red.].

⁴ W tym artykule termin „cyfryzacja” oznacza stosowanie technologii cyfrowych do tworzenia dokumentów elektronicznych (np. czasopism elektronicznych) jako formy pierwotnej (*digital, born digital*), w odróżnieniu od digitalizacji, polegającej na zmianie formy analogowej na cyfrową, np. przez skanowanie. Możliwe jest także inne rozumienie tych terminów, np. digitalizacji jako przekształcanie tego, co jest zapisane analogowo (czyli skanowanie), a cyfryzacji jako przekształcanie tego, co ma być przesłane (sygnału analogowego) (Schmidt, 2011, s. 52).

jej użytkowników. System bada i przechowuje wzorce zachowań informacyjnych użytkowników, realizowanych w trakcie wyszukiwania. Podobnie jak Google, wymaga dużego oprogramowania i rozbudowanego sprzętu.



Rys. 3. Integracja zasobów informacyjnych poprzez interfejs biblioteki

Organizacyjnie jest to skomplikowane przedsięwzięcie, gdyż niezbędne są porozumienia ze wszystkimi dostawcami danych/metadanych. Dostawcy muszą rozumieć swój interes w tym przedsięwzięciu, nawet jeżeli ich zasoby łączone są z zasobami firm konkurencyjnych. Wszyscy zyskują zadowolonych użytkowników, mających możliwość stosowania bardziej przyjaznych narzędzi wyszukiwania metainformacji, zakończonego dostępem do pełnych tekstów dokumentów (Way, 2010, p. 219). Jest w tym jednocześnie element konkurencji w stosunku do produktów Google i innych wyszukiwarek globalnych. Tego typu rozwiązania oferują m.in. firmy EBSCO (Discovery Service⁵), SerialsSolutions (Summon⁶), OCLC (WorldCat Local⁷) oraz ExLibris (Primo Central⁸). Produkty te mogą być na jeszcze wyższym poziomie integrowane ze źródłami faktograficznymi i usługami oferowanymi przez bibliotekę w formie tzw. poradników (ang. guides) (Derfert-Wolf, 2011).

Zastosowanie pojedynczego indeksu do wszystkich zasobów oferowanych przez bibliotekę jest krokiem we właściwym kierunku. Zmniejsza liczbę odrębnych miejsc niezbędnych do odwiedzenia w bibliotece (również zdalnie) w celu uzyskania w miarę kompletnych wyników wyszukiwania. Jednocześnie zmniejsza się liczba interfejsów, które trzeba znać, używać i przełączać się między nimi, do jednego interfejsu, w dodatku funkcjonującego w sposób, do którego użytkownicy są coraz bardziej przyzwyczajeni – jedno okienko do wpisywania wyrażenia wyszukiwawczego, je-

⁵ Zob. <http://www.ebscohost.com/discovery/eds-about>.

⁶ Zob. <http://www.serialssolutions.com/summon/>.

⁷ Zob. <http://www.oclc.org/worldcatlocal/default.htm>.

⁸ Zob. <http://www.exlibrisgroup.com/category/PrimoCentral>.

den przycisk do rozpoczęcia wyszukiwania. Wyniki są sortowane według rankingów i szybko wyświetlane w formie podobnej jak w wyszukiwarce. Rozwiązanie to, ze względu na swoją przewagę technologiczną, nadaje się do zastosowania w urządzeniach mobilnych, które mogą stanowić przyszłość technologii cyfrowych.

WNIOSKI

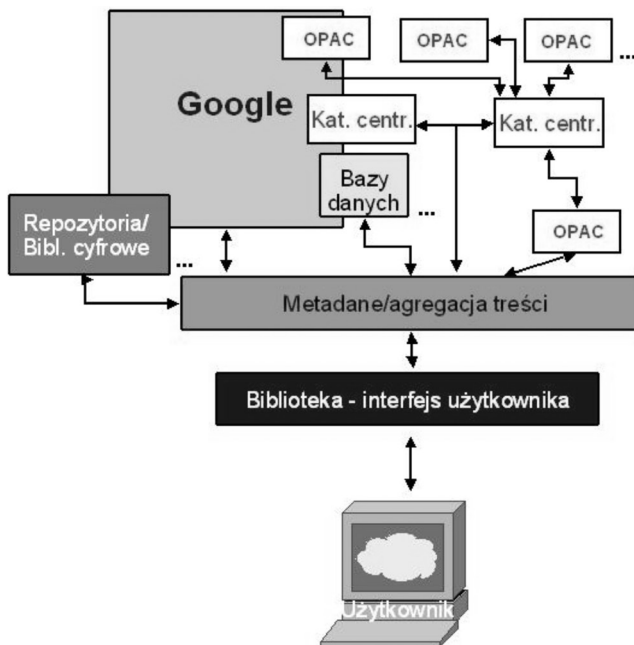
Przedstawiona sytuacja wskazuje na istnienie dwóch, niezależnych rozwiązań, służących integracji wyszukiwania w naukowych zasobach Webu: indeksu Google Scholar, który obejmuje już niemal 100% czasopiśmiennictwa naukowego i indeksów tworzonych na lokalne potrzeby poszczególnych bibliotek z zasobów, które biblioteka zdecydowała się udostępnić użytkownikom, tworząc je samodzielnie lub kupując.

Podstawową różnicą pomiędzy tymi dwoma rozwiązaniami jest sposób uzyskiwania metadanych o udostępnianych zasobach. Google opiera swą działalność na maszynowym indeksowaniu pełnych tekstów dokumentów, w tym artykułów z czasopism i metadanych bibliotecznych. Jego indeks zawiera metadane tworzone na podstawie pełnych tekstów i innych metadanych, traktowanych jak teksty. Z tego w dużej części wynikają problemy opisywane przez część autorów (Jacso, 2010, p. 176; Nunberg, 2009), a dotyczące błędów w opisach Google Scholar, takich jak i mylenie nazw autorów z innymi nazwami, błędne podawanie dat wydania itp. Błędy te jednak nie muszą mieć krytycznego znaczenia dla użytkowników, gdyż są weryfikowalne poprzez jedno kliknięcie prowadzące do pełnego tekstu.

Opisane systemy biblioteczne opierają się na wykorzystaniu ustrukturyzowanych metadanych, tworzonych przez bibliotekarzy i dokumentalistów. W takim przypadku przydatny jest schemat metadanych stanowiący „najmniejszy wspólny mianownik”, czyli taki, do którego da się sprowadzić bardziej rozbudowane schematy. W tym celu często używany jest Dublin Core, zalecany przez OAI-PMH. Scalanie metadanych może powodować problemy opisane wcześniej w tym artykule, ale ich jakość pozostaje wysoka: stosowana struktura metadanych i wiedza indeksujących uniemożliwia na ogół mylenie autorów z np. tytułami. Jest to więc bogate źródło metadanych i danych o zapewnionej, wysokiej jakości. Nie może to jednak być zasób na skalę Webu, jak czasem sugeruje nazwa nadawana przez producentów oprogramowania, gdyż taką możliwość zapewniają tylko wyszukiwarki.

W obu przypadkach w celu udostępnienia pełnego tekstu dokumentu stosowane są resolvery typu OpenURL. Resolvery te funkcjonują jako pośrednicy pomiędzy metadanymi a pełnymi tekstami dokumentów. Narzędzia te odgrywają podstawową rolę w serwisach komercyjnych i pełnotekstowych bazach danych. W wyszukiwarce Google Scholar mają znaczenie tylko pomocnicze, gdyż może ona indeksować bezpośrednio pełne teksty artykułów na stronach wydawców i podobnych serwisów. Dzięki temu zdarza się, że Google Scholar jest w stanie skierować użytkownika do odpowiedniego tekstu nawet w przypadku, gdy resolver zawodzi. OpenURL przydatny jest więc z punktu widzenia Google Scholar przede wszystkim w przypadku, gdy nie ma bezpośredniego dostępu do

pełnych tekstów zgromadzonych w bazie danych oraz wtedy, gdy należy umożliwić dostęp użytkownikom z określonej instytucji, znajdującym się poza jej siedzibą (np. poza kampusem), do zasobów udostępnianych przez tę instytucję (uczelnię lub jej bibliotekę) (Chen, 2010a, p. 82).



Rys. 4. Pełna integracja zasobów informacyjnych przez interfejs biblioteki

Oba rozwiązania opierają się na obecności pełnych tekstów naukowych w Internecie. Wynika stąd, że im więcej tekstów ma udostępniać biblioteczny system zintegrowany w skali Webu, tym więcej zasobów staje się dostępnych dla indeksów wyszukiwarki. Inicjatywa związana z bibliotekami odniesie sukces w przypadku, gdy użytkownicy będą chcieli ograniczyć wyszukiwanie do zasobów wysokiej jakości, oferowanych przez ich macierzysty system biblioteczny.

Można wyobrazić sobie kolejny etap realizacji obu rozwiązań opisanych w tej części artykułu, polegający na ich dalszej integracji. Interfejs wyszukiwawczy biblioteki naukowej powinien dawać dostęp nie tylko do zasobów udostępnianych przez nią (tworzonych samodzielnie i kupowanych), ale także do wyników wyszukiwania wyszukiwarki, czyli zasobu Internetu (rys. 4). Wyniki wyszukiwania zasobów kupowanych przez bibliotekę powinny zostać zintegrowane z wynikami utworzonymi poprzez indeks wyszukiwarki, najlepiej wyspecjalizowanej, takiej jak Google Scholar. Użytkownik powinien otrzymać jedną listę metadanych wyników ze wszystkich zasobów dostępnych we własnej bibliotece, zasobów ogólnodostępnych w innych bibliotekach (tu przydatne są takie serwisy, jak WorldCat i NUKAT), a także z indeksów wyszukiwarki, które także zawierają część zasobów udostępnianych w bibliotekach. Wówczas można mówić o wyszukiwaniu w pełni zintegrowanym w skali Webu. W obrębie

metadanych należy odpowiednio pogrupować dane o dostępie, aby użytkownik wiedział, co i gdzie może otrzymać (biblioteka własna, biblioteka inna, Internet). Naturalnie nie do wszystkich zasobów każdy otrzyma bezpośredni dostęp. Będzie on zależał od statusu użytkownika, poziomu zalogowania itp. Za dostęp do części materiałów trzeba będzie zapłacić. Do tego jednak użytkownicy są przyzwyczajeni, bo wyszukiwarki też nie dają bezwzględnego dostępu do wszystkich wyszukanych zasobów. Przy okazji może się okazać, że wiele materiałów udostępnianych komercyjnie, w Sieci jest również dostępnych bezpłatnie.

ZAKOŃCZENIE

Obecnie znikają granice pomiędzy zasobami metadanych i danych tworzonymi i gromadzonymi w bibliotekach, zamawianymi u dostawców zewnętrznych oraz odnajdowanymi dzięki wyszukiwarkom internetowym. Wydaje się, że dyskusja o tym, co jest lepsze – baza danych, w tym katalog biblioteczny (OPAC) czy wyszukiwarka internetowa – szczęśliwie dobiega końca. Jak zwykle okazało się, że najlepsze efekty daje połączenie korzyści z wszystkich dostępnych rozwiązań. Dzięki temu, że współczesne środowisko informacji ma zasięg globalny, możliwa stała się pełna integracja wszelkich zasobów, od katalogów stosunkowo niewielkich bibliotek lokalnych (na przykład uczelnianych), poprzez katalogi centralne różnego poziomu (głównie krajowe), aż po największe światowe kolosy, takie jak OCLC i Biblioteka Kongresu. Największe jednak nadzieje wiązać należy z tworzeniem czegoś, co nazwać należy globalną biblioteką hybrydową, w której wspólnie funkcjonują zasoby tradycyjne i cyfrowe, a z nimi ich metadane.

Współczesne biblioteki wstąpiły na drogę digitalizacji nie tylko opisów własnych zasobów w postaci katalogów online, ale także digitalizacji samych zasobów. Sposób opisu tych zasobów nie jest najważniejszy, pod warunkiem, że metadane istnieją, powstają w oparciu o standardowe formaty danych i tworzone są według powszechnie przyjętych zasad. Takie metadane umożliwiają współdziałanie, a więc pełną wymianę danych. Bibliotekarze stali się także hybrydowi – z jednej strony zarządzają oni zbiorami tradycyjnymi, tworząc ich cyfrowe surogaty, z drugiej przejęli funkcje wydawców obiektów cyfrowych i ich metadanych. Biblioteki są instytucjami, które łączą te dwa (analogowy i cyfrowy) światy. Natomiast wyszukiwarki lepiej obsługują wyszukiwanie obiektów cyfrowych, powstających poza działalnością bibliotek i tak zostanie do czasu, gdy różnice pomiędzy wszystkimi wymienionymi narzędziami do końca zanikną.

„Uwolnienie” danych bibliotecznych i udostępnienie usług sieciowych pozwalających na ich wykorzystanie powoduje, że ich użycie nabiera nowego wymiaru. Zamiast zamykać informację w systemie bibliotecznym należy pozwolić, aby pracowała ona na wartość dodaną, oferowaną użytkownikom bez względu na ich miejsce pobytu, lokalizację informacji i preferencje co do rodzaju usług informacyjnych, zarówno oferowanych przez Google, Amazon, portal internetowy (w tym biblioteczny), dostawcę baz danych czy sieć społeczną. W rezultacie dane biblioteczne są skalane z innymi danymi, z innych źródeł, pozwalając na tworzenie całkiem

nowych aplikacji i usług. Stąd wynika potrzeba współpracy na nowym poziomie bibliotekarzy z informatykami, wydawcami, wyszukiwarkami, dostawcami treści i baz danych.

Pełna integracja zasobów, pomimo iż jest przedsięwzięciem bardzo skomplikowanym i daleko do jej zakończenia, nie jest jednak celem ostatecznym. Co prawda maksymalne ułatwienie dostępu do zasobów jest bardzo istotne, bez niego trudno oczekiwać zadowolenia użytkowników, jednak sam dostęp przestał wystarczać. Już obecnie dostęp do informacji jest lepszy niż kiedykolwiek wcześniej. Obecnie, dla zwiększenia efektywności pracy uczonych, dydaktyków i ich studentów niezbędne jest dostarczanie usług, działających na dostępnych treściach. Oprócz dostarczania treści, przyszłością bibliotek są usługi i narzędzia pozwalające na mashup treści. Powinny one umożliwiać takie działania, jak: adnotowanie, cytowanie, porównywanie i odnajdywanie różnic, tworzenie kolejnych wersji, tworzenie zestawień, tworzenie chmury tagów, skorowidzów i indeksów, prowadzenie analizy morfologicznej, szukanie przeciwności, podobieństw, obrazowanie graficzne, podkreślanie, włączanie do sylabusów, opisywanie za pomocą słowników kontrolowanych, przedstawianie na mapie, drukowanie, kupowanie, wypożyczanie, przeglądanie, zapamiętywanie, wyszukiwanie, rozpowszechnianie, streszczanie, tagowanie, podążanie za pracami autora, podążanie za cytowaniami, tłumaczenie i wiele innych (Morgan, 2009). Warunkiem realizacji tych działań jest posiadanie pełnego dostępu do treści i to w formie elektronicznej. Obecnie taki dostęp, w wyniku zmasowanej digitalizacji i cyfryzacji, jest coraz bardziej powszechny, dzięki czemu powstają prace badawcze oparte na nowych metodach badań kolekcji liczących miliony zdigitalizowanych tekstów (Michel et al., 2011, p. 176). Metody te otwierają możliwości prowadzenia prac badawczych niedostępne nigdy w przeszłości. Wymienione funkcje i im podobne są wyznacznikami sposobu działania bibliotecznych katalogów kolejnych generacji. Implementacja serwisów i systemów opisanych w tym artykule jest krokiem w tym kierunku.

BIBLIOGRAFIA

- American Library Association (2009). *Library Association comments on the proposed settlement* [online]; [dostęp: 29.10.2010]. Dostępny w World Wide Web: <<http://wo.ala.org/gbs/wp-content/uploads/2009/05/Library-Association-Comments-on-the-Proposed-Settlement.pdf>>.
- Breeding, Marshall (2005). Plotting a new course for metasearch. *Computers in Libraries*, vol. 25, no. 2, pp. 27-29.
- Budd, John (2009). Academic library data from the United States: an examination of trends. *Libres*, vol. 19 no. 2, pp. 1-21.
- Burke, Jane (2010). *Discovery versus disintermediation: the New reality driver by today's end-user* [online]. VALA 2010 15th Biennial Conference and Exhibition [dostęp: 5.11.2010]. Dostępny w World Wide Web: <http://www.vala.org.au/vala2010/papers2010/VALA2010_57_Burke_Final.pdf>.
- Burchard, Maria; Kasprzyk, Agnieszka (2007). *Cały świat widzi NUKAT* [online]. Wrocławskie spotkania bibliotekarzy polonijnych [dostęp: 4.11.2010]. Dostępny w World Wide Web: <http://www.ebib.info/publikacje/matkonf/mat17/burchard_kasprzyk.php>.
- Chen, Xiaotian (2010a). The declining value of subscription-based abstracting and indexing services in the new knowledge dissemination era. *Serials Review*, vol. 36, no. 2, pp. 79-85.
- Chen, Xiaotian (2010b). Google Scholar's dramatic coverage improvement five years after debut. *Serials Review*, vol. 36, no. 4, pp. 221-226.

- Cisek, Sabina (2009). Dzielenie się wiedzą w Internecie. *Bibliotheca Nostra*, vol. 19, nr 3/4, s. 33-42.
- Cox, John; Cox, Laura (2010). *E-only scholarly journals: overcoming the barriers* [online]; [dostęp: 20.11.2010]. Dostępny w World Wide Web: <http://www.publishingresearch.net/documents/E-only_reportRIN2010.pdf>.
- Derfert-Wolf, Lidia (2011). Specjalista informacji 2.0? Bibliotekarz dziedzinowy 2.0? Nowa forma przewodników po zasobach – LibGuides. *Biuletyn EBIB* [online], nr 1 (119) [dostęp: 21.02.2011]. Dostępny w World Wide Web: <<http://www.ebib.info/2011/119/a.php?derfert>>.
- Eden, Bradford (2010). The new user environment: the end of technical services? *Information Technology and Libraries*, vol. 29, no. 2, pp. 93-100.
- Franko, Jerzy (2011). Digitalizacja dokumentów piśmienniczych – strategie rozwoju. W: *Cyfrowy świat dokumentu*. Pod red. Henryka Hollendra. Warszawa: CPI, s. 124-187.
- Gauder, Brad, ed. (2010). *Perceptions of libraries*, 2010. Dublin, Ohio: OCLC.
- Górny, Mirosław (2009). Komercyjne funkcje systemów informacji naukowej. W: *Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem*. Pod red. Marii Kocójowej. Kraków: IINiB UJ, s. 31-36.
- Head, Alison; Eisenberg, Michael (2009). *What today's college students say about conducting research in the digital age*. Washington: Univ. of Washington.
- Herrera, Gail (2007). MetaSearching and beyond: implementation experiences and advice from an academic library. *Information Technology and Libraries*, vol. 26, no. 2, pp. 44-52.
- Hollender, Henryk (2011). Zaczniemy przyglądać się Primo. Spotkanie w Bibliotece Narodowej. *Bibliotekarz*, nr 1, s. 28-29.
- Holley, Rose (2010). Crowdsourcing: how and why should libraries do it? *D-Lib Magazine* [online] vol. 16, no. 3/4 [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.dlib.org/dlib/march10/holley/03holley.html>>.
- Housewright, Ross; Schonfeld, Roger (2008). *Ithaka's 2006 studies of key stakeholders in the digital transformation in higher education* [online]; [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.ithaka.org/ithaka-s-r/research/Ithakas%202006%20Studies%20of%20Key%20Stakeholders%20in%20the%20Digital%20Transformation%20in%20Higher%20Education.pdf>>.
- Jacsó, Péter (2010). Metadata mega mess in Google Scholar. *Online Information Review*, vol. 34, no. 1, pp. 175-191.
- Labovitz, Craig [et al.] (2009). *Atlas Internet observatory 2009 annual report* [online]. NANOG47 Conference [dostęp: 5.11.2010]. Dostępny w World Wide Web: <http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf>.
- Law, John (2008). *Observing students researchers in their native habitat* [online]. VALA 2008 14th Biennial Conference and Exhibition [dostęp: 5.11.2010]. Dostępny w World Wide Web: <http://www.valaconf.org.au/vala2008/papers2008/45_Law_Final.pdf>.
- Levinson, Paul (2010). *Nowe nowe media*. Kraków: Wydaw. WAM.
- Luce, Richard (2008). *A new value equation challenge: the emergence of eResearch and roles for research libraries* [online]. CLIR Report 142 [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.clir.org/pubs/reports/pub142/luce.html>>.
- Mazurek, Jolanta (2009). Drogi czytelniku, jak korzystasz z biblioteki cyfrowej? Próba odpowiedzi. *Biuletyn EBIB* [online] nr 5 (105) [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.ebib.info/2009/105/a.php?mazurek>>.
- McClure, Randall; Link, Kellian (2009). How do you know that? An investigation of student research practices in the digital age. *Portal: Libraries and the Academy*, vol. 9, no. 1, pp. 115-132.
- Michel, Jean-Baptiste [et al.] (2011). Quantitative analysis of culture using millions of digitized books. *Science*, vol. 331, no. 6014, pp. 176-182.
- Mischo, William (2005). Digital libraries : challenges and influential work. *D-Lib Magazine* [online] vol. 11, no. 7/8 [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.dlib.org/dlib/july05/mischo/07mischo.html>>.
- Morgan, Eric (2009). *Web-scale discovery indexes and next generation library catalogs* [online]; [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.library.nd.edu/daiad/morgan/musings/web-scale/>>.
- Nahotko, Marek (2010a). *Komunikacja naukowa w środowisku cyfrowym. Globalna biblioteka cyfrowa w informatycznej infrastrukturze nauki*. Warszawa: Wydaw. SBP.

- Nahotko, Marek (2010b). Metadane biblioteczne XXI wieku. *Przegląd Biblioteczny*, z. 2, s. 165-178.
- Nunberg, Geoffrey (2009). Google's Book Search: a disaster for scholars. *The Chronicle of Higher Education* [online]. August, 31 [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://chronicle.com/article/Googles-Book-Search-A/48245/>>.
- Osińska, Veslava (2008). Wizualizacja i mapowanie przestrzeni danych w bibliotekach cyfrowych. *Toruńskie Studia Bibliologiczne*, nr 1, s. 167-176.
- Ożóg, Janina ; Potempa, Aureliusz (2009). Multiwyszukiwarka baz własnych katalogu online Biblioteki Głównej Uniwersytetu Ekonomicznego w Krakowie. W: *Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci*. Materiały konferencyjne pod red. M. Kocójowej. Kraków: IINiB UJ, s. 96-105.
- Research Information Network (2009). *Overcoming barriers: access to research information content* [online]; [dostęp: 20.11.2010]. Dostępny w World Wide Web: <<http://www.rin.ac.uk/barriers-access/>>.
- Rosa, Cathy de [et al.] (2005). *Perceptions of libraries and information resources: a report to the OCLC membership* [online]; [dostęp: 27.10.2010]. Dostępny w World Wide Web: <<http://www.oclc.org/reports/2005perceptions.htm>>.
- Schmidt, Kazimierz (2011). Digitalizacja, cyfryzacja czy dyskretyzacja i jaka? W: *Cyfrowy świat dokumentu*. Pod red. Henryka Hollendra. Warszawa: CPI, s. 50-63.
- Sroka, Marek (2009). The Google Library Project, Internet Archive and Open Content Alliance: struggle for universal access to digital collections from the American perspective. W: *Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci*. Materiały konferencyjne pod red. M. Kocójowej. Kraków: IINiB UJ, s. 16-20.
- Tenopir, Carol (2009). The value gap. *Library Journal* [online] July 17 [dostęp: 5.11.2010]. Dostępny w World Wide Web: <<http://www.libraryjournal.com/article/CA6668190.html>>.
- Way, Doug (2010). The impact of Web-scale discovery on the use of a library collection. *Serials Review*, vol. 36, no. 4, pp. 214-220.
- Wisner, William (2001). Librarianship enters the twilight. *Library Journal*, no. 126, p. 68.
- Wisniewski, Jeff (2010). Web scale discovery: the future's so bright, I gotta wear shades. *Online*, vol. 34, no. 4, pp. 55-57.

MAREK NAHOTKO

Institute of Library and Information Studies
Jagiellonian University
e-mail: marek.nahotko@uj.edu.pl

FEDERATED SEARCH IN INFORMATION RESOURCES

KEYWORDS: Federated information retrieval. Internet search engines. Library catalogs. WorldCat. Google Books. Google Scholar.

ABSTRACT: The author presents activities aimed at the integration of retrieval environment of scholarly information available on Internet and in research libraries. As federated search is not sufficiently effective in global networks, two other solutions are discussed: integration by means of Google Scholar search engine and Web-scale federated search via library web interface. The paper is concluded with the list of merits and drawbacks of both methods and a proposal for their integration into one solution.

Tekst wpłynął do Redakcji 9 listopada 2010 r.