

ADDING SEARCH FOR WEBSITE: A STUDY OF REMOTE SEARCH ENGINES

by

Pujar S M *
Mundgod M B **
Munnolli Satish ***

ABSTRACT

Attempt has been made to identify and study various remote search services available on the Internet. Functions, facilities, benefits, limitations and features of remote search engines are briefed. Also discusses experiences with these search engines. Concludes search services are must for surfers for browsing required information efficiently in minimal possible time.

Keywords: Remote Search Engines, Search Services

* Assistant Librarian, Indira Gandhi Institute of Development Research, Mumbai

** Librarian, Gammon India Ltd., Mumbai

*** Librarian, Asian Paints (India) Ltd., Mumbai

0 Introduction

Information on the Internet is growing at such a speed that it is humanly impossible to remember each and everything. To facilitate surfers on the Internet in finding out required information 'Search engines' made their presence. Like search engines for searching different websites, now it has become absolute necessity to add search facility for individual websites. It is because a website or homepage contains different pages, which may range from few pages to hundreds of pages. Secondly, to help visitors of website to find out quickly what they want. Without search facility it is difficult to find required information unless one spends sufficient time. It acts as a 'mirror' for entire information covered in the website, just like a catalogue of library.

Search engines for a website can be added in two ways. First, by using a search software, which can be stored on one's own webserver and secondly, by using remote site search services of others, where in the search software will be stored on their webserver. Here in this paper an attempt has been made to discuss remote search engines.

1 What is a Remote search engine?

It is a tool, which locates information from the website. It works on similar lines of any other search engines-that, indexes a website and builds a database. Only the difference is it does not index entire web like google, altavista etc. When the query is entered in the search box, it checks its index, matches and retrieves relevant piece of information. In other words it acts as a searchable index of a website. Its services are available only to the registered members at cost or without cost depending upon the policy of service providers. Re-indexing of site is done on a regular basis at fixed intervals as scheduled by the member or service provider. Figure-1 shows the working of Remote search engine.

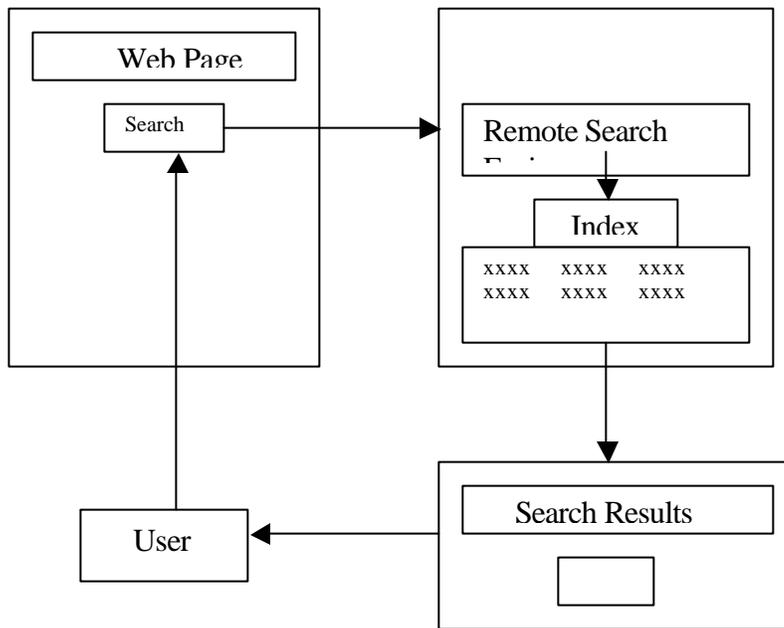


Figure 1

2 Why Remote search engine?

Remote search engines are very easy to work with compared to stand alone search service i.e. by using a search software on own webserver, because it does not involve much technical expertise like CGI programming. It also have many more advantages like index of a website is going to be stored on a remote server causing no load on the server, easy administration, minimal initial investment or no investment at all if the site is small. It is also easy to switch over to other service provider if organization or webmaster is not happy with the services.

2.1 Functions

Like any other search engine, remote search engines too function by using three major elements viz. the spider or a web indexing robot or crawler, the index and a search engine. All these will be stored on a remote server. When search expression is entered into search box of local web page, the local web page sends message to the remote server.

In turn remote server compares the search expression with its index terms, sorts them according to relevance and creates an HTML page with the results. The figure-2 shows the sample of results page. When surfer's clicks on the result link, they will see the page from site. Though the results displayed are of remote services, not of local website, it looks like as if results are displayed from the local website itself.

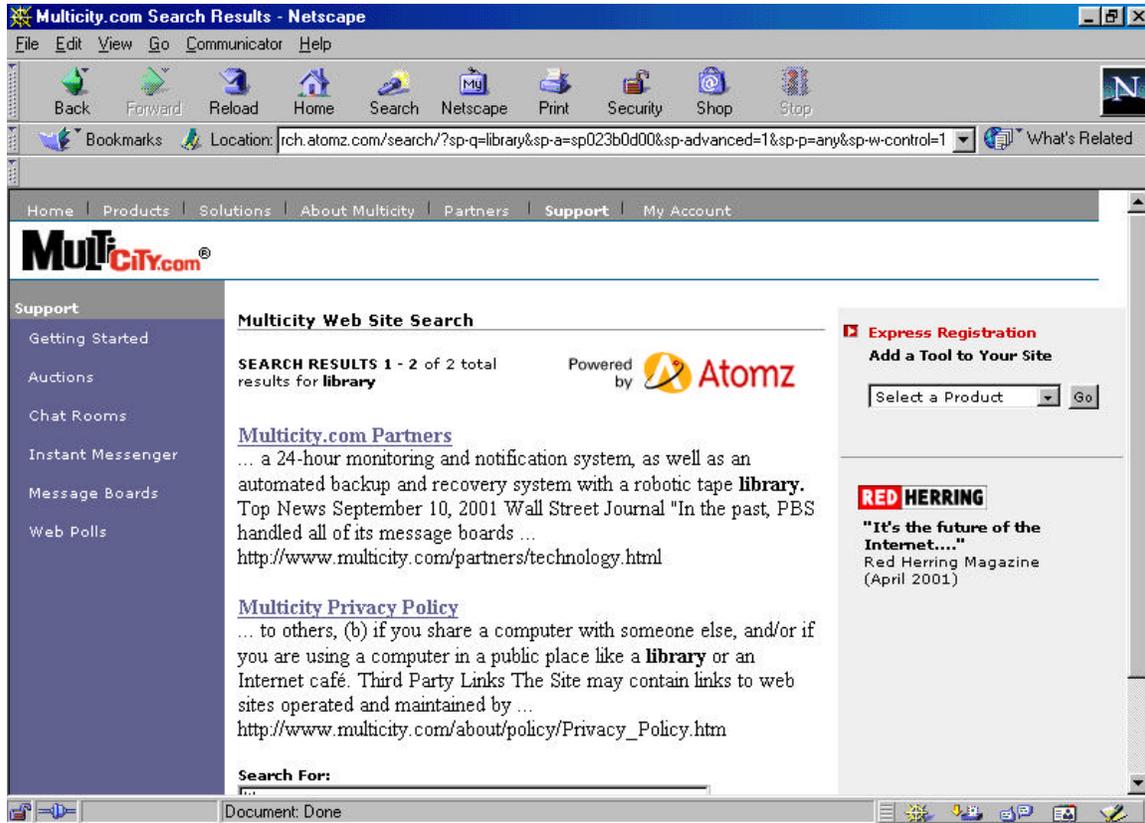


Figure-2 Sample Results page

2.2 Facilities

Remote search engines provide facility for customization of search forms, indexing of pages and results page. The search form usually will be either simple or advanced. Simple search form comes with a text field, search button with an additional option for searching contents from individual pages. Advanced search facility in addition to the above comes with more options like Boolean search, date ranges, etc. The facility of indexing pages gives more flexibility for the webmaster to index only certain portion or pages of the site. The results page customization is very important because of best user experience; the search results page should look and feel like rest of the site. Customization features of this page include page color, background, text and link color and images viz. logo, images, etc. The customization features makes person to feel that results are obtained from the same site which he is searching, though they are not. Some of these remote search engines also send periodic e-mail to members about important

keyword searches done on their site. This helps Webmaster to know the feedback from the surfers and helps him to do the modifications and updations.

2.3 Benefits

The benefits one gets by using remote search engines for their site are many.

- i. There is no need to learn CGI programming or server systems. In other words one never need to learn any software, worry about its compatibility, permissions etc.
- ii. Easy administration. One can administer it from service provider's webpage with username and password.
- iii. No load on organizations own server as search engine requires sufficient disk space and CPU time. Remote search services are usually maintained by highly expert staff with excellent connectivity and round the clock administration
- iv. Minimal initial investment or no investment at all
- v. Easy to switch over to other search engines if organization or webmaster is not happy with the search service

2.4 Limitations

Remote search engines are having following limitations.

- i. Less control over the indexing; if the site changes quite frequently as most of these services will not index site that often
- ii. Advertising on search results page or one must pay for the services every month
- iii. Dependence on outside service. Sometimes it may affect the speed of retrieval if the server is busy
- iv. Less capacity to index as many of these services can index somewhere between 200 to 5000 pages

- v. Provide only few special features like indexing special file formats, adding thesaurus etc.
- vi. Can not be used on intranet as their services need Internet connection and also security of data is at stake
- vii. Multi-site indexing in many cases is not possible

2.5 How to put it on the homepage

It is very simple to add search form on a homepage or website. Remote search engine services provide readymade HTML template of search form on their site. Which needs to be simply 'Copy' from their site and 'Paste' somewhere on the webpage where webmaster likes to put it. Usually it will be put on index or first page of a website. One can also customize it as per the site requirements such as adding logo, background color and so on. Different HTML templates are offered depending upon the type of search facility selected. i.e. simple or advanced.

3 Remote Vs Stand alone Search engine

There is a lot of difference between both remote and stand-alone search engines. Both have their own pros and cons. Following are the major differences between two.

Remote Search Engine	Stand alone/Local Search Engine
Uses remote search engine software to index site	Uses stand alone search software, which could be priced or free
Runs on a outside server	Runs on a local server
Less technical to implement	More technical, needs knowledge of programming
Less control on indexing	More control on indexing
Saves disk space as indexes are stored on a remote server	Disk space required as indexes are stored on a local server
Low up-front cost	More up-front cost
Fewer special features of indexing	More special features of indexing
Advertising on results page in many cases if service is free	No advertising
In many cases multi-site indexing is not possible	Multi-site indexing is possible i.e. indexing of other's sites

4 Remote search engines available on the web

Today good number of remote search engines are available on the web. Some are totally free, some are free for limited number of pages and some are priced. Most of these are

very easy to administer. Some of the important Remote Search Engine services are covered here.

4.1 Atomz [<http://www.atomz.com>]

Atomz is a powerful and customizable tool, and by the same time it is quite impressive too. It indexes site very fast as soon as one register for the service. It even lets exclude folders or specific URL's or pages from the search. It offers both simple and advanced sets of search forms for webmasters to insert into pages. Search of individual sub directories or pages is also possible. Its services are free up to 500 pages and above that one has to pay for the services. No advertisements appear on results page for free service except a small logo of Atmoz. Weekly update or on demand indexing is possible. It is available for both Internet and intranet. Paid version carries extra features and there will be no display of its logo. Other than text it indexes and searches .PDF and .MP3 files. Synonyms and Stop words listings for search is possible. It provides good indexing logs and search reports and also weekly status report via email.

4.2 Freefind [<http://www.freefind.com>]

Freefind offers both free and paid service. Free service has no fixed page limit but advertisements are displayed on search results page. It takes bit more time to index a site. But it does send an email when indexing is finished. It also offers more scheduling options compared to other services. Re-indexing can be scheduled as frequently as possible, but daily re-indexing is not guaranteed for free accounts. Its setup is very easy and allows multiple-site indexing. Its search form provides option to match 'all words' or 'any word'. Easy customization of results page is possible. Also creates automatic site map for the site. Freefind is available only to the Internet accessible web servers. Paid version indexes .PDF files.

4.3 Fusionbot [<http://www.fusionbot.com>]

Fusionbot offers service in five packages. They are free, bronze, silver, gold and platinum. Service is free up to 250 pages on free package with advertisements and remaining services are priced. Automatic or on demand re-indexing is possible on a monthly basis. Basic customization of results page and reporting service is available. It indexes and searches .PDF documents. This service is not available for free accounts. It also generates automatic site map for the site. It is available only to the Internet accessible web servers.

4.4 Google [<http://www.google.com>]

Google site search is a service offered by famous Google search engine. It offers three types of services viz. free, paid and free university program. This service is available only

to those sites, where index is available on 'Google' search engine database. Free version carries advertisement on results page. Simple customization of results page is also possible. Paid version and University program does not carry any advertising on results page. Free University program is available only to the educational institutes. It is available only to Internet-accessible web servers. It too indexes multiple sites. Search is being done using Google syntax. The search form carries Google logo and option of searching site or Internet. It offers minimal search reporting. Indexes different kinds of documents.

4.5 Indexmysite4free [<http://www.indexmysite4free.com>]

It too offers both free and paid service. Up to 500 pages in total with multiple site indexing is free of cost, but advertisements will be placed on search results page. Easy customization of results page is possible and it also indexes and searches different types of documents like .PDF, .DOC, Postscript etc. Re-indexing schedule can be done on weekly or monthly basis. It is available only to the Internet accessible web servers. Masking of subdirectories from indexing is possible.

4.6 Minisearch [<http://www.spacesurfer.com/minisearch/>]

Offers free service with advertisements on results page. It is very easy to administer and provides nice indexing reports. Easy customization of results page is possible. Available only to the Internet accessible web servers. It offers simple search form to search the site. Search results per page can be limited to certain number of records.

4.7 PicoSearch [<http://www.picosearch.com>]

It offers both free and paid service. Service is available in two models viz. free, professional or premium. Indexing up to 1500 pages is free with no advertisements on results page but with Picosearch logo. It indexes multiple file types like .pdf and Ms-Office file extensions, but this facility is not available for free service. Automatic re-indexing too is not available for free accounts. Extensive customization of results page is possible. Only manual re-indexing facility is available. Picosearch is available only to the Internet accessible web servers. Multiple indexing of sites is possible and it offers both simple and advanced search form facilities.

4.8 SiteLevel [<http://intra.whatuseek.com>]

It offers three types of service viz. Personal, Business and Government, educational and non-profit. Its basic service up to 1000 pages is free with advertisements on results page for all the three types. Impressive customization like background color, font type, search box and others are possible. It offers both simple and advanced search forms. Masking of certain URL's or pages from indexing is possible. Re-indexing on demand can be done once a day. Synonyms and Stop words listings for search is possible. It is available only to the Internet accessible web servers. It also offers search reports and statistics.

4.9 SiteSearch

[<http://www.searchengine.com/english/path.cgi?&path=/products/&cl=english&cu=uk>]

SiteSearch offers free service up to 30000 pages with advertisements and logo on results page. Allows manual and automatic re-indexing of site on daily, weekly and monthly basis. Also sends automatic notification after indexing site. Masking of certain pages of site from indexing and impressive customization of results page is possible. It provides excellent reporting of searches, logs and information regarding visitor operating system, browser and IP address. Synonyms and Stop words listings for search and multi site indexing is possible. It can also index password-protected site or pages.

4.10 Webinator [<http://www.master.com/texis/master/app/home.html>]

Webinator is available in two forms viz. software and as a remote search service. Both are available free of cost, but the later is free up to 10000 pages with advertisements on results page. Search form provides both options of searching website and Internet. It is easy to administer and customize. Searching of individual pages or sub directories within a website is possible. Also searches and indexes different file formats like .PDF, Word Perfect and MS-Word documents.

5 Features of Remote Search Engines

Following table shows major features of remote search engines.

Service	Free up to	Advertisement or Logo	Index info	Match words	File Type*
Atomz	500 pages	Logo	Stored/ e-mail report	Any (Can use + and -)	HTML, Text, .PDF and MP3
Freefind	No page limit	Advertisement	Stored report	Choice of any or all	HTML, Text and .PDF
Fusionbot	250 pages	Advertisement	Stored report	All (Can use +, - and *)	HTML, Text and .PDF
Google	No page limit	Advertisement	NO	All (Can use - and """)	HTML, Text, .PDF and MS Extensions
Indexmysit e4free	1000 pages	Advertisement	Stored report	All	HTML, Text, .PDF and Postscript
Minisearch	No page limit	Advertisement	YES	All	HTML and Text

Picosearch	1500 pages	Logo	Stored/ e-mail report	Choice of any or all (Can use +, - and Boolean)	HTML, Text, MS- Word, Excel and Postscript
Sitelevel	1000 pages	Advertisement	Stored report	All	HTML and Text
SiteSearch	30000 pages	Advertisement	Stored report	Choice of any or all (Can use "" and Boolean)	HTML and Text
Webinator	10000 pages	Advertisement	e-mail report	All	HTML, Text, .PDF, MS-Word and Word Perfect

** Indexing and searching of file types .PDF, MS-OFFICE extensions, Word Perfect and Post Script are available in most cases only in paid version*

6 Experiments with Remote Search Engines

In order to get better understanding of remote search engines **Atomz and Google are experimented.** Search responses are tested and results are compared with the remote search engines discussed above, using searchtools.com test search. Search response of most of the sites was very good, but one has to keep in mind that many times it also depends upon the speed of Internet connection.

From the observation it is found that Atomz has many features for indexing, customization and reporting compared with Google. Google allows minimal customization and it does not give any kind of report for searches. Google Site search uses syntax from its popular search engine for indexing sites. Administration of both can be done by using Username and Password. Google free SiteSearch option is a sub-index of Google's main index. Updates are determined by Google and there is no guarantee of pages crawled. Whereas, Atomz updates indexes sites at a fixed interval as scheduled by the member or on demand. It ensures updated pages getting indexed.

Atomz for each of the search results gives topic heading with a link, summary, date and time on which the information is made available on the site, URL address, % of relevance and highlights the search term. It gives few options on results page such as sorting results by date and hiding of summaries. Atomz displays ten records per page with a caption of total number of records retrieved on top and provides search form to make new search.

Freefind provides for each of the search results topic heading with a link along with a serial number, brief description, URL address and highlight the search term. Free find displays ten records per page with a caption of total number of records retrieved on top and provides search form with the options to make new or refine or web search. Also gives links to search tips and site map to help users in searching information.

Fusionbot for each of the search results gives topic heading with a link along with a serial number, brief summary, % of relevance, URL address, date and time on which information is updated and file size. Fusionbot displays ten records per page with a caption of total number of records retrieved with a time on top and provides search form to make a new search. Also provides links to search tips, site map and different search engines like Yahoo, Google, Altavista and Galaxy to search for information using the same search term.

Google site search includes in each of the search results retrieved topic heading with a link, brief description, URL address, file size, link for cache and similar pages, option to convert .PDF or MS-Word or Power Point documents in to HTML. If the document retrieved is .PDF, indicates it by mentioning .pdf in square [] brackets.

Indexmysite4free includes topic heading with a link, abstract, URL address, ranking, table showing matched records and their details, site map and link for more like records for each of the search results. It displays ten records per page with an option display in multiples of ten or all and provides search form to make a new search.

PicoSearch provides topic heading with a link, URL address, brief description and high light the search term for each of the search results. It displays ten records per page and provides search form with an option to do search within results and for doing new search. It also provides link for help.

SiteLevel includes in each of the search results retrieved topic heading with a link, very brief description, URL address, file size, date and time and highlight the search term. Records displayed are ten per page with a caption showing total number of records retrieved. It also provides search form to do new search with an option to search Internet.

Webinator gives topic heading with a link, brief description, % of relevance, file size, link for finding similar records, matching info and show parents and serial number for each of the search results. It also gives an option to sort records by date. It displays ten records page and provides search form to do new search.

7 Conclusion

Search engine for a website is like an index for a book. It is a must for any website having pages. Remote search services are boom for webmasters or organizations, as they spend little or nil for the service. Also they are very easy to administer and implement without having much technical knowledge. As a library professional, it will be more convenient for us as we need not worry much about the technicalities involved in it.

However, by the same time one should have the basic knowledge of HTML to work with remote search services. The main disadvantage with remote search services is that they need Internet connection to function. In this paper thrust is given more on free of cost services provided by remote search service providers.

8 References

1. CHAMPEON (STEVEN). Finding the right search engine. Express Computer. August 27, 2001.
p.16-17
2. DREIER (TROY). Finding: Site search. PC Magazine January 31, 2001.
<http://www.zdnet.com/products/stories/reviews/0,4161,2679429,00.html>
3. PLAT (NINA). Search engines for intranet. <http://www.llrx.com/features/nina.htm>
4. PUJAR (S M) and MUNDGOD (M B). Search engines: A total perspective, presented at First ASIST conference on Internet Resources and Librarianship held in Bangalore between August 12-14, 2001
5. RAPPOPORT (AVI). Adding search to your site.
<http://hotwired.lycos.com/webmonkey/00/09/index2a.html>
6. RAPPOPORT(AVI). Search engines: The hunt is on.
<http://www.networkcomputing.com/1120/1120f1.html>
7. Review of remote search services.
<http://www.searchtools.com/reviews/remotesearch/>