

## SUBJECT ACCESS ON INTERNET : HIGHLIGHTS OF THE METASEARCH ENGINES

*Wendy Tan*

*Head Cataloger  
Wexler Library  
Hunter College, City University of New York  
New York, U.S.A.*

### Abstract

Metasearch engines can facilitate Internet searchers to retrieve useful results in an efficient way. How do these search tools differ from the index search engines? This article examines ten popular engines and provides some tips for using them.

### Keywords :

Internet search engines ; World Wide Web ; Information retrieval

### What Are These Tools?

Metasearch engines make it possible to perform an Internet search through many search engines at once. In general, they don't have their own databases, but they act as a bridge to visiting a few popular search engines simultaneously and top responses from each engine will be returned. However, some metasearch engines may list a number of engines and allow a user to search one database at a time directly from that page.

### Advantages vs. Disadvantages

#### A. Advantages

##### 1. Efficiency

For the past few years, Web content has grown by leaps and bounds. With all of this information available, getting what you were looking for became very difficult. Although the popularity of search engines brought some order out of chaos, since each engine uses its own algorithms, it is very common that the same query produces different results between two engines. Therefore, the pattern for an experienced searcher is to try some favorite search engines and pick top responses found by those engines. This process is exactly what metasearch provides.

## 2. A good overview and selection

Some metasearch engines offer options for sorting the results, removing duplicates, ranking the remaining items and listing them according to scores. This type of engine is very useful for getting an idea of what is available on the Internet from a variety of search engines.

### B. Disadvantages

#### 1. Small sampling

One of the down sides of metasearching is the lack of comprehensive results. Users may steer clear of metasearch engines when thorough research is desired.

#### 2. No access to special features of individual engines

Advanced searching features available in many search engines are not recommended when metasearch engines are used. The search protocol of metacrawlers is to make the query relatively simple to be executed successfully across the board.

Below are the brief descriptions of ten major metasearch engines.

### **Ask Jeeves**

<http://www.askjeeves.com/>

#### Key features

1. Besides searching other Internet resources, Ask Jeeves has its own database which is built by research staff.

2. Questions can be typed in plain English. Whole sentence is allowed, and sometimes it works better than a phrase or Boolean searching. Spelling check is also available.

#### **How it works**

Uses type in a question in plain English and Ask Jeeves shows a list of a few possible answers. After the user selects the best match, and clicks <go to> next to the entry, Ask Jeeves responds with one web site which should answer the question.

#### **Information provided in results index screen**

A few entries with drop down choices respectively are returned first, and the user has to select one of them to get complete information of the site.

### **C4 Cyber 411**

<http://www.cyber411.com/>

#### **Key features**

1. Coverage : Altavista, DejaNews, Excite, Galaxy, Goto, Hotbot, Looksmart, Lycos, Magellan, Northern Light, PlanetSearch, Search.com, Thunderstone, Webcrawler, What-U-Seek, Yahoo, etc. 16 popular engines are available for searching. A few or all of them may be selected.

2. Options for customization :

a. Display results that are received within (10 ; 20 ; 30 ; 40 ; 50 ; 60 ; 90 ; 120 second) ;

b. Number of concurrent connections (1-10).

3. Cyber 411 supports many foreign languages.

#### **How it works**

Users may type the query in a single word, strings of words enclosed in quotation marks, or using Boolean operators (AND, OR, NOT, NEAR) to combine words and results are gathered from the engines selected. Duplicate URLs are removed and all entries are shown at once.

#### **Information provided in results index screen**

Name of the engine, title and URL.

### **Dogpile**

<http://www.dogpile.com/>

#### **Key features**

1. Coverage : Yahoo! Lycos' A2Z, Excite guide, Go To Com, Planetsearch, Thunderstone, What U Seek, Magellan, Lycos, Web-crawler, InfoSeek, Excite, AltaVista, etc. 13 engines are available.

2. Options for customization :

- a. Medium to be searched (The Web ; Usenet ; FTP ; News wires ; Weather ; Stock quotes).
- b. waiting time (10-60 seconds).
- c. The order of the engines to be examined.

#### **How it works**

1. Users may use Boolean operators (AND, OR, NEAR, NOT) to combine words and phrases. However, NEAR, OR, and NOT should be used with caution because many engines do not support them. Using no connector, AND will be inserted.

2. Quotes and parentheses are allowed. For those engines which don't support their use, they will be ignored.

#### **Information provided in results index screen**

It depends on individual engines. Some engines display titles only ; some have titles, URLs, summaries ; some can provide very detailed formats such as titles, rating, reviews, URLs, and references etc.

### **Highway 61**

<http://www.highway61.com/>

#### **Key features**

1. Coverage : Yahoo, Lycos, Webcrawler, Infoseek and Excite, etc. 5 engines are available.

2. Options for customization :

- a. Search as Boolean (or ; and).



- b. Your patience level (I'm a reasonable person ; Hurry up! You loser! Please try and make it quick! Time is a relative thing ; Take your time! I'm going to bathroom).
- c. How many hits (Lots ; Bury me).
- d. Will the armadillo make it across the road? (I don't bet on road kill ; I pray that he does ; No, I'm afraid he won't ; He doesn't stand a chance).

#### How it works

1. Users may type a query with a single word or multiple words. "OR" is assumed if no connector between words. "AND" has to be chosen from options if all of the terms need to be included.

#### Information provided in results index screen

Name of the engine, title, summary, score (on a scale of 100) and URL.

### Inference Find

<http://www.infind.com/>

#### Key features

1. Coverage : AltaVista, Excite, Infoseek, Lycos, Webcrawler, Yahoo etc. 6 engines are available.

2. It retrieves the maximum number of results each engine will allow, merges the results, removes redundancies, and then groups the related items together. It is the only metasearch tool which offers "clustering" feature.

3. Option for customization : Max time (1 ; 5 ; 7 ; 10 ; 30 seconds).

4. Searching is also available in German and French.

#### How it works

Boolean operators "AND" "OR" "NOT" may be used to make more exact queries. However, since some engines such as Infoseek and Yahoo ignore Boolean operators, this function is sometimes inconsistent and the accuracy of results cannot be assured.

#### Information provided in results index screen

It classifies items into different sites such as commercial sites ;

educational sites etc. Within each site, only titles are listed.

### **Mamma**

<http://www.mamma.com/>

#### **Key features**

1. Coverage : AltaVista, Excite, Infoseek, Lycos, WebCrawler, Yahoo, etc. 7 major engines are available.
2. Options for customization :
  - a. Search options (The Web ; Usenet ; Newswires ; Stock symbol ; company name ; MP3 files ; picture ; sound).
  - b. Show summaries (v ; blank).
  - c. Search page titles (v ; blank).
  - d. phrase search (v ; blank).

#### **How it works**

1. Common operators (AND ; OR ; NOT ; + ; -) can be used.
2. When a particular name is desired, choose the "Phrase search" to make the searching more effective. Conducting a "Search page titles" is a way of targeting a specific web page. All of the operators and "-" "+" signs still apply when "title search" is selected.

#### **Information provided in results index screen**

The name of the engine ; title ; score ; summary ; URL.

### **Metacrawler**

<http://www.metacrawler.com/>

#### **Key features**

1. Coverage : Lycos, Infoseek, WebCrawler, Excite, AltaVista, Yahoo, Thunderstone etc. 7 engines are available.
2. Options for customization (a special software is required for this purpose).
  - a. Default interface (regular ; power ; low bandwidth).
  - b. Default method (any ; all ; as a phrase).
  - c. Results from (everywhere ; North America ; Europe ; Asia ;

- Australia ; South America ; Africa ; Antarctica ; U.S. Educational site ; U.S. commercial site ; U.S.Govt).
- d. Results per source <10 ; 20 ; 30>.
- e. Results per page <10 ; 20 ; 30>.
- f. Timeout <5 ; 10 ; 15 ; 30 ; 60 seconds, 2 minutes>.

### 3. Ranking

Metacrawler combines and normalizes the scores given by each engine and presents them with the scores of a standardized scale.

#### How it works

In addition to the basic “any words” “all words” and “as a phrase” options, users may use a special search syntax (and ; or ; not ; + ; -) to describe the desired results very specifically.

#### Information provided in results index screen

The number of results from each engine Metacrawler contacted is displayed first, followed by the display of items. The information included in each item is : name of the engine, title, summary, URL, rating (on a scale of 1000 points).

### Metafind

<http://www.metafind.com/>

#### Key features

1. Coverage : AltaVista ; Excite ; HotBot ; Infoseek ; Planet-search ; Webcrawler etc. 6 engines are available.
2. Options for customization :
  - a. Wait a maximum of <10 ; 20 ; 30 ; 40 ; 50 ; 60 seconds>.
  - b. <Sort by keywords ; Do not sort ; Sort by engine ; Sort alphabetical ; Sort by domain ; Keywords>.
  - c. Show descriptions <v ; blank>.

#### How it works

Users may use AND, OR, NEAR, NOT, ( ), “ ” in the search. When no connector is inserted, AND is the default one. However, among the six engines, Infoseek does not support Boolean operators, so results from this engine may not be very accurate when operators

are applied in the query.

#### Information provided in results index screen

The name of the engine; title; summary; URL.

#### Profusion

<http://profusion.ittc.ukans.edu/>

#### Special features

1. Coverage : AltaVista, Infoseek, Open Text, Excite, Lycos, WebCrawler, HotBot, Magellan, Yahoo, etc. 9 engines are available. However, users have a variety of options (choose best 3; choose fastest 3; Choose all, Choose manually).

2. Options for customization :

- a. Search mode (Default; Boolean; phrase).
- b. Search in (The Web; Usnet).
- c. Summary option (with summary; without summary).
- d. Check links (0; 5; 10; 15; 20; 25; 30; 35; 40; 45; 50).  
(Although "Check links" is one of the options, users are advised NOT to take advantage of it, because it will make searching less efficient).

#### How it works

Boolean operators—And (&); OR (|); NOT (!); NEAR (~) may be used if "Boolean mode" is selected. Either uppercase letters or symbols can be used. Since Boolean searching is NOT supported in InfoSeek, magellan, Open Text, Yahoo, it is advisable not to choose them when Boolean queries are intended.

Exactmatch phrases may be requested by changing the search mode to "Phrase". Double quotes are not needed while doing phrase searching.

#### Information provided in results index screen

Ranking (based on a scale of 1.0); title; URL; summary (if chosen).



## Savvy Search

<http://williams.cs.colostate.edu:1969>

### Key features

1. Coverage : AltaVista, CSTR, DejaNews, Excite, Einet Galaxy, Four 11, FTP Search, InfoSeek, InReference, Internet Movie, Database, LinkStar, LookUp! Lycos, Megellan, NlightN, Pathfinder, Point Search, Shareware.com, Tribal Voice, WebCrawler, Who where, Yahoo, Yellow Pages, etc. 24 engines are available.

2. Options for customization include :

- a. Search for documents containing {all query terms, as a phrase ; all query terms ; any query term}
- b. Retrieve {10 ; 20 .30 ; 40 ; 50} results from each search engine
- c. Display results in {brief ; normal ; verbose} format
- d. Integrate results

When the option "Integrate results" is selected, duplicate items are eliminated, and results are not separated by search engine. The drawback for this selection is the prolonged response time.

3. There are 14 foreign languages available.

### How it works

Although Boolean operators AND, OR are supported, there is no need to insert them between words. They will be applied automatically depending on the selection "all query term" or "any query term". If "all query terms, as a phrase" is selected, results in which the query terms appear next to each other will be retrieved. Currently operator "NOT" is not supported.

### Information provided in results index screen

"Display options" in the customization is a way to tell Savvy Search how much information you want to be included in the index screen.

Brief format—search engine, title, URL.

Normal format—Author, title, URL, size, score, date, type,



summary.

Verbose format—Besides the information for normal format, timing and other search engine statistics are all included.

#### References

- Hearst, Marti A. "Interfaces for Searching the Web," *Scientific American*, (March 1997) : 68-72.
- Lynch, Clifford. "Searching the Internet," *Scientific American*, (March 1997) : 50-51.
- Lynch, Clifford. "The Internet: Bringing Order from Chaos," *Scientific American*, (March 1997) : 52-56.
- "Metacrawlers and Metasearch Engines," (1998) <http://www.searchenginewatch.com>.
- Petters, Richard, Sikorski, Robert. "Smarter Searching," *Science*, 277 (Aug.15, 1997) : 976-977.
- Rao, Ramana et al. "Rich Interaction in the Digital Library," *Communications of the ACM*, 38 : 4 (1995) : 29-39.
- Schatz, Bruce R. "Information Retrieval in Digital Libraries: Bringing search to the Net," *Science*, 275 (January 1997) : 327-334.
- Scoville, Richard. "Find It on the Net," *PC World*, 14 (January 1996) : 124-130.