

Osińska, Veslava / Osiński, Grzegorz

Efektywne wyszukiwanie informacji w sieci WWW

Acta Universitatis Nicolai Copernici. Bibliologia 4 (340), 385-403

2000

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.

Veslava Osińska, Grzegorz Osiński

Efektywne wyszukiwanie informacji w sieci WWW

Wstęp

W celu odnalezienia pożądaných informacji w Internecie użytkownik ma do wyboru dwie usługi: serwis wyszukiwawczy oraz katalog tematyczny. Pierwsza, popularnie zwana „szperaniem”, daje potencjalną możliwość odnalezienia wszystkich zasobów sieciowych o podanych przez użytkownika warunkach. Natomiast katalogi tematyczne pozwalają przeglądać odpowiednie działy, posegregowane już przez administratorów i redaktorów serwisów, mają one jednak istotną wadę – aby odszukać w gąszczu kategorii szczegółową informację, potrzeba dużo czasu, a wyniki nie zawsze są zadowalające. Do najpopularniejszych w Polsce katalogów można zaliczyć serwisy: Polishworld, Wirtualna Polska, WOW, Polska¹. Odnośniki do najważniejszych polskich katalogów tematycznych umieszczone są m.in. na stronie głównej Uniwersytetu Mikołaja Kopernika w Toruniu². Ze względu na małą użyteczność wykorzystywania posegregowanej już informacji dla celów naukowych i poznawczych, które stawiamy sobie przy omawianiu narzędzi sieciowych, nie będziemy szczegółowo omawiali struktur budowy katalogów tematycznych, skupimy się natomiast na serwisach wyszukiwawczych. Dają nam one możliwość nie tylko konstruowania odpowiednich zapytań, ale również tworzenia własnych katalogów tematycznych. W ten naturalny sposób możemy wyszukiwać informacje oraz tworzyć katalogi tematyczne w celu ich późniejszego wykorzystania lub udostępnienia innym użytkownikom. Aby jednak móc sprostać takim zadaniom,

¹ Adresy wymienionych serwisów: www.polishworld.com, wp.cnt.pl, www.wow.pl, www.polska.pl

² <http://www.uni.torun.pl>

konieczna jest dobra znajomość działających w sieci serwisów informacyjnych oraz możliwość twórczego ich wykorzystania dla własnych celów.

Prawie każdy, kto szuka ważnych informacji, nie wiedząc, jak do nich dotrzeć, zaczyna od uruchomienia przeszukiwarki i postawienia zapytania. Serwisy informacyjne, popularnie nazywane wyszukiwarkami, służą nie tylko początkującym użytkownikom. Wszystko jedno czy szuka się adresu e-mailowego, artykułu z grupy tematycznej (*newsgroup*), programu shareware'owego³ czy też specjalnej oferty jakiejś firmy: bez wsparcia w postaci odpowiednich narzędzi wyszukiwawczych jest się na straconej pozycji.

Wyszukiwarki można określić jako ogromne bazy danych, które na podstawie słów kluczowych wyszukują strony WWW, które je zawierają. Zasada działania takich serwisów wyszukiwawczych polega na codziennym skanowaniu zasobów internetowych, w trakcie którego zachodzi automatyczne indeksowanie w specjalnych bazach danych znalezionych stron, rejestrowanie zauważonych zmian oraz bezpłatne udostępnianie użytkownikom stworzonych katalogów słów kluczowych⁴. Gdy wpisujemy w odpowiednie pole wyszukiwarki słowo identyfikujące interesujący nas temat, w wyniku wyszukiwania otrzymamy listę adresów sieciowych (URL)⁵ odnoszących się do znalezionych dokumentów. Największe, ogólnościatowe serwisy wyszukiwawcze (np. Altavista) dysponują bazami liczącymi ponad 85 milionów stron WWW. Narzędzia te pozwalają na zadawanie pytań, następnie przetwarzają wprowadzone zapytania i prezentują w odpowiedzi tytuły wyszukanych stron, zwykle z krótkim opisem ich zawartości, oraz odsyłaczami hipertekstowymi.

Poza aktualnością i kompletnością bazy danych o jakości wyszukiwarki decyduje przede wszystkim elastyczność systemu formułowania pytań. Dzięki oferowanej przez wszystkie przeszukiwarki możliwości używania logicznych operatorów AND (spójnik „i”), OR (spójnik „lub”) oraz NOT (spójnik „nie”) można sprecyzować pytanie i dzięki temu zredukować listę odpowiedzi na dany temat. Mimo wspomagania ze strony przeszukiwarek i tematycznie uporządkowanych katalogów „polowanie” na informacje jest zajęciem dość żmudnym.

Kto żąda precyzyjnych wyników wyszukiwania, musi oczywiście opanować specyficzne mechanizmy zadawania pytań. Problem zazwyczaj pojawia się, kiedy trzeba użyć nie jednego, a kilka słów kluczowych. Wtedy należy pamiętać o kilku zasadach:

– Zastosowanie małych liter w poszukiwanym wyrażeniu wygeneruje odpowiedź, gdzie podane słowa pisane są literami o dowolnej wielkości. Natomiast

³ Shareware – rodzaj dystrybucji oprogramowania, którego można używać za darmo w okresie próbnym.

⁴ W. Duch, *Fascynujący świat programów komputerowych*, Poznań 1997, s. 351–380.

⁵ Uniform Resource Locator – system adresowania używany w sieci WWW. Składa się z protokołu, np. http, dwukropka, dwóch znaków „/”, następnie adresu serwera – DNS i ścieżki do konkretnego zasobu na serwerze.

użycie dużych liter spowoduje znalezienie tych dokumentów, gdzie wyrazy są pisane dużymi literami. Na przykład: wpisując *Kowalski*, szukamy ciągu znaków wyrażenia rozpoczynającego się koniecznie z dużej litery: *KOWALSKI* lub *Ko-wAlski* itp.

– Jeśli szukamy zdań składających się z kilku wyrazów, to lepiej je umieścić w cudzysłowach „”.

– Aby kazać szperaczowi szukać kontekstowo, należy wpisać możliwie proste zdanie bez cudzysłowów. To spowoduje odnalezienie maksymalnej liczby stron zawierających jakiegokolwiek z wpisanych słów.

– Słowa, które mają wystąpić na stronie, poprzedzamy plusem (+), odwrotnie wyrazy, których ma nie być – minusem (-). Symbol gwiazdki „*” – dowolny ciąg znaków – umieszczony za przedrostkiem (minimum trzyliterowym) spowoduje wyszukanie dokumentów zawierających jedno ze słów, które można utworzyć na bazie przedrostka.

2. Wyszukiwanie nie tylko dla zaawansowanych

Znaki używane w zapytaniach

Dla połączenia kilku słów kluczowych nie zawsze trzeba odwoływać się do opcji wyszukiwania zaawansowanego. Można spróbować metody polegającej na używaniu odpowiednich symboli bądź znaków.

NAZWA	SYMBOL	OPIS	SZPERACZE
CUDZYŚLÓW	„ ”	Tego symbolu używa się do oznaczenia ścisłego zdania lub frazy.	Wszystkie
Operator PLUS	+	Operator plus umieszcza się przed słowem lub zdaniem, które powinno wystąpić na stronie z odpowiedzią.	Wszystkie
Operator MINUS	-	Operator ten poprzedza wyrazy, które mają być pominięte na wyszukanych stronach.	Wszystkie
Operator mnożenia	*	Gwiazdki używa się do odnalezienia dokumentów, zawierających słowa, zaczynających się lub kończących się jednym ciągiem znaków.	Altavista, Lycos, Hotbot
Przecinek	,	Przecinka używa się do rozdzielania nazw własnych w wyrażeniu.	Infoseek
Filtr		Filtr pomaga zawęzić wyniki wyszukiwania; stosuje się go, aby sprecyzować hasło wyszukiwania.	Infoseek

Poniżej podajemy kilka przykładów formułowanych zapytań stworzonych z wykorzystaniem przedstawionych powyżej znaków. Szukając nazwy własnej bądź

ściśle brzmiącego wyrażenia, aby szybko pozbyć się stron z losowym trafieniem słów składowych wyrażenia, należy ująć je w cudzysłów: „Ogniem i mieczem”, „Głos Uczelni”, „Komitet Badań Naukowych”. Wyszukiwarka Infoseek oferuje możliwość znalezienia na jednej stronie kilku niezależnych nazw własnych, nie umieszczając ich w cudzysłowach, a rozdzielając przecinkiem: specyfikacja *Kubuś Puchatek, Prosiaczek, Tygrysek* znajdzie strony z bohaterami bajki o Kubusiu Puchatku. Aby zdobyć informację medyczną na temat np. chorób wewnętrznych z wyjątkiem przypadków kardiologicznych, wystarczy użyć odpowiednich znaków: *+choroby +wewnętrzne –kardiologia*. Należy zwrócić szczególną uwagę na wstawioną spację przed znakami plus i minus zaczynając od drugiego wyrazu. Ponieważ symbol gwiazdki oznacza dowolny ciąg znaków, możemy nim zastąpić słowa bądź litery podając jako przedrostek minimum 3 znaki. Wspecyfikowanie *regu** spowoduje szukanie stron zawierających takie wyrazy jak *regulamin, regulaty, regulacja, regulowanie* itp.

Operatory logiczne (BOOLE'A)

Większość wyszukiwarek toleruje wpisanie operatorów logicznych w wyrażeniu wyszukiwawczym dopiero w trybie zapytań złożonych (Advanced). Dozwolone jest także użycie symboli zastępczych operatorów⁶.

OPERATOR	SYMBOL	OPIS	SZPERACZE
1	2	3	4
AND (spójnik „i”)	&	AND łączy dwa lub więcej wyrazów w celu zawężania obszaru poszukiwania. Zostaną wymienione tylko strony, zawierające wszystkie wyrazy zawarte w wyrażeniu.	Wszystkie
OR (spójnik „lub”)		OR łączy dwa wyrazy i wybiera wszystkie dokumenty, zawierające jeden z nich wybiórczo.	Wszystkie
NOT ANDNOT	!	NOT będzie szukał listy z występującymi wyrazami, które zostały wpisane przed NOT, a wykluczającymi występowanie wyrazów, wpisanych po NOT.	Altavista, Infoseek, Lycos, Excite, Netoskop Hotbot
Segregator	()	Nawiasy są pomocne dla konstruowania bardziej złożonych zapytań, kiedy trzeba segregować relacje między słowami kluczowymi.	Altavista, Netoskop, Wirtualna Polska, Hotbot, WebCrawler ⁷
NEAR	~	NEAR wskazuje na bliską odległość (czasem z zadaniem parametrem wyrażanym ilością słów) pomiędzy szukanymi wyrazami.	Altavista, Lycos Webcrawler, Netoskop

⁶ Brad Hill, *World Wide Web Searching for Dummies*, San Diego 1998, s. 15–44.

⁷ <http://www.webcrawler.com/Help/html>

1	2	3	4
NEXTTO		Umożliwienie znajdowania dokumentów z wyrazami występującymi w zadanej kolejności.	Netoskop, Lycos
@	@	Wyszukanie stron z zadaniem wyrażeniem na wyspecyfikowanym serwisie.	Netoskop

Jeśli chcemy przejrzeć w Internecie strony o roślinach i zwierzętach, można wpisać *flora AND fauna*. Operatora NOT używamy zawsze razem z operatorem AND: wpisujemy *elektrownie AND NOT atomowe*, po to, ażeby odszukać strony o wszystkich elektrowniach poza atomowymi. Wojna w Jugosławii trwa z różnym natężeniem od wielu lat, ażeby uzyskać informacje o ostatnim konflikcie w Kosowie, powinno się umiejętnie odseparować opisy poprzednich wydarzeń w tym regionie: (*wojna OR nality*) *AND NATO AND (Yugoslavia OR Serbia) AND NOT Sarajevo*.

Aby zapoznać się ze stronami opisującymi działania Ministerstwa Edukacji Narodowej w związku z wprowadzaną reformą szkolnictwa, wpisujemy *reforma szkolnictwa NEAR Ministerstwo*.

Pola meta

Pola meta – inaczej skróty – to mechanizmy ułatwiające poszukiwaczom wynajdowanie słów kluczowych w tekście źródłowym⁸ strony WWW. Mimo że z tekstu pomocy niektórych szperaczy wynika, że pola meta stosowane są jako opcja wyszukiwania prostego, jednak niewiele osób posługuje się tego typu narzędziem, dlatego postrakujemy go jako zaawansowane. Metoda ta pozwala na poszukiwanie tytułów stron, URL-i, linków oraz jakichkolwiek innych dodatkowych informacji, które są określone za pomocą HTML-owych pól meta. Nazwa pola musi być napisana małymi literami, a od zapytania musi być oddzielona tylko dwukropkiem (bez odstępów).

Dla wszystkich przeglądarek pola meta pełnią podobne funkcje. Dla przykładu, szukając wyszczególnionego tytułu strony (w kodzie HTML tekst zawarty pomiędzy słowami <TITLE> i </TITLE>) w Altavista, Infoseek oraz Hotbot używamy pola *title:tytuł*, w Yahoo – *t:tytuł*, określonej domeny⁹ w adresie *url:domena* lub *domain:domena* bądź *u:domena* odpowiednio.

Coraz częściej spotyka się rozwiązanie szybkiej obsługi pól meta polegające na umieszczeniu na stronie WWW mnóstwa okienek wyboru (rys.). Poprzez takie graficzne przedstawienie środowiska praca dla przeciętnego użytkownika sprowadza się do prostego wyboru poszczególnych opcji. Wystarczy wpisać tylko tekst w odpowiednich polach, a serwis połączy je w logiczne zdanie.

⁸ Źródłem WWW jest dokument napisany w języku HTML (Hyper Text Markup Language).

⁹ Domena lub DNS (Domain Name Serwer) jest adresem znakowym grupy komputerów podłączonych do jednej sieci lokalnej.

Return to *Fewer Options*

1 **Look For**
Search for pages containing the search term(s) in the form specified.

2 **Language**
Limits results to a specific language.

3 **Word Filter**
Limit results to pages containing/not containing the words specified.

4 **Date**
Limit results to pages published within a specified period of time.

5 **Pages Must Include**
Return only pages containing the specified media types or technologies.

6 **Location/Domain**
Return only pages in specific domains (wired.com, doj.gov), top-level domains (.edu, .com), and/or specific continents or countries.
(See the [domain and country code index](#) for more detailed info.)

7 **Word Stemming**
Searches for grammatic variations of your search term. Ex: Searches for "thought" will also find "think" and "thinking".

8 **Return Results**
Choose the number of results to be displayed and the length of each description.

all the words

any language

must contain the words

must not contain the words

more terms (+)

anytime

After or on

January 1, 1998

image audio MP3 video
 Shockwave Java JavaScript ActiveX
 VRML Acrobat VB Script Win Media
 extension: (gif, .bd...)

Continent

anywhere

Domain

(.com, .edu) website: (wired.com, etc.)
country code: (.uk, .fr, .jp)

Any Page
 Top Page
 Personal Page
 Page Depth

Enable Word Stemming

10

full descriptions

SEARCH clear settings

Ostatnio obserwuje się tendencję do konstruowania „inteligentnych” przeszukiwarek, czyli poszerzenia stopnia kojarzenia kontekstu zapytań. Dla angielskojęzycznych internautów pomocna może być funkcja „prostych zapytań w jęz. angielskim” typu „how many birds are there in the world?” lub „best pizza in San Francisco” i otrzymanie sensownej odpowiedzi. Tu należy używać poprawnej gramatyki oraz ortografii.

Altavista www.altavista.digital.com www.altavista.pl

Popularna i wydajna wyszukiwarka firmy Digital oprócz stron WWW umożliwia również przeszukiwanie grup dyskusyjnych – w tym celu należy ustawić parametr The Web opcji Search na Usenet.

– Opcja Simple – wyszukiwanie proste umożliwia stosowanie w wyrażeniu pytającym metody rozróżniania dużych/małych znaków, cudzysłowów, plusa, minusa, symbolu gwiazdki.

– Opcja Advanced¹⁰ – wyszukiwanie zaawansowane pozwala na użycie pomiędzy słowami kluczowymi operatorów AND, OR, NOT, NEAR i nawiasów w celu układania wyrażen złożonych. Funkcja NEAR szuka w odległości nie większej niż 10 słów.

– Opcja Display the results – pozwala na wybór formatów stron wynikowych. Standard Form – lista stron zawiera tytuły i krótkie opisy odnalezionych stron, datę wprowadzenia danego dokumentu do bazy oraz URL-e. Compact Form – dostarcza zwięźlejszego opisu, natomiast Detailed Form – bardziej szczegółowego.

– Kolejność wyświetlania odnalezionych dokumentów odpowiada zgodności ich zawartości z podanym wyrażeniem. Pierwsze pozycje na tej liście zajmują więc strony najlepiej opisujące zagadnienia.

– Jak wspomniano wyżej, słowa meta można dodawać do wyrażenia szukanego (używając operatorów) w celu sprecyzowania pożądaných informacji i zawężenia listy odpowiedzi.

Słowo kluczowe	Funkcja
1	2
anchor:tekst	Znajduje strony, gdzie odnośniki do kolejnych stron zawierają tekst <i>tekst</i> .
applet:klasa	Znajduje strony z appletami ¹¹ wyspecyfikowanej klasy.
domain:domena	Znajduje strony z wymienioną domeną: np. dla „domain:pl” w odpowiedzi zostaną wyświetlone polskie strony.
host:nazwa	Znajduje strony na komputerze określonym jako <i>nazwa</i> .

¹⁰ http://www.altavista.com/av/content/help_advanced.htm

¹¹ Applety – miniprogram napisany w języku Java i osadzony na stronie WWW, uruchamiany w momencie przeglądania przez kompilator Java, podobnie skrypty napisane w języku JavaScript – zespół instrukcji, określający jak komputer ma wykonać zadanie.

1	2
image:plik	Szuka stron z wstawionym obrazkiem o nazwie <i>plik</i> . (p. Uwaga).
link:tekstURL	Szuka stron odwołujących się do adresów <i>tekstURL</i> .
text:tekst	Wyszukuje strony z występującym ciągiem znaków <i>tekst</i> w dowolnym miejscu, wykluczając nazwy obrazków, linków, URL.
title:tytuł	Wyszukuje strony o tytule <i>tytuł</i> .
url:adres	Wyszukuje strony z podanym fragmentem adresu.

Uwaga: Szukający plików graficznych podając nazwę pliku, powinien brać pod uwagę, że serwis przeszukuje całe ścieżki dostępu do plików. Dlatego w odpowiedzi możemy otrzymać nie tylko strony zawierające nazwę poszukiwanego obiektu, ale również strony posiadające w ścieżce dostępu do rysunku poszukiwany ciąg znaków.

Infoseek www.infoseek.icm.edu.pl www.infoseek.com

Ten serwis wyszukiwawczy posiada lokalny oddział w Polsce, kataloguje wyłącznie strony WWW i listy dyskusyjne polskiego Internetu. Bardzo ciekawą funkcję oferuje strona zwrotna serwisu: Opcja „Poszukaj wśród otrzymanych wyników” (search within these results), umożliwiającą zawężenie obszaru kolejnego wyszukiwania do stron uzyskanych w wyniku poprzedniego przeszukiwania.

Wyniki wyszukiwania można zawęzić do określonej lokalizacji, czyli miejsca serwera stron WWW za pomocą opcji: Search by location.

– Opcja Simple – wyszukiwanie proste oferuje możliwości używania następujących znaków: rozróżniania dużych i małych znaków, cudzysłówów, plusa, minusa, przecinka i dwukreska (filtra).

– Opcja Advanced¹² – wyszukiwanie zaawansowane daje możliwość używania kombinacji warunków pojawienia się wyrazów na stronie. Pewnym ułatwieniem obsługi serwisu szczególnie dla użytkowników nie mających styczności z operatorami logicznymi jest rozmieszczenie okienek wyboru opcji, precyzujących akcję, miejsce i terminy wyszukiwania. Okno Search oferuje 4 możliwości wyboru miejsca szukania wyrazu: DOKUMENT – w tekście dokumentu WWW, TITLE – w tytule strony WWW, URL – w adresie URL strony, HYPERLINK – w napisach hiperlinków do stron.

W okienku Web for pages wybór opcji MUST, SHOULD lub SHOULD NOT zostanie zrozumiany przez przeszukiwarke, iż poszukiwany ciąg znaków powinien, może lub nie może wystąpić na stronie wynikowej. Za pomocą opcji dostępnych w okienku Contain the definiujemy rodzaj wyrażenia wyszukiwanego. Jeśli żądamy, żeby wpisane wyrazy były potraktowane jako grupa słów, uruchamiamy

¹² <http://infoseek.icm.edu.pl/special.html>, <http://infoseek.go.com/Help.html>

opcję WORD'S, jako zdanie – wybieramy PHRASE, jako nazwa własna – wciskamy NAME. W następnej serii pól można dodatkowo ustalić warunki wyświetlenia bądź wykluczenia na liście wyników stron o wybranych domenach: COM, GOV, EDU, NET, ORG.

W oknie Show my results istnieje możliwość określenia, w jakiej formie mają być zaprezentowane strony końcowe. SHOW/HIDE SUMMARIES oznacza żądanie, aby w odpowiedzi przy każdej pozycji było wyświetlone/ukryte, krótkie podsumowanie na temat strony. W oknie Number of Results podawana jest liczba pozycji na stronie.

Pole Search by collection odwołuje się do tradycyjnie ułożonego przez serwis Infoseek katalogu tematycznego, skąd można wybrać dział wyszukiwania. Search by location pozwala na określenie lokalizacji wyszukiwanych stron.

Wyszukiwanie po polach meta. Jeśli nie określono nazwy pola meta, w dokumencie HTML przeszukiwana jest zawartość wierszy <TITLE>, <SUMMARY> i <BODY>. Przeszukiwarka udostępnia następujące pola meta:

Nazwa pola	Opis
link	link wewnątrz dokumentu
site	strony na określonym serwerze
url	URL dokumentu
title	tytuł dokumentu
alt	słowo wewnątrz tabeli

Chcąc np. przejrzeć strony różnych wydziałów i instytucji UMK, czyli zawierające ten sam fragment domeny, wpisujemy *url:uni.torun.pl*. Chcąc obejrzeć strony na serwerze polskiej TV o tematyce sportowej wpisujemy *sport site:www.tvp.com.pl*.

Yahoo! www.yahoo.com

Yahoo jest typowym katalogiem tematycznym, posiadającym możliwość przeszukiwania omówionych na jego stronach serwisów. Wszystkie znajdujące się w bazie danych serwisu strony WWW są bowiem katalogowane nie przez automat przeszukujący Internet, lecz przez specjalną redakcję. Na stronie głównej poniżej pola zapytań umieszczony jest wykaz kategorii katalogu. Zaletą tej wyszukiwarki jest kombinacja wyboru kolejnych grup tematycznych, które można wykorzystywać również jako słowa kluczowe, co przyspiesza dostęp do wybranego zagadnienia.

Dużą popularnością cieszy się strona serwisu, którą można zobaczyć po naciśnięciu przycisku Cool Links. Redaktorzy Yahoo umieścili w tym miejscu dokumenty, które ich zdaniem wyróżniają się wśród skatalogowanych stron. Z poziomu strony głównej można również przejść do serwisów lokalnych lub krajowych.

– Po wciśnięciu opcji **Advanced**¹³ pojawia się inna strona: z możliwościami przełączania metody szukania i wyboru obszaru szukania. Każąc wyszukiwarce wziąć pod uwagę wszystkie wpisane wyrazy – **MATCHES ON ALL WORDS** w polu A search method – odpowiada to spójnikowi **AND**, w efekcie strony wynikowe będą zawierały wszystkie te wyrazy. Podobnie, klikając na **MATCHES ON ANY WORDS**, wybiera się spójnik słów **OR**. Wpisane wyrazy w okienku można kazać potraktować programowi jako zdanie – opcja **AN EXACT PHRASE MATCH**. Opcjonalnie można w wyrażeniu wyszukiwawczym wpisać wymienione operatory bądź cudzysłowy i zaznaczyć opcję ręcznego sterowania procesem kojarzenia zapytań **INTELLIGENT DEFAULT**. W polu A search area, żeby wskazać na szukanie na stronach **WEB**owych, zaznacza się opcję **WEB Sites**, lub pierwszym krokiem szukanie wśród już posegregowanych w serwisie kategorii tematycznych – wybiera się odpowiednio **YAHOO CATEGORIES**.

– Osoby korzystające z tej wyszukiwarki w zapytaniach mogą używać zdań, ujmując je w cudzysłowach, a także plusa, minusa, symbolu gwiazdki i operatorów logicznych **AND**, **OR**. Ponadto w szczególnych przypadkach można odwołać się do pól meta: adresu **URL** („u:”) lub tytułu („t:”).

Lycos www.lycos.com

Lycos umożliwia przeszukiwanie zasobów Internetu zarówno poprzez zadawanie pytań, jak i przeglądanie katalogu tematycznego, którego kategorie nadrzędne znajdują się na głównej stronie serwisu. Serwis oferuje jeszcze gotowe opcje wyszukiwania na mapie świata danego miasta – opcja City-Guide lub adresów e-mail o podanych personaliach poszukiwanych osób (niestety, mieszkających w USA) – PeopleFind.

– Podobnie jak **AltaVista** serwis udostępnia przeglądanie wyników w różnych formach: zwrotne strony będą wyświetlane jako tytuły dokumentów – Just the links, albo dodatkowo zawierały pierwsze słowa tekstu i informacje o ich wielkości – Standard descriptions albo jeszcze streszczenie ich zawartości – Detailed descriptions.

– Tryb Advanced Search¹⁴ zawiera sporo okienek wyboru opcji zaawansowanych, co przyspiesza i uprzyjemnia pracę użytkowników. W tym trybie można stosować operatory **AND**, **OR**, **NOT**, a również operator **ADJ** wskazujący, iż wyrazy w wyszukanych dokumentach muszą wystąpić obok siebie w dowolnej kolejności, operator **NEAR**, określający odległość pomiędzy wymienionymi wyrazami – maksymalnie 25 słów oraz operator **BEFORE**, który żąda, żeby pierwsze ze słów wystąpiło przed drugim.

– W polu Search for trybu zaawansowanego szukania można deklarować wyszukiwanie dokumentów zawierających wszystkie wpisane wyrazy lub przynaj-

¹³ <http://www.yahoo.com/search/syntax?.html>

¹⁴ <http://www.lycos.com/help/search-help.html>

mniej jeden dowolny w wyrażeniu (operatory AND, OR); kolejność słów, odległość, dokładne brzmienie zdania. Można wykorzystać uporządkowany tematycznie katalog serwisu zaznaczając pasujący dział na liście i rozwijając go tematycznie. Na przykład chcąc odnaleźć piosenkę zespołu, trzeba po pierwsze ustawić opcje Search na Sounds, po drugie – wpisać w polu nazwę grupy i kliknąć Go Get It. W podobny sposób można znaleźć zdjęcia lub rysunki, wybierając opcję Pictures. Uściślenie obiektu przeszukiwania dokonuje się przez zaznaczenie odpowiedniej opcji w polu Search the; do wyboru są dokument WWW, tytuł, adres URL i określenie domeny. Bardzo ciekawe funkcje zawarte są w polu Sort results by ranking importance: określa się stopień ważności (niski/średni/wysoki) właściwości dokumentu: odnalezienia na stronie wszystkich słów, częstotliwości powtarzania słów, spotkania wyrazu na początku tekstu, w tytule, powiązania słów, zachowania ściślego porządku słów itp.

NEToskop www.chip.pl

Serwis znajdujący się na stronie głównej redakcji pisma informatycznego „Chip”¹⁵ potrafi katalogować strony zapisane we wszystkich standardach kodowania polskich liter. Katalogowane strony niezależnie od sposobu kodowania „ogonków” zamieniane są na postać bez polskich liter, wskutek czego nie są one potem obecne w podawanych przez serwer odpowiedziach. W celu zadania pytania wystarczy w znajdującym się na stronie głównej polu edycyjnym wpisać poszukiwane wyrażenie, a następnie wcisnąć przycisk Szukaj.

Pytania można zadawać w jednym z dwu trybów – przybliżonym lub dokładnym. W trybie dokładnym serwis wyszuka strony zawierające podane w zapytaniu słowo w identycznym brzmieniu. Drugi tryb pracy umożliwia wyszukiwanie dokumentów zawierających wyrazy np. o zmienionych końcówkach, co jest bardzo istotne w przypadku stron polskojęzycznych. Wyboru trybu dokonuje się na stronie głównej serwisu poprzez wybranie odpowiedniej wartości opcji Wyszukiwanie.

Opcja Odpowiedzi pozwala na zdefiniowanie czy wyświetlone przez NEToskop nagłówki stron powinny zawierać tylko ich tytuły wraz z odsyłaczami, czy też zarazem ich krótkie opisy. Złożone pytania wymagają umieszczenia operatorów logicznych (a w NEToskopie to są AND, OR i ANDNOT) oraz nawiasów. Można też użyć operatora NEAR, pozwalającego na wyszukanie stron, na których wyrazy znajdują się w odległości najwyżej 3 słów od siebie, operatora NEXTTO, który wymaga, żeby wyrazy na wyszukanych stronach znajdowały się tuż obok siebie i w zadanej kolejności.

NEToskop jest jedynym serwisem, który wprowadził operator @, umożliwiający wyszukanie informacji z ograniczeniem do stron konkretnego serwisu (serwera).

¹⁵ <http://netoskop.chip.pl/~netoskop/info.html>

HOTBOT www.hotbot.com

Ten szperacz jest jednym z najbardziej wydajnych serwisów wyszukiwawczych. Aktualnie jego baza danych¹⁶ zawiera ponad 110 milionów skatalogowanych stron. Serwis posiada kilka dodatkowo bardzo interesujących opcji, znacznie rozszerzających możliwości w zakresie wyszukiwania. Środowisko nieco różni się od pozostałych serwisów¹⁷: graficzne rozwiązanie wyboru pól meta znacznie upraszcza definiowanie warunków szukania żądanej informacji. Niestety w pasku stanu adres aktualnie podświetlonej pozycji jest przekłamywany.

– Skorzystanie z serwisu przy ustawieniu standardowym sprowadza się do wyboru odpowiednich opcji w 3 okienkach wyboru.

– Kliknięcie przycisku More Search Options usprawnia poszukiwanie poprzez udostępnienie rozszerzonych funkcji wyszukiwania. Na rys. przedstawiona została strona zaawansowanego wyszukiwania. Żeby ułatwić czytelnikowi zrozumienie przeznaczenia oferowanych opcji na tej stronie, posłużymy się szczegółowym opisem każdej po kolei. Okno podstawowe ma tytuł Search The Web, gdzie wpisuje się wyrażenie wyszukiwawcze. Do uruchomienia procesu szukania na tej stronie umieszczone zostały dwa przyciski Search: po wciśnięciu pierwszego na czerwonym tle i wyświetleniu listy wyników pozostaje strona z możliwościami zaawansowanymi, drugi przycisk Search na tle zielonym wywołuje w finale powrót do strony głównej serwisu.

1. Okno LOOK FOR z następującymi opcjami:

- all the words – oznacza żądanie, żeby wszystkie słowa podane w wyrażeniu zostały odnalezione na stronie, odpowiednikiem opcji jest używanie operatora AND lub plusa w wyrażeniu;
- any of the words – przy wyborze tej opcji przynajmniej jedno ze słów powinno być na stronie, innym rozwiązaniem jest użycie operatora OR pomiędzy słowami;
- exact phrase – wyszukuje ściśle zdania, co odpowiada ujęciu wyrażenia w cudzysłowach;
- the page title – powoduje, że szukany wyraz znajdzie się w tytułach stron WWW;
- the person – wskazuje na wyszukiwanie osoby;
- links the URL – będą przeszukiwane hyperlinki do stron;
- Boolean phrase – opcja umożliwia w wyszukiwanym wyrażeniu użycie operatorów logicznych.

2. W oknie LANGUAGE można wybrać język stron wynikowych. Niestety nie ma opcji wyszukiwania stron w języku polskim.

3. Oprócz wyspecyfikowanego wyrażenia (w głównym oknie szukania) można podać dodatkowe wyrazy. W tym celu należy najpierw w oknie WORD FILTER w polu pierwszym wybrać warunek precyzujący zawartość strony wynikowej.

¹⁶ Czyli z datą 04.1999

¹⁷ <http://hotbot.com/help/tips/default.asp>

- must contain – oznacza, że strona powinna zawierać wpisany wyraz;
- should contain – strona może zawierać wpisany wyraz;
- should not contain – strona nie może zawierać podanego wyrazu.

Następne pole przeznaczone jest w celu uściślenia rodzaju wyrazu:

- word's, czyli jako słowo/grupa słów;
- person, czyli jako osobę;
- phrase, czyli jako ścisłe zdanie.

W polu poniżej wpisać należy ten dodatkowy wyraz. Cały ten proces definiowania powyższych warunków w tymże oknie wobec wyrażenia można powtórzyć, jeśli istnieje potrzeba podania jeszcze jednego wyrazu lub grupy wyrazów. Kliknięcie przycisku More terms spowoduje wyświetlenie większej liczby tych pól w celu wpisania większej liczby oczekiwanych wyrazów.

4. W polu Anytime okna DATE można określić odcinek czasu wstecz, kiedy została utworzona lub zmodyfikowana szukana strona. Zaznaczając następne pole After/Before, precyzujemy czas tej czynności wobec podanej daty. Opcjonalnie w wyrażeniu wyszukiwawczym możemy wpisać after:*dzień/miesiąc/rok* lub before:*dzień/miesiąc/rok*.

5. Bardzo ciekawe możliwości zawarte są w następnym oknie PAGE MUST INCLUDE.

- Image – zaznaczamy, jeśli spodziewamy się na stronie grafiki w formatach GIF, JPG¹⁸. Innym sposobem jest dopisanie w wyrażeniu wyszukiwawczym do głównego warunku AND feature:image;
- Audio – szukanie na stronach z muzycznym obramowaniem. W wersji tekstowej odpowiadałoby to feature:audio;
- Video – wyselekcjonowanie stron z plikami wideo. Odpowiednio można wpisać w wyrażeniu pole meta feature:video;
- Shockwave – przeglądanie stron z plikami shockwave¹⁹;
- Acrobat – będą skanowane strony z plikami Acrobat Reader²⁰, tekstowo: feature:acrobat;
- VRML – wykrywanie stron w formacie VRML²¹. Opcjonalnie możemy wpisać w polu szukania feature:vrml;
- ActiveX – szukanie kontrolki ActiveX²²; wariantem tekstowym jest feature:activex;

¹⁸ Rozszerzenia plików graficznych, najczęściej GIF, JPG.

¹⁹ Pliki zawierające schematy dźwiękowe.

²⁰ Dokument specyficznego formatu, mający rozszerzenie PDF, tworzony w programie Acrobat Writer firmy Adobe.

²¹ Język (Virtual Reality Modeling Language) do modelowania wirtualnej rzeczywistości na stronach Internetu.

²² Element aktywny w językach programowania, obsługujący takie funkcje jak grafika, animacja i dźwięk.

- Java – definiowanie warunku umieszczenia na stronie appletów Javy. Innym sposobem jest opisanie właściwości: `feature:applet`;
- JavaScript, VB Script – znajdowanie stron z osadzonymi skryptami w języku JavaScript lub VBScript; innym opisem tego warunku jest `scriptlanguage:język`; bez wyszczególnienia języka pole meta wyglądałoby `feature:script`;
- MP3 – znajdowanie plików MP3²³;
- WinMedia – wyszukiwanie plików multimedialnych w formacie MS Windows;
- Extension – ograniczenie wyszukiwania do stron zawierających pliki z wyspecyfikowanym rozszerzeniem w okienku obok. Tekstowo ten warunek wyglądałby z użyciem pola meta `outgoingurltext:rozszerzenie`.

W wersji graficznej nie ma możliwości podania wymogów szukania na stronach zawierających tabele oraz w tytułach stron. Dlatego należy pamiętać o składni pól meta: `feature:table` oraz `title:wyrac` odpowiednio.

6. Użytkownik w celu przyspieszenia akcji szukania może wybrać w oknie LOCATION/DOMAIN określenie lokalizacji serwera, gdzie została umieszczona strona WWW albo poprzez miejsce – Continent, albo poprzez adres domenowy – Domain. Innym rozwiązaniem podania adresu domenowego jest wpisanie w wyrażeniu wyszukiwanym `domain:[domena]`, gdzie dopuszcza się trzypoziomowe domeny (3 kropki) ze wskazanym ciągiem znaków.

7. Okno PAGE DEPTH zawiera opcje wyboru miejsca w tekście, gdzie ma być szukane wyrażenie:

- Any page – w miejscu dowolnym;
- Top of Page – u góry strony;
- Personal Page – tylko na stronach osobistych;
- Page Depth – czyli „głębokość strony” – własny parametr (zdefiniowany przez redaktorów serwisu), mierzony w procentach, podawany jest w polu Page Depth, i szacujący wartość informacji na danej stronie dla użytkownika (max 99).

8. Dla angielskojęzycznych internautów okno WORD STEMMING może się okazać bardzo przydatne, gdyż toleruje różne odmiany wpisanych wyrazów w wersji angielskiej.

9. Sposób opisu stron wynikowych zależy od ustawienia opcji w oknie RETURN RESULTS:

- Full descriptions – podająca pełny opis;
- Brief descriptions – podająca opis skrócony;
- URL's only – podanie tylko adresów URL stron wynikowych.

Aby wygodniej przeglądać zawartość strony, w odpowiednim polu definiujemy liczbę wyświetlanych pozycji na jednej stronie.

²³ Pliki muzyczne Real Player zapisane w formacie kompresji MP3, w takim formacie są archiwizowane np. audycje radiowe i nagrania płytowe.

3. Opis metod wyszukiwania konkretnych informacji

Wyszukiwanie wieloparametryczne

ZADANIE: znaleźć strony bibliograficzne F. Chopina z jego zdjęciem i utworami w postaci plików dźwiękowych.

a) EWOLUCJA SKŁADNI WYRAŻENIA WYSZUKIWANEGO (na przykładzie Altavisty opcja wyszukiwania zaawansowanego, język stron – dowolny)

Lp.	Wyrażenie	Wynik	Komentarz
1	„Fryderyk Chopin” OR „Frederic Chopin”	5313	Wariant imienia w polskim i obcym brzmieniu
2	(„Fryderyk Chopin” OR „Frederic Chopin”) AND biblio*	163	Nawiasy potrzebne po to, żeby warunek AND odnieść do dwóch wersji imion.
3	(„Fryderyk Chopin” OR „Frederic Chopin”) AND biogr*	479	Stron z biografią 3 razy więcej niż stron z bibliografią.
4	(„Fryderyk Chopin” OR „Frederic Chopin”) AND image:chop*	218	Zawężanie listy wynikowej do stron z ilustracjami pod nazwami (zgodzanie) rozpoczynającymi się „chop”.
5	(„Fryderyk Chopin” OR „Frederic Chopin”) AND image:chop* AND biblio*	2	Strony z bibliografią i prawdopodobnie portretami Chopina.
6	(„Fryderyk Chopin” OR „Frederic Chopin”) AND biblio* AND image:chop*	2	Jak wyżej z zamianą kolejności słów; wynik bez zmian.
7	(„Fryderyk Chopin” OR „Frederic Chopin”) AND image:chop* AND biogr*	35	Strony z biografią i prawdopodobnie portretami Chopina.
8	(„Fryderyk Chopin” OR „Frederic Chopin”) AND biogr* AND image:chop*	31	Jak wyżej z zamianą kolejności słów. Na różnicę wpływa umieszczenie wyrazu „biografia” w tytule strony.

Należy odróżniać sekwencję logiczną 1 w tabeli od na przykład potraktowanych wybiórczo imienia i nazwiska: „*Fryderyk OR Frederic AND Chopin*”, co daje w wyniku 7364 stron, albo wyrażenia „*Fryderyk OR Frederic NEAR Chopin*”, jeszcze bardziej skracającej listę do 6470 ze względu na odległość.

Intuicja w podaniu nazw obrazków jest niezbędna; w przypadku wyrażenia 4 wynik okazał się bardziej satysfakcjonujący niż gdybyśmy szukali „*image:fryd**” z wynikiem równym tylko 1. Altavista nie jest przydatnym serwisem, kiedy chcemy odseparować strony z plikami dźwiękowymi: nie ma na razie (można spodziewać się, że w miarę rozwoju serwisu autorzy naprawią podobne wady programowe) zdefiniowanego odpowiedniego pola meta.

b) PORÓWNANIE WYNIKÓW WYSZUKIWANIA ZAAWANSOWANEGO STRON WWW POŚWIĘCONYCH F. CHOPINOWI PRZEZ OPISANE WYSZUKIWARKI

Bardzo zróżnicowane funkcje i możliwości serwisów wyszukiwawczych utrudniają analizę porównawczą w postaci tabeli. Opisywanie szczegółów akcji wyszukiwawczej każdego szperacza powinno użytkownikowi ułatwić wybór odpowiedniego środowiska w celu znalezienia żądanej informacji.

Altavista

Wyniki i opis wyszukiwania podane są w punkcie a).

Infoseek

Opcje i liczba okien pozwalają na ułożenie tylko niektórych kombinacji wyrazów.

Ustawienie dla imion opcji SHOULD, dla Chopin – MUST, co jest równoznaczne wypisaniu „*Fryderyk Frederic +Chopin*”, podaje bardzo szeroką listę odpowiedzi – 20 119 stron, na których na pewno będzie występować nazwisko kompozytora. Zamiana kolejności „*+Chopin Fryderyk Frederic*” silnie zredukuje listę: wynik – 298.

Po wpisaniu wyrażenia „*Fryderyk Chopin*” jako zdania i ustawieniu metody jako MUST, dla „*Frederic Chopin*” odpowiednio jako – SHOULD otrzymamy tylko 259 pozycji. Zamieniając zawartość okienek wpłyniemy na ważność i częstotliwość występowania imion w obcym brzmieniu, w wyniku uzyskamy już 1870 stron. Niestety nie ma możliwości wyspecyfikowania wyrażenia wyszukiwawczego tak jak w przypadku Altavisty, w punkcie 1 w tabelce.

Yahoo

Jedynymi wyrażeniami, które zdołała odszukać ta niewydajna w tym przypadku wyszukiwarka, były „*Fryderyk Chopin*” z wynikiem 1 oraz „*Frederic Chopin*” – z 9 stronami.

Lycos

Wynik wyszukiwania w opcji zaawansowanej wyrażenia „*Fryderyk Chopin*” równa się 196 dla „*Frederic Chopin*” – 902 strony, zdecydowanie mniej niż wyszukiwała Altavista. Dla polskiego imienia znaleziono tylko 3 strony z rysunkami, dla obcego – 0, dzięki zaś wykryto na jednej stronie. Wypisanie wyrażenia z wymaganiem obecności wyrazu Chopin na stronie „*Fryderyk Frederic +Chopin*” daje w wyniku 33436 stron – o 13 000 stron więcej niż w Infoseeku.

Hotbot

Frazę „*Fryderyk Chopin*” wyszukiwarka znalazła aż na 450 stronach oraz „*Frederic Chopin*” – na 1490 odpowiednio. Łącząc te wyrażenia spójnikiem OR jak w punkcie 1 tabeli dla Altavisty otrzymujemy 1900 stron. Po dodaniu warunku obecności ilustracji na stronach „*image:chop**” odpowiedź podawana jako 1400 pozycji (p. 4 w tabeli).

Żeby porównać z pozostałymi punktami w tabeli należy dopisać wyrazy „*biblio**” – wynikiem będzie 5 stron, bądź „*biogr**” – 13 stron, natomiast dołączając warunek ilustracji na stronie otrzymujemy zero wyników.

Tej wyszukiwarce udało się odnaleźć udźwiękowane strony z wyrażeniem logicznym 1 (w tabeli) – wynalazła w rezultacie 47 stron; po dodaniu do wyrażenia wyrazów „*biblio*” bądź „*biogr*” wylistowano odpowiednio 11 i 4 strony.

Odnalezienie osoby w społeczności Internetu

W celu sprawdzenia opcji wyszukiwania osób należy wybrać osoby dobrze znane, publiczne, ale nie działaczy politycznych, żeby wykluczyć strony doniesień prasowych. Wytypowaliśmy dziennikarkę, aktora i muzyka.

ZADANIE: Wynaleźć strony z nazwiskiem Hanny Smoktunowicz, Bogusława Lindy lub Grzegorza Ciechowskiego.

Hotbot: z zaznaczoną opcją szukania osoby (PERSON) odnalazł odpowiednio 3, 48 i 31.

Yahoo: poprawnie wpisuje się w polu FIRST imię, w polu LAST – nazwisko. Nie odnaleziono żadnej strony z wymienionymi nazwiskami, ponieważ na razie pole wyszukiwania osób w tym serwisie jest ograniczone do terenu USA.

Lycos, Altavista, Infoseek: taki sam negatywny wynik.

Odnalezienie określonego artykułu w Internecie

ZADANIE: „Wpadł mi kiedyś w ręce artykuł o teorii ewolucji z ilustracjami odtworzonych woskowych figur praludzi. Nie pamiętam tytułu czasopisma lecz tylko wydawnictwo: Prószyńskiego. Chciałbym znaleźć ten artykuł”.

Do tego celu świetnie nadaje się serwis NEToskop. Należy wytypować charakterystyczne słowo dla tego artykułu, np. *neandertalczyk*, oraz przypomnieć adres serwisu Prószyńskiego. Wpisując w polu wyszukiwania *neandertalczyk@proszynski.com.pl* otrzymujemy całą listę pozycji: na początku wyświetlone zostały książki, poniżej czasopisma. Po krótkim podglądzie udało się odnaleźć odpowiedni numer „Wiedzy i Życia”.

Taki sam wynik, ale po dłuższym czasie wyświetla wyszukiwarka Hotbot, jeśli wpisujemy w polu **Domain** domenę serwisu Prószyńskiego.

Wynikowe porównanie na podstawie badań możliwości przeszukiwarek

Opisane przykłady posłużyły jako testy analizy możliwości wyszukiwania zaawansowanego serwisów informacyjnych. Czas trwania procesu wyszukiwania nie jest odpowiednim parametrem do porównywania, ponieważ testy były prowadzone w różnym czasie, na różnych komputerach z różniącymi się łączami i kartami sieciowymi w zależności od dostępu. Jednak da się zauważyć bezwzględną szybkość działania Hotbot i Altavisty. Na podstawie opisanych powyżej procesów przeszukiwania spróbujemy porównać inne charakterystyki.

Funkcje	Altavista	Infoseek	Hotbot	Lycos	Yahoo	Netoskop
1	2	3	4	5	6	7
Rozróżnianie małych i dużych liter	•	•	•			
Tolerowanie znaku *	•		•	•	•	
Wpisanie złożonych wyrażeń logicznych	•••		••	•	•	
Szukanie wariantów/odmian słowa	••	•	••			
Przeszukiwanie wg określonych pól	•••	•	•••	•		
Wymóg od użytkownika znajomości tych pól	•					
Wyłowienie wyspecyfikowanych domen	•	•	•	•	•	•
Korzystanie z katalogów tematycznych serwisu	•	•	•	•	•••	
Wyszukiwanie multimedialnych	••		•••	•		
Opcje wyboru języków stron (w tym polskiego)	•••	•	•	•		
Szperanie wyłącznie na polskich stronach						•

Zróznicowanie sprawności tych funkcji w poszczególnych serwisach informacyjnych można ocenić za pomocą liczby punktów: od niskiego stopnia • poprzez średni •• do wysokiego stopnia sprawności – •••. Największe możliwości (czyli trzy punkty) pod względem przeszukiwania według pól meta oferują przeszukiwarki Altavista i Hotbot, najwydajniej skonstruowane katalogi mają Yahoo i Lycos. Podobne porównanie sprawności wyszukiwania dla różnych serwisów, ale tylko angielskojęzycznych, można znaleźć w artykule [3].

Zakończenie

Podany powyżej krótki przegląd dostępnych narzędzi poszukiwawczych, zwanych popularnie wyszukiwarkami, wykazuje, że nie tak łatwo o jednoznaczną klasyfikację i ocenę zarówno sprawności wyszukiwania podanych indeksów, jak i możliwości dostosowania dostępnych struktur składniowych do potrzeb użytkownika. Należy stwierdzić, iż dynamiczny wzrost liczby dostępnych zasobów sieciowych zmusza użytkowników do stosowania coraz bardziej wysublimowanych metod filtrowania informacji. Służą ku temu właściwie stosowane operatory logiczne, pola meta czy też korzystanie z dodatkowych – zaawansowanych funkcji wyszukiwawczych. Trudno decydować jednoznacznie, który z opisanych serwisów wyszuki-

wawczych jest najlepszy, ale podana w końcowej części pracy tabela porównawcza może być pomocna w wyborze właściwego dla konkretnego zadania. W chwili obecnej największą sprawność wykazują wyszukiwarki Altavista i Hotbot, łącząc w sobie zarówno łatwość obsługi, jak i dużą ilość skatalogowanej informacji. Jednak, biorąc pod uwagę szybki rozwój całej sieci internetowej, sytuacja ta może ulec zmianie, szczególnie w kierunku polskich serwisów, które obecnie znajdują się jeszcze w początkowej fazie rozwoju. Właściwe więc wydaje się systematyczne śledzenie i sprawdzanie działania wyszukiwarek, testowanie i porównywanie wyników poszukiwań oraz śledzenie informacji na ten temat publikowanych w różnych ośrodkach zajmujących się badaniem i przetwarzaniem informacji.