ROLE OF META SEARCH ENGINES IN WEB- BASED INFORMATION SYSTEM: FUNDAMENTALS AND CHALLENGES

Subarna Kumar Das

Abstract

Meta search engines are powerful tools that search multiple search engines simultaneously. Unlike crawler-based search engines such as Google, AllTheWeb, AltaVista and others, meta search engines generally do not build and maintain their own web indexes. Instead, they use the indexes built by others, aggregating and often post-processing results in unique ways. Meta search engines accept your query, and send it out to multiple search engines in parallel. They are often quite fast, using private "backdoor" servers made available by the search engines they query. They get this privileged status thanks to revenue sharing agreements. There are several advantages to using meta search engines, though they're not always the most appropriate tool. The most obvious advantage is that you can get results from multiple search engines without having to visit each in turn. Apart from the time savings, there is some evidence that this gives your search a broader scope, since each individual search engine's index differs from all others. Present paper attempts to highlights the role of meta search engine in web- based information retrieval

Keywords: Search Engine, World Wide Web, Meta Search Engine, Metacrawler Information Storage,

1. Introduction

A search engine is a program designed to help find information stored on a computer system such as the World Wide Web, inside a corporate or proprietary network or a personal computer. The search engine allows one to ask for content meeting specific criteria (typically those containing a given word or phrase) and retrieves a list of references that match those criteria. Search engines use regularly updated indexes to operate quickly and efficiently. Without further qualification, search engine usually refers to a Web search engine, which searches for information on the public Web. Other kinds of search engine are enterprise search engines, which search on intranets, personal search engines, which search individual personal computers, and mobile search engines. However, while different selection and relevance criteria may apply in different environments, the user will probably perceive little difference between them.

2. Meta Search Engine: Meaning

1. A search engine that compiles its results from the databases of other search engines Search engines that automatically submit your keyword search to several other search tools, and retrieve results from all their databases.

2. Convenient time-savers for relatively simple keyword searches (one or two keywords or phrases in "A meta search engine (also known as metacrawler or multi-threaded engine) is a search tool that sends your query simultaneously to several search engines (SEs), Web directories (WDs) and - sometimes - to the so-called Invisible (Deep) Web, a collection of online information not indexed by traditional search engines. After collecting the results, the meta search engine (MSE) will remove the duplicate links and, according to its algorithm, combine/rank the results into a single merged list.
3. In a meta-search engine, you submit keywords in its search box, and it transmits your search simultaneously to several individual search engines and their databases of web pages. Within few seconds, you get back results from all the search engines queried. Meta-search engines do not own a database of Web pages; they send your search terms to the databases maintained by search engine companies.
4. A meta-search engine is a search engine that sends user requests to several other search engines and/or databases and returns the results from each one. They allow users to enter their search criteria only one time and access several search engines simultaneously. Since it is hard to catalog the entire web, the idea is that by searching multiple search engines you are able to search more of the web in less time and do it with only one click. The ease of use and high probability of finding the desired page(s) make meta search engines popular with those who are willing to weed through the lists of irrelevant 'matches'. Another use is to get at least some results when no result had been obtained with traditional search engines.
5. Meta search engines create what is known as a virtual database. They do not compile a physical database or catalog the web. Instead, they take a user's request, pass it to several other heterogeneous databases and then compile the results in a homogeneous manner based on a specific algorithm.
6. A search engine that displays results from multiple search engines

3. Meta search Engine: what are the fundamentals

No two meta search engines are alike. Some only search the most popular search engines while others also search lesser-known engines, newsgroups and other databases. They also differ in how the results are presented and the quantity of engines that are used. Some will list results according to search engine or database. Others return results according to relevance, often concealing which search engine returned which results. This benefits the user by eliminating duplicate hits and grouping the most relevant ones at the top of the list. Ithaki.net uses more than 450 different engines to find results while Dogpile, allows users the flexibility to choose exactly how results are presented. Another metasearch engine, Qksearch, will let users search using any of the 3 commonly deployed metasearch methods. Yet, another meta search engine, LemmeFind processes and ranks the results from the primary search engines using its own (yet unpublished) methods and presents them to the users in a straight forward fashion. Search engines frequently have different ways they expect requests submitted For example, some search engines allow the usage of the word "AND" while others require "+" and others only require a space to combine words. The better metasearch engines try to synthesize requests appropriately when submitting them. Results can vary between metasearch engines based on a large number of variables. Still, even the most basic metasearch engine will allow more of the web to be searched at once than any one stand-alone search engine.Unlike the individual search engines and directories, the meta search engines (a) Do not have their own databases and (b) Do not accept URL submissions

4. Use of Meta search Engines

a. For quick and dirty searches. If you want an answer fast, you may have better luck querying multiple engines simultaneously.

b. For broad and shallow searches. Meta searching is an excellent approach if the purpose of your search is to get an overview of a topic.

c. To assess potential keywords for an unfamiliar subject. What better way to discover search terms than to see how they appear in a cross section of documents across the web?

d. To see how different engines handle the same query. This is an excellent way to get to know the "personalities" of different search engines — their strengths, weaknesses, and types of queries they handle best. Meta search engines present results in two ways. One way is to simply list ten or so results from each engine queried with no additional post-processing. Dogpile works this way, listing results from three engines at a time.Other meta search engines analyze the results and then rank them according to their own rules, combining results from multiple engines into a single, unified list. IxQuick, Metacrawler and Vivisimo are examples of this type of result aggregating meta search engines.

5. Meta search Engine:Types

1. "Real" Meta search Engine
2. "Pseudo" Meta search Engine type I
3. "Pseudo" Meta search Engine type II
4. Search Utilities

5.1 "Real" Meta Search Engine

These "real" MSEs simultaneously search the major search engines, aggregate the results, eliminate the duplicates and return the most relevant matches.

a. Ez2Find (formerly ez2www) [http://Ez2Find.com/]

Searches the best SEs - AlltheWeb, Google, AltaVista, Teoma, Wisenut - and directories - Yahoo and Open Directory. Through its "Advanced Search" function it also searches a small part of the Invisible (Deep) Web. It also searches news, newsgroups, MP3, images and many, many more. Provides excellent results in a very neat interface.

b. Vivísimo [http://www.vivisimo.com/]

Uses the clustering technology, meaning matches are organized in folders. Don't like the frames? Just modify the size of both the upper and the left frames. This MSE was created by researchers at Carnegie-Mellon University. Advanced searching options available: exact phrase, Boolean operators, fields searching (domain, host, title, URL, etc.) and more..

c. InfoGrid [http://www.infogrid.com/]

Provides excellent results in an easy to read layout, despite a very confusing Home Page crowded and with frames. This MSE searches the 'big ones' including Google, AlltheWeb, Yahoo! and Open Directory. It also searches newswires, auctions, discussion forums, MP3, FTP files and more.

d. Infonetware [http://www.infonetware.com/]

Searches the Web and provides relevant results, organized in topics, in a very clean interface. This MSE is based in Edinburgh, Scotland .

e. IBoogie [http://www.iboogie.com/]

Uses a minimalist design. This MSE "performs intelligent clustering of results". It searches the Web, the Invisible (Deep) Web, images, video and audio files.

f. Metaseek [http://www.pcdigest.net/metasearch/en/main.shtml]

Excellent MSE from Ukraine. Searches major international and local search engines. Besides the Web you can search images, MP3, FTP files, news and more. You can use "Phrase" ("), "natural language" processing, Boolean logic and field searching (by URL, title, site/domain or link). .

g. Fazzle [http://www.fazzle.com/]

Uses an excellent selection of search engines and directories. This MSE provides relevant results in a relatively crowded interface. For each result you can see the search engine where the hit was found, and its ranking.

h. Query Server [http://www.queryserver.com/web.htm]

Searches an impressive list of 11 SEs - everything important except Google. But don't worry: Query Server searches Yahoo!, Netscape and AOL, all partially powered by Google. This is another example of the clustering technology. Highly customizable metasearch tool.

i. Vinden.NL [http://www.vinden.nl/]

Searches "the best," providing very good results in a clean interface.

j. Meta Bear [http://www.metabear.com/]

Provides relevant results from both international and Russian sites.

k. Web Scout [http://www.webscout.com/]

Searches the Web, news, newsgroups, auctions, MP3 files and jobs.

l. Experts Avenue [http://www.expertsavenue.com/]

Searches different search engines simultaneously for Web pages, auctions, jobs and forums and provides very relevant results in a neat interface. Enables online language translation of Web pages.

m. Suchspider. de [http://www.suchspider.de/meta-suchmaschinen/]

Searches a whopping 100 (!) international SEs and WDs. Google, AlltheWeb, Open Directory, you name it. You can sort the results by relevance, source or - much better - grouped by domain name.

n. EmailPinoy [http://www.emailpinoy.com/]

Sends your query to 15 search engines. Don't use Kanoodle, ah_ha.com and GoClick pay-per-click search engines, because you'll get irrelevant results. For better results enclose phrases in quotation marks. .

o. 1 SECOND [http://www.1second.com/]

Searches a good selection of 14 major SEs and WDs, throws out the duplicates and summarizes the results in a neat listings page. Use the Advanced Search if you want to customize the search, especially the timeout of the search engines.

p. My Prowler [http://www.myprowler.com/]

Searches over a dozen search engines, news, images, audio/MP3, music videos, auctions and various other sites.

q. Gimenei [http:// www.gimenei.com/]

Use of the "Advanced Search" option is strongly recommended. You can customize the results page, including my favorite option, "All Results" in one page.

r. Search 66 [http://www.search66.com/]

Groups together pages from the same domain. Beautiful. To avoid SEs timeouts, select the "Speed": "Comprehensive." Obviously, you'll get more results from this excellent Australian MSE .

(a) Internav [http://www.internav.com/]
(b) NetXplorer [http://www.netxplorer.de/]
(c) Metengine [http://www.metengine.com/]
(d) One2Seek [http://www.one2seek.com/]
(e) Ithaki [http://www.ithaki.net/]
(f) Fossick [http://www.fossick.com/Search.htm]
(g) Pandia [http://www.pandia.com/powersearch/index.html]
(h) meta EUREKA [http://www.metaeureka.com/

(i) Widow [http://www.widow.com/]
(j) Meta 360 [http://www.meta360.com/]
(k) 7 Meta Search [http://www.7metasearch.com/]
(l) Metor [http://www.metor.com/]
(m) Ixquick [http://www.ixquick.com/]

s. Unimpressive meta search engines

The following is a list of some unimpressive meta search engines. These MSEs do not provide the breadth of coverage offered by the sites recommended above. Each of these has its own flawed characteristics, but generally they are old and have not kept up with the latest capabilities or they suffer from too many functional problems

(a) Metacrawler [http://www.metacrawler.com/]
(b) Dogpile [http://www.dogpile.com/]
(c) Mamma [http://www.mamma.com/]
(d) Pro Fusion [http://www.profusion.com/],
(e) Bytedog [http://www.bytedog.com/]
(f) il motore [http://www.ilmotore.com/]
(g) METASEEK.NL [http://www.metaseek.nl/]
(h) ApocalX [http://www.search.apocalx.com/])

5.2 The type I "Pseudo" Meta Search Engine

It sends the query to the search engines, and then presents the results grouped by search engine in one long, easy to read scrollable list. The best MSEs in this category are:

(a) Mall Agent [http://www.mallagent.com/web.html]
(b) qb Search [http://www.qbsearch.com/]
(c) Better Brain [http://www.betterbrain.com/] My Net Crawler [http://www.mynetcrawler.com/]
(d) NBCi [http://www.nbci.msnbc.com/]
(e) Planet Search (Sherlock Hound) [http://www.planetsearch.com/]
(f) Rede Search [http://www.redesearch.com/] 1 BLINK [http://www.1blink.com/]
(g) Search Wiz [http://www.searchwiz.com/]Search Fido [http://www.searchfido.com/

5.3 "Pseudo" Meta Search Engines Type II

There are two types of Type II "Pseudo" MSEs: a) You type your query one time and then select the search engines. One browser window will open for each SE selected. The best are:

(a) Multi-Search-Engine.com [http://www.multi-search-engine.com/]

(b) GoGettem [http://www.gogettem.com/]
(c) Search Bridge [http://www.searchbridge.com/]
(d) The Info [http://www.theinfo.com/]
(e) Net Depot [http://www.netdepot.org/]

The best of this type are:

(a) Alpha Seek [http://www.alfaseek.com/]
(b) Westlaser [http://www.westlaser.com/]
(c) Dan's No Overhead Search Thingy [http://www.danielc.com/thingy.html]
(d) Express Find [http://www.expressfind.com/]
(e) Freeality [http://www.freeality.com/meta.htm].

5. 4 Search Utilities

These are downloadable meta search tools that search multiple search engines. Results are collated and ranked for relevancy with redundancies removed. They are not free but most of them have a free trial version available. Example

(a) Copernic [http://www.copernic.com/en/index.html]
(b) Arrow Search [http://www.rt-software.co.uk/arrow_search/]
(c) SearchRocket [http://www.searchrocket.com/]
(d) WebFerret [http://www.ferretsoft.com/index.html].

6. The Big Four Meta Search Engines

6.1 Dogpile

It is designed as a mainstream consumer site, with a simple presentation. Results are grouped by provider, typically with Overture and FindWhat results presented first. This presentation makes it easy to compare and contrast results from different search engines for the same query, and is one of Dogpile's most useful features.Dogpile offers no advanced search capabilities, though you can limit your search to Web search; Images from Ditto and Fast Image; Audio/MP3 from Astraweb, Fast Audio and MP3Board; Files from Fast FTP; News from Dogpile Newscrawler and Fast News; or Multimedia files from Fast.You can also select, to an extent, which search engine results you want to be displayed first, or not displayed at all, using the "custom search" form.

6.2 Metacrawler

"Metacrawler is a hardcore site for the sophisticated searcher," said InfoSpace's Baur. Unlike Dogpile's approach of presenting results grouped by their original source, Metacrawler blends results based on relevancy and performance. Speed is emphasized — if a particular engine isn't processing a query quickly, its results won't be included in your Metacrawler results. The source of each result is clearly labeled. This lets you see that a particular result came from several search engines, or

perhaps just a single one. Seeing the attribution displayed like this offers clues that can help you decide whether to click through to the underlying page or not. For example, if a particular result is attributed to Fast, the Open Directory, and Find What, you can be relatively confident that it's an authoritative site that they've all "agreed" is a highly relevant match for your query

6.3 Excite and Webcrawler

When InfoSpace purchased Excite and Webcrawler, the intention was to keep the user experience as similar as possible to those provided by Excite, primarily to maintain the high traffic levels enjoyed by each site. To a large extent, the company has succeeded in its goal. Excite still offers most of the portal features that have long made it an appealing start page, such as news, stock quotes and so on. Webcrawler still has its clean, simple look.What's changed, though, is the underlying search engine for both properties. Search results for both are now blended metasearch results, similar to Metacrawler results. Neither Excite nor Webcrawler offer Metacrawler's advanced customization features, though you can limit your search to the web, news or photos.Results are identical for both properties, with one key exception: Webcrawler results are ad-free, apart from the links served by Overture, FindWhat and Ah-Ha. No banners; no annoying pop-ups. Of course, you don't have access to the personalization features offered by Excite, but the ad-free environment is refreshing.Bottom

7. Other Meta Search Engines

(a) C4http://www.c4.com

C4 allows meta searching against several major search engines.

(b) IcySpicyhttp://www.icyspicy.com/

IcySpicy is both a meta search engine offering results from Google, Overture, MSN, WiseNut, FindWhat, etc., and a collection of useful directory links and search forms for package tracking, movie locations, and so on.

(c) Moonmist http://www.moonmist.info/index.asp

Moonmist allows you to do a general or country specific meta search. Results include links to site info and a link to the Wayback machine's archived copies of the underlying result page.

(d) Searchy.co.uk http://www.Searchy.co.uk

Searches 15 U.K. engines. The advanced search form allows you to change the order that results are presented, either by speed or manually to suit your own preferences.

(e) TeRespondo http://sl.terespondo.com

Spanish metacrawler that searches on the more popular search engines. Motor de busqueda que busca tu consulta en los buscadores mas populares en espanol.

(f) Turbo10 http://turbo10.com/

Turbo10 is a metasearch Engine accesses both traditional web search engines and some invisible web databases, with a very speedy interface.

(g) Watson for the Macintosh http://www.apple.com/downloads/macosx/internet_utilities/watson.html

Watson is a "Swiss Army Knife" with nineteen interfaces to web content and services — an improvement on Sherlock, with nearly twice as many tools, including Google Searching.

(h) Widow Meta Searchhttp://www.widow.com

Easy to choose exactly which search engines you wish to query, and the same listings found at multiple search engines are combined together.

(i) SearchIQ http://www.zdnet.com/searchiq/directory/multi.html

Not a meta search engine but instead reviews of meta search engines.

8. Challenges faced by search engines

The search engine are facing some challenges:

- The web is growing much faster than any present-technology search engine can possibly index). In 2006, some users found major search-engines became slower to index new webpages.
- Many webpages are updated frequently, which forces the search engine to revisit them periodically.
- The queries one can make are currently limited to searching for key words, which may result in many false positives, especially using the default page-wide search. Better results might be achieved by using a proximity-search option with a search-bracket to limit matches within a paragraph or phrase, rather than matching random words scattered across large pages.
- Dynamically generated sites may be slow or difficult to index, or may result in excessive results, perhaps generating 500 times more webpages than average. Example: for a dynamic webpage which changes content based on entries inserted from a database, a search-engine might be requested to index 50,000 static web pages for 50,000 different parameter values passed to that dynamic webpage.
- Many dynamically generated websites are not indexable by search engines; this phenomenon is known as the invisible web.
- Some search-engines do not rank results by relevance, but by the amount of money the matching websites pay.
- In 2006, hundreds of generated websites used tricks to manipulate a search-engine to display them in the higher results for numerous keywords. This can lead to some search results being polluted with linkspam or bait-and-switch pages which contain little or no information about the matching phrases. The more relevant webpages are pushed further down in the results list, perhaps by 500 entries or more.

9. Conclusion

Meta search engines look pretty much the same up front, but their approach to presenting results varies widely. Meta-search engines do not have their own databases, results are only as good as the quality of the sources the engines query. As they improve their capabilities and become better known, the number of people using them will continue to grow. Ensuring that your Web site is found not only by search engines but also by meta-search engines can help you tap into an audience of millions. Traditional search engines only indexed a fraction of the Web. Meta-search engines covered much more. It's powerful because one search can highlight strengths of a number of top search portals, such as Google, AltaVista, LookSmart, and Yahoo whatever the challenges faced by search engines in web- based system.

References

1. http://www.dogpile.com
2. http://searchenginewatch.com
3. www.clickz.com
4. www.marketingterms.com/dictionary/metasearch_engine
5. www.infolibrarian.com
6. NACLIN 2003 (Proceedings)
7. http://en.wikipedia.org/wiki/Metasearch_engine
8. http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.
9. www.dogpile.com
10. http://www.web-design-in-new-york.com/meta_search_engines.html
11. www.searchengineguide.com
12. http://www.evolt.org/node