

Natalia Pamuła-Cieślak


Instytut Informacji Naukowej i Bibliologii
Uniwersytet Mikołaja Kopernika w Toruniu
e-mail: pamula@vp.pl

Wybrane wyszukiwarki zasobów naukowych w Internecie

Wprowadzenie

Zgromadzona w Internecie informacja przeraża swym ogromem. Oczywiście jest, że sieć to medium o tyle rozległe, co różnorodne, a jej użytkownik może się szybko zagubić – zarówno w ilości zgromadzonej tam informacji, jak i jej klasyfikacji, ocenie i przydatności. Na tym tle wyróżniającym się problemem stało się wyszukiwanie informacji naukowej – akademickiej, edukacyjnej, dydaktycznej – licznie przecież zgromadzonej w Internecie. Ten typ informacji charakteryzuje przede wszystkim wysoka jakość merytoryczna, językowa i bibliograficzna, ponieważ powstałe na gruncie nauki i dla niej przeznaczone zasoby tworzone są zgodnie z zasadami obowiązującymi tradycyjne publikacje i źródła naukowe. Elektroniczna informacja naukowa to czasopisma dostępne on-line, pełne teksty oraz abstrakty artykułów z nich pochodzących, bazy danych gromadzące informację edukacyjną i naukową, katalogi biblioteczne i bibliografie, roczniki statystyczne, serwisy pełnotekstowe, e-booki i e-printy, materiały konferencyjne, a także źródła nietekstowe, jak mapy i systemy nawigacyjne, rysunki techniczne, dane numeryczne i teleadresowe.

W świecie tak rozległej informacji, nie tylko naukowej i edukacyjnej, najpopularniejszym dotychczas narzędziem wyszukiwawczym, o wciąż niesłabnącej popularności, są komercyjne wyszukiwarki internetowe (ang. *search engines*). Nie są one jednak w stanie udzielić swym użytkowni-



kom informacji relewantnej. Dają raczej ogólne spojrzenie na poszukiwane zagadnienie, nakreślając bardzo rozległe terytorium do dalszych poszukiwań¹. Wyszukiwarki koncentrują się na ilości zaindeksowanych stron internetowych. Nie mają jednak żadnych mechanizmów kontroli i oceny treści takich zasobów. Ich wadą jest również brak aktualizacji danych. Nierzadko podawana w rankingu odpowiedzi informacja jest zdezaktualizowana z powodu zmian, które zaszły na stronie lub w jej adresie. Wyszukiwarki nie mają także możliwości indeksowania płatnych zasobów dostępnych on-line, są one bowiem fragmentem Ukrytego Internetu (ang. *Invisible Web*)². Dla nieefektywnych wyszukiwarek komercyjnych istnieje jednak kilka alternatywnych sposobów i narzędzi wyszukiwawczych. Niewątpliwie ważną grupą narzędzi ułatwiających dostęp do informacji naukowej on-line są grupy bądź listy odnośników, czyli różnego rodzaju katalogi specjalne, które zestawiają zasoby informacyjne według tematów lub rodzajów informacji³. Często użytkownik odnajduje je na witrynach bibliotek oraz innych instytucji – uczelni wyższych i stowarzyszeń o charakterze naukowym. Pochodnymi tego typu inicjatyw są kolekcje odnośników w postaci serwisów tematycznych (ang. *subject gateways*)⁴ oraz bibliotek wirtualnych.

Jedną z alternatyw dla wyszukiwania informacji o charakterze naukowym są narzędzia dedykowane takim właśnie zasobom, zwane wyszukiwarkami naukowymi (ang. *science search engines*). Niniejszy artykuł prezentuje działanie kilku z nich.

Scirus for scientific information only – <http://www.scirus.com>

Zgodnie z pełną nazwą tego narzędzia, Scirus to wyszukiwarka zasobów naukowych w Internecie. Prawa autorskie do tego narzędzia posiada Elsevier B.V. – światowy wydawca naukowy, mechanizmy wyszukiwawcze zaś dostarcza norweska firma Fast Search and Transfer, której

¹ J. Mason i in., *INFOMINE: Promising Directions in Virtual Library Development*. First Monday [on-line] 2000, vol. 5, no. 6 [dostęp 30 września 2008]. Dostępny w World Wide Web: http://firstmonday.org/issues/issue5_6/mason/index.html.

² Ch. Sherman, G. Price, *The Invisible Web: Uncovering Information Search Engines Can't See*, Medford–New Jersey 2003, s. 73.

³ L. Derfert-Wolf, *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways*. Biuletyn EBIB [on-line] 2004, nr 6 (57) [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2004/57/derfert.php>.

⁴ Tamże.

partnerami technologicznymi są koncerny: AT&T, CareerBuilder, Dell, Freeserve, IBM i Reuters⁵. Wyszukiwarka Scirus indeksuje naukowe, edukacyjne, techniczne i medyczne informacje zgromadzone w World Wide Web. Znajduje najnowsze raporty, recenzowane artykuły, patenty, preprinty oraz czasopisma niewyszukiwane przez inne wyszukiwarki. Scirus oferuje unikalne funkcje zaprojektowane specjalnie dla naukowców oraz osób poszukujących informacji naukowych i zawiera filtry, które pomagają odrzucać strony ich niezawierające⁶. Przeszukuje witryny WWW, serwery preprintowe, archiwa tekstów elektronicznych, repozytoria, a także bazy patentów i czasopism. Wyszukiwarka Scirus koncentruje się na zasobach naukowych, dlatego rejestruje:

- 83 mln stron w domenie .edu,
- 25 mln stron w domenie .org,
- 10 mln stron w domenie .ac.uk,
- 22 mln stron w domenie .com,
- 6,5 mln stron w domenie .gov,
- 68 mln stron z kręgu STM⁷ oraz uniwersytetów i innych uczelni wyższych z całego świata⁸.

Ponadto Scirus indeksuje zasoby takich autorytatywnych⁹ serwisów, jak: ArXiv.org, BioMed Central, Caltech Coda, Cogprints, Crystallography Journals Online, CURATOR, Digital Archives, DiVa, Project Euclid, HKUST Institutional Repository, Institute of Physics Publishing, LexisNexis, Medline, MIT Open Course Wave, NASA, NDLTD, Organic Eprints, PsyDok, PubMed Central, RePEc, ScienceDirect, Scitation, SIAM, T-SPACE¹⁰.

Wyszukiwanie za pomocą Scirusa nie jest skomplikowane, choć oferuje on kilka opcji niedostępnych w tradycyjnych wyszukiwarkach internetowych. Okno do wyszukiwania prostego, dostępne na stronie głównej wyszukiwarki, umożliwi sformułowanie prostego zapytania lub też frazy ujętej w cudzysłów. Formularz wyszukiwania zaawansowanego po-

⁵ *White Paper: How Scirus Works* [on-line]. Scirus – for scientific information [dostęp 30 września 2008]. Dostępny w World Wide Web: http://scirus.com/press/pdf/WhitePaper_Scirus.pdf.

⁶ Tamże.

⁷ STM (ang. *Science, Technology and Medicine*) – określenie stron naukowych z kręgu tematycznie dotyczącego nauk ścisłych, technicznych i medycyny.

⁸ Stan na dzień 30 czerwca 2008 r.

⁹ Autorytatywny, czyli zawierający odpowiednio duży zasób wartościowych informacji na poszukiwany temat.

¹⁰ *Scirus Help* [on-line]. Scirus – for scientific information [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://scirus.com/html/help/index.html>.

zwala z kolei na zawężenie obszaru wyszukiwania (np. do tytułu artykułu, tytułu czasopisma, autora, numeru ISSN) oraz na stosowanie odpowiednich operatorów logicznych. Dzięki temu możliwe staje się wyszukiwanie według dat wydania materiałów, typów zasobów (np. artykuły, patenty), formatów zapisu danych (np. pdf) czy konkretnych serwisów (czasopiśmienniczych, jak np. „Crystallography Journals Online”, lub internetowych, jak np. NASA). Problemu nie stanowi także ograniczenie zakresu wyszukiwania do określonej dziedziny wiedzy.

Ranking odpowiedzi generowany przez Scirusa pod względem wizualnym nie różni się znacząco od innych tego typu list, choć zawiera kilka swoistych elementów i oryginalnych funkcji. Na górze listy wyświetlona zostaje statystyka wyszukiwania, dotycząca ogółu uzyskanych odpowiedzi i zawierająca informacje, z jakich źródeł one pochodzą. Ranking sortowany jest według relewantności. Istnieje jednak możliwość uporządkowania rezultatów według dat publikacji materiałów. Opis wyszukanego dokumentu składa się z tytułu (lub jego namiastki, kiedy dokument tytułu nie ma) będącego łączem hipertekstowym do zasobu. Jeśli dokument zaopatrzone jest w metadane, wyświetlana jest informacja o jego autorze oraz pozostałe informacje bibliograficzne. Każdy wyszukany dokument zaopatrzone jest w adnotację treściową. W sytuacji, gdy artykuł pochodzi z zasobów partnerów współpracujących ze Scirusem, u dołu opisu pojawia się stosowne logo lub informacja o tym. Jeśli dokument nie jest dostępny w formacie html, obok opisu widnieje informacja, w jakim formacie został zapisany. Istnieje możliwość zachowania (zapisania) wybranych opisów na dysku bądź w katalogu wskazanym przez użytkownika, przesłania rezultatów pocztą elektroniczną lub ich eksportu do dowolnego programu. Z lewej strony listy odpowiedzi umieszczone jest menu pozwalające na wybór dodatkowych opcji, tj.:

- wyświetlenie informacji, ile znalezionych dokumentów pochodzi z zasobów czasopiśmienniczych, ile z preferowanych źródeł internetowych, a ile z pozostałej części sieci WWW,
- ograniczenie wyników wyszukiwania do jednej z wyżej wymienionych kategorii,
- zawężenie wyników wyszukiwania do jednego ze wskazanych formatów zapisu danych,
- doprecyzowanie terminu wyszukiwawczego (poprzez odpowiedni kontekst),
- redefiniowanie – poszerzanie i zawężanie wyników wyszukiwania.



Scirus zaoferował bibliotekom oraz ich użytkownikom nową funkcjonalność swojego serwisu. Biblioteki, decydując się na współpracę z tą wyszukiwarką, mogą ułatwić swoim użytkownikom dotarcie do pełnych tekstów naukowych dostępnych w płatnych serwisach przez nie subskrybowanych. Służy do tego opcja dostępna w preferencjach (na stronie głównej serwisu) – Library Partner Links. Z rozwijanej listy można dokonać wyboru biblioteki, do której zasobów elektronicznych użytkownik chciałby mieć dostęp. Wówczas na liście odpowiedzi pojawią się narzędzia pozwalające na bezpośredni dostęp do pełnych tekstów wyszukanych dokumentów.

Poza omówionymi funkcjami Scirus proponuje także dostęp do najnowszych doniesień naukowych pojawiających się w serwisie New Scientist (opcja ta dostępna jest ze strony głównej wyszukiwarki pod nazwą Latest Scientific News). W ramach Scirusa uruchomiony został dodatkowo serwis internetowy – Scirus Topic Pages. Jest to portal społecznościowy typu wiki, w którym eksperci umieszczają autorytatywne omówienia bieżących problemów badawczych z różnych dziedzin wraz z odsyłaczami do tematycznie powiązanych z nimi artykułów naukowych i stron internetowych¹¹. Serwis można przeglądać według ostatnio dodanych streszczeń oraz za pomocą list alfabetycznych autorów i tematów. Obecnie funkcjonuje on jeszcze w wersji testowej.

Google Scholar – <http://scholar.google.com>

Google Scholar jest jednym z narzędzi wyszukiwawczych udostępnionych przez potentata na rynku wyszukiwania w Internecie – firmę Google. Ma to niemałe znaczenie dla użytkowników, którzy w dużej mierze korzystają z głównego narzędzia wyszukiwawczego stworzonego przez tę firmę – komercyjnej wyszukiwarki Google. O jej popularności decyduje efektywność wyszukiwania, związana z unikalnym mechanizmem rankingowym Page Rank, oraz prostota interfejsu wyszukiwawczego¹². Ponieważ wyszukiwarka Google nie była dostosowana do poszukiwań informacji naukowej w Internecie, w listopadzie 2004 r. udostępnione zosta-

¹¹ *Scirus topic pages* [beta] [on-line]. Scirus – for scientific information [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://topics.scirus.com>.

¹² R. Hock, *The Extreme Searcher's Guide to Web Search Engines*, Medford–New Jersey 2001, s. 111.

ło narzędzie Google Scholar¹³ umożliwiające realizację zapytań o charakterze naukowym. Do dziś (wrzesień 2008 r.) wyszukiwarka ta dostępna jest w wersji testowej beta.

Mottem wyszukiwarki Google Scholar jest cytata z Isaaca Newtona: „Stań na ramionach olbrzymów”. Przesłanie to sugeruje, że Google Scholar korzysta z doświadczeń i mechanizmów wyszukiwawczych potężnego narzędzia, jakim jest wyszukiwarka Google, w takim samym stopniu, w jakim wielki naukowiec wspierał się na doświadczeniach swoich poprzedników. Na przynależność do rodziny produktów Google wskazuje także interfejs wyszukiwawczy, który jest równie prosty i przejrzysty jak okno wyszukiwarki komercyjnej.

Google Scholar jasno precyzuje swoje cele: jest wyszukiwarką internetową zasobów naukowych, która z jednego miejsca w sieci umożliwia wyszukiwanie wielu źródeł, recenzowanych czasopism naukowych, rozpraw, książek, abstraktów i artykułów pozyskanych z wydawnictw akademickich, stowarzyszeń naukowych, repozytoriów, zbiorów preprintów, uniwersytetów i innych instytucji edukacyjnych¹⁴. Chcąc być jedną z najlepszych wyszukiwarek naukowych w Internecie, podjęła inicjatywę współpracy z wydawcami tekstów naukowych, którym proponuje zwiększenie „widzialności” ich publikacji w sieci poprzez umieszczenie w bazie wyszukiwarki pełnych tekstów posiadanych przez nich publikacji. Zachowują oni pełne prawa i kontrolę zarówno nad tymi zasobami, jak i sposobem udostępniania ich użytkownikom. Wymagane jest jedynie, aby wydawca, zgłaszając tekst, udostępnił wszystkim użytkownikom Google Scholar abstrakt naukowy, tak by mogli się oni zorientować, czy tekst odpowiada ich potrzebom informacyjnym. W ten sposób Google Scholar chce przybliżyć użytkownikom dużą część Ukrytego Internetu. Niestety, brak informacji o tym, jacy wydawcy podjęli współpracę z wyszukiwarką.

Google Scholar wychodzi naprzeciw także użytkownikom bibliotek z całego świata, proponując współpracę również i tym instytucjom naukowym. Jest to pierwsza wyszukiwarka naukowa, która zaproponowała tego typu możliwość. Korzystanie z opcji pozwalającej na dostęp do zasobów bibliotecznych odbywa się podobnie jak w Scirusie. Należy wybrać

¹³ M. Weryho, *Stań na ramionach gigantów, czyli Google Scholar*. Biuletyn EBIB [on-line] 2005, nr 2 (63) [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2005/63/weryho.php>.

¹⁴ *About Google Scholar* [on-line]. Google Scholar [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://scholar.google.pl/intl/en/scholar/about.html>.

preferencje wyszukiwawcze, a następnie w polu „Linki biblioteczne” wpisać nazwę żądanej biblioteki. Niestety, Google nie udostępniła listy instytucji współpracujących i dlatego wprowadzanie nazw bibliotek jest często działaniem w ciemno.

Wyszukiwanie w Google Scholar nie odbiega od wyszukiwania w Google. Różnią się natomiast rezultaty wyszukiwania i ich prezentacja. Wyraźnie wyodrębniają się dwie formy dostępu do znalezionych zasobów:

- on-line – tytuł dokumentu ma postać hiperłącza prowadzącego do pełnego tekstu elektronicznego lub jego abstraktu,
- off-line – tytuł dokumentu nie przybiera postaci hiperłącza, lecz formę tradycyjnego, niewyróżnionego graficznie fragmentu tytułu; w większości przypadków w ten sposób prezentowane są cytaty dokumentów niedostępnych on-line, które zostały pozyskane metodą analizy cytowań na podstawie przypisów i bibliografii załącznikowych pozostałych zasobów zaindeksowanych w Google Scholar. Użytkownik, korzystając z opcji „Cytowane przez”, ma możliwość dotarcia do wszystkich publikacji cytujących wyszukany przez niego dokument. Poszerza to zakres prowadzonych poszukiwań.

Jeśli wyświetlony na liście rankingowej dokument został opublikowany w kilku wersjach i miejscach w Internecie, prezentowana jest wersja optymalna, a u dołu opisu dokumentu pojawiają się odsyłacze do pozostałych wariantów tekstu – preprintów, abstraktów, materiałów konferencyjnych lub innych opracowań. Dla każdego wyszukanego dokumentu można przeprowadzić dodatkowe wyszukiwanie w obrębie całego Internetu. Opis dokumentów wyszukanych przez Google Scholar jest opisem w miarę pełnym. W przypadku ustalenia autorstwa tekstu informacje takie zostają czytelnie podane. Nierzadko pojawia się także tytuł, rocznik i numer czasopisma, z którego pochodzi dany artykuł, a także informacja o wydawcy (zarówno przy pełnych tekstach, jak i przy dokumentach o ograniczonym dostępie).

Poza swoistymi elementami opisu wyszukanych zasobów Google Scholar ma także te same funkcje, jakie znane są z komercyjnej wyszukiwarki Google, tj.:

- własne wersje wyszukiwanych dokumentów umożliwiające korzystanie z kopii przechowywanych na serwerze wyszukiwarki, w przypadkach, kiedy dokument oryginalny jest chwilowo niedostępny,
- możliwość oglądania dokumentów zapisanych w niespecyficznych formatach (np. pdf, ppt) jako dokumentów html.



Wyszukiwarka Google Scholar pozwala również na prowadzenie wyszukiwania zaawansowanego (pełnotekstowo w dowolnym miejscu artykułu lub jego tytule, według autora artykułu, czasopisma, w którym został zamieszczony, i daty publikacji). Google Scholar jest wyszukiwarką naukową, metodą indeksowania, wyszukiwania i wyświetlania wyników najbardziej zbliżoną do tradycyjnych wyszukiwarek internetowych. Stanowi to niewątpliwą zaletę, gdyż zachęca do korzystania z tego narzędzia stałych klientów komercyjnej wyszukiwarki Google. Ponadto niezwykle cenna wydaje się inicjatywa współpracy z wydawcami i bibliotekami. Zwłaszcza ta ostatnia możliwość wskazuje, jak wiele bibliotekarze mogą zrobić dla propagowania i lepszego udostępniania literatury naukowej w Internecie. Trwają prace nad udostępnieniem zasobów zaindeksowanych w Google Scholar w głównym interfejsie wyszukiwawczym komercyjnej wyszukiwarki Google.

Google Book Search – <http://books.google.com>

◯ oprócz wyszukiwarki Google Scholar istnieje kilka innych narzędzi Google pozwalających na wyszukiwanie specjalistycznej informacji naukowej. Jednym z nich jest Google Book Search. Jest to z jednej strony rezultat inicjatywy digitalizacyjnej firmy Google i kilku naukowych bibliotek amerykańskich i brytyjskich, z drugiej – wynik współpracy wydawców naukowych i wielu bibliotek na świecie. Wyszukiwarka dysponuje interfejsem wyszukiwawczym Google, który charakteryzuje się – jak we wszystkich narzędziach tego koncernu – dużą prostotą, intuicyjnością obsługi oraz ujednoliceniem opcji. Jest on dostępny także w polskiej wersji językowej.

Jak wskazuje nazwa, Google Book Search pozwala na wyszukiwanie książek. Mogą one przybierać tu postać zeskanowanych, lecz w pełni przeszukiwanych, kopii cyfrowych zapisanych i udostępnianych w formatach pdf i html lub mogą być prezentowane we fragmentach (gdy prawa autorskie i wydawnicze nie wygasły). Książki udostępnione w ograniczonym podglądzie można jednak kupić (Google prezentuje wyniki wyszukiwania z księgarni internetowych) czy zlokalizować w konkretnej bibliotece poprzez współpracujące z serwisem katalogi centralne poszczególnych krajów¹⁵ albo światową bazę WorldCat. Takie rozwiązanie możliwe jest dzie-

¹⁵ W Polsce katalogiem współpracującym z Google Book Search jest NUKAT.

ki metadanych pobieranym z katalogów bibliotecznych oraz systemowi lokalizacji położenia geograficznego użytkownika korzystającego z wyszukiwarki.

Formularz zapytań zaawansowanych w Google Book Search pozwala na wybór trybu dostępu do książki (pełne teksty, widok ograniczony, katalogi biblioteczne) oraz zawężenie kryteriów wyszukiwawczych do języka dokumentu, tytułu, autora, wydawcy, daty publikacji oraz numeru ISBN.

Nadrzędnym celem twórców wyszukiwarki książek Google jest współpraca z możliwie największą liczbą wydawców i bibliotek. Ma to umożliwić utworzenie kompletnego, łatwego w przeszukiwaniu, wirtualnego i wielojęzycznego katalogu bibliotecznego¹⁶. Z pewnością takie narzędzie wyszukiwawcze jest potrzebne do uporządkowania zasobów książkowych w Internecie, bo bez wątplenia stanowi może cenne źródło pozyskiwania informacji o charakterze naukowym.

Podsumowanie

Omówione powyżej wyszukiwarki naukowe są jedynie wybranymi przykładami powstających od niedawna tego typu narzędzi. Ich istnienie jest dowodem na to, że wyszukiwanie informacji naukowej w sieci stało się prężnie rozwijającym się trendem. Wydaje się, że tendencja ta podyktowana jest ogromnymi potrzebami użytkowników, którzy dostrzegli, że globalne wyszukiwarki internetowe nie zapewniają dostępu do zasobów naukowych. Warto zauważyć, że rynek narzędzi do wyszukiwania naukowego podlega także monopolizacji i nie przynosi wymiernych korzyści. Pod koniec maja 2008 r. swoje istnienie zakończyły dwie wyszukiwarki (naukowa i książek) tworzone przez Microsoft – Live Search Academic oraz Live Search Books. Choć koncern zapewnia, że wyszukiwanie naukowe zostało włączone do interfejsu głównego innego produktu Microsoftu – Live Search¹⁷, niestety nie jest to zauważalne w rezultatach wyszukiwania, ponieważ nadal dominują tu informacje z ogólnodostępnej, komercyjnej części sieci.

¹⁶ Program biblioteczny Google Books Search Library Project [on-line]. Google Book Search [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://books.google.com/googlebooks/library.html>.

¹⁷ S. Nadella, *Book search winding down*. W: *Live Search. The official blog of the Life Search Team at Microsoft* [on-line]. May 23, 2008, 7:45 AM [dostęp 30 września 2008]. Dostępny w World Wide Web: <http://blogs.msdn.com/livesearch/archive/2008/05/23/book-search-winding-down.aspx>.

Z całą pewnością wyszukiwarki naukowe ułatwiają dostęp do wysokojakościowych źródeł i opracowań, choćby poprzez wyposażenie interfejsu wyszukiwawczego w szereg funkcji, których pozbawione są wyszukiwarki tradycyjne. Istotne staje się więc propagowanie tych specjalistycznych narzędzi w środowiskach akademickich, gdyż służą one pomocą w wyszukiwaniu odpowiednich informacji oraz szybkim i prostym ich uwiarygodnianiu.



Selected academic search engines on the Web

Abstract

The article reviews selected scientific and scholar search engines which are available in the Internet: Scirus, Google Scholar and Google Book Search. It pays attention to their functions, usage and utility for academic environment which is connected with searching for information in the various fields of science.

