Journal of Scientific & Industrial Research
Vol. 67, October 2008, pp.739-746

# Information retrieval on Internet using meta-search engines: A review

Manoj M* and Elizabeth Jacob
Computational Modelling and Simulation Unit, National Institute for Interdisciplinary Science & Technology, Trivandrum 695019

*Received 13 August 2007; revised 07 July 2008; accepted 10 July 2008*

Meta-search engines (MSEs) on Internet have improved continually with application of new methodologies. Understanding and utilisation of MSEs are valuable for computer scientists and researchers, for effective information retrieval. This paper reviews functionality, working principles, growth, evolution and popularity of various MSEs.

**Keywords:** Collection fusion techniques, Information retrieval, Meta-search engine, Web searching

## Introduction

Though automatic information retrieval[#1] (IR) existed before World Wide Web[1] (WWW), post-Internet era has made it indispensable. IR is sub field of computer science concerned with presenting relevant information, gathered from online information sources to users in response to search queries[2]. Various types of IR tools have been created, solely to search information on Internet. Apart from heavily used search engines (SEs), other useful tools are deep-web search portals, web-directories and meta-search engines (MSEs). Among various IR tools[3] available, most people use SEs to locate web pages. Using SEs, an index is searched rather than entire Web. Index is created and maintained through ongoing automated web searching by programs commonly known as spiders. Web-directories are databases of web sites compiled and maintained by humans. Since web-directory content is hand picked by humans, results have high accuracy but a typical web-directory's index size will be only a fraction of that of a SE and content can easily become outdated. Open-Directory-Project (ODP)[#2] is biggest web-directory available today. Deep-Web or Invisible web[4] consists of publicly accessible pages with information in databases such as catalogues and references on WWW that are not indexed by SEs. Much research has emerged[5] in recent years on information in invisible web sites. A MSE (multi-threaded SE) such as Dogpile, SavvySearch, Metacrawler, Profusion and Inquirus, is a search tool that sends a query simultaneously to several SEs and consolidates all results, thereby saving time.

This paper reviews working of MSEs and their significance in IR on Internet.

*Author for correspondence
E-mail: www.manoj.m@gmail.com

## Meta Search Engines (MSEs)

### Search Engines (SEs)

MSEs are derived from general SEs. SE indexes contents of entire WWW. Vertical SEs or speciality SEs[6] alleviate problems of precise IR on web to some extent by using focused databases in a particular knowledge domain, which cannot be applied to general topics. There are many vertical SEs (CiteSeer, eBizSearch, IceRocket, FindSounds, FileSearching and WebSeer). A spider (crawler or robot) explores hyper-linked documents of web, searching and gathering web pages to index. Index and copy of documents themselves are stored in a database. SE accepts a query and creates a list of links to web documents matching query and presents it to the user. Though this logical architecture is basically same for all SEs, almost all modern SEs use computer cluster and massive parallelism to handle heavy loads and to provide fail-safe functionality.

Since birth of modern Internet in early 1990s, need for IR led to growth, dominance and death of various SEs[7,8] like Wandex, Aliweb, Excite, webcrawler, Lycos, AltaVista, Inktomi, AskJeeves and Northern Light. Majority (80%) of Internet users are hooked on to three engines-Google, Yahoo and MSN-Search[9]. Google[10], developed in Stanford University around 1998, used concept of link popularity and PageRank[11] as its main-

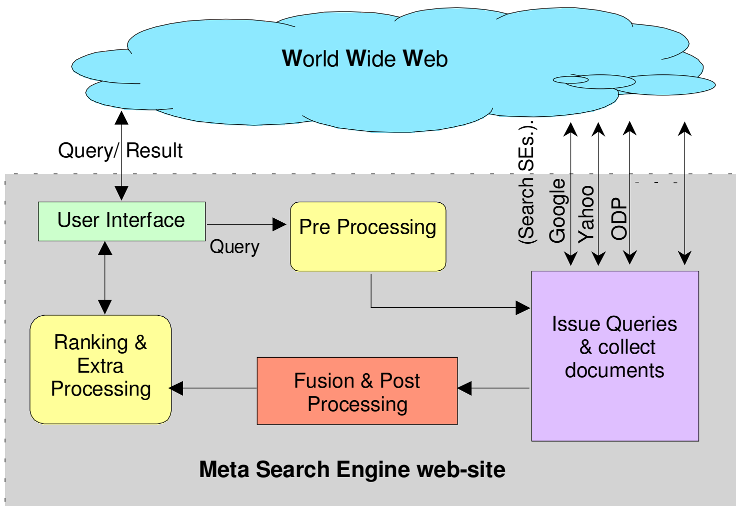


Fig. 1 — Architecture of a typical meta-search engine

ranking algorithm. Yahoo[#3], launched in 1994 by Stanford University, started out as a listing of personal favourite web sites with URL and description of each page. MSN-Search (Live Search) is an SE owned by Microsoft, launched in 1999 and was powered by results from Looksmart and Inktomi till 2004, after which it uses its own crawler-based index.

***Limitations of Search Engines (SEs)***

Size of Internet[12] in terms of data continues to grow exponentially. No single SE indexes more than about one-third of 'indexable web', and combining results of 6 SEs yields about 3.5 times as many documents on an average as compared to the results from only one engine. SEs do not index sites equally and no engine indexes more than about 16% of web[13,14]. Major engines index less than half of web and average overlap between engines is very small[15,16] (<1.4% of total coverage). Indexable web is approx. 11.5 billion pages. Intersection of Google, MSN, Ask/Teoma and Yahoo indexes is 28.85%, or about 2.7 billion pages, and their union is about 9.36 billion pages[17]. Even if 2 SEs use same databases, search results may vary, because each SE uses its own ranking algorithm. Deep (non indexable) web often contains large amounts of data[5], whose major part is not available through traditional SEs. A combination of retrieval paradigms[18] brings improvements in IR results. Coverage limitations, non-uniform user interfaces, query limitations and duplicates lower effectiveness of SEs. This has led to the development of MSEs.

**Development of Meta-Search Engines (MSEs)**

MSEs are useful for all web-surfers. By searching several databases at the same time, MSEs find more hits for a single query. MSEs provide integration of results from different SEs, comparison of rank positions and advanced search features on top of commodity SEs like, clustering, question answering and personalized results. MSEs use other IR tools (web-directories, search engines, deep-web search tools) in real-time to conduct web-search. By using an MSE, one gets a snapshot of top results from a variety of IR tools.

MSEs can be classified as real, semi-pseudo, pseudo and client-side. Real MSEs (MetaCrawler) appear like a traditional SE and work in the server. Semi-pseudo MSEs (Better Brain) queries multiple SEs and present results grouped by engine in a scrollable easy to read list. Pseudo MSEs (NetDepot) open multiple SE pages simultaneously in multiple browser windows/frames. In a client-side MSE (Copernic), components reside at user's machine. Semi-pseudo and pseudo MSEs don't process any data. Though client side MSEs have advantages like moving all resource usage to client machines; it is usually outweighed by disadvantages like the need for frequent updating and client software installation.

***Working of Meta-Search Engines (MSEs)***

MSEs (Fig. 1) perform following tasks sequentially: i) accept a user query; ii) query processing; iii) launch multiple queries; iv) collect and merge results (collection fusion); and v) present post-processed results to the user.

*Querying*

MSEs build query interface as a static HTML page. In information retrieval dialogue, expression of users' information needs is a self-refining process achieved by a progressive query interface. Starting page usually consists of some common controls like Least-Common-

Denominator (LCD) interface and during user query process, query pages change according to users' needs. Apart from input query, many MSEs have advanced search options such as selecting search list of needed search sources from a given list, depth of search, maximum allowable time for search in IR tools etc.

*Collection Fusion of Results*

Merging of results collected by MSE is called collection fusion[19], which combines retrieval results from multiple, independent collections into a single result such that effectiveness of combination approximates effectiveness of searching entire set of documents as a single collection. Collection fusion is stated as:

Given a query Q, information servers $I_1, I_2, ..., I_C$ and N number of documents to be retrieved, find values of $\lambda_1, \lambda_2, ..., \lambda_C$ such that $\sum_{i=1}^{C} \lambda_i = N$ and $\sum_{i=1}^{C} F_Q^{I_i}(\lambda_i)$ is maximum, where $F_Q^I(S)$ is distribution of relevant documents in retrieved set, which is a function of the number of retrieved documents S. To approximate $F_Q^I(S)$, it is assumed that all collections have an equal number of relevant documents. Another approach is to select cut-off values such that documents with N greatest similarities across all collections are retrieved provided similarity measures are comparable.

Relevant document distribution (RDD) of a query Q is modeled by averaging RDDs of the k most similar training queries, which have highest cosine similarity with Q. A maximization procedure finds value for $\lambda_i$ for each collection that maximizes number of relevant documents retrieved. Within each collection, $\lambda_i$ documents are ordered. Query clustering fusion strategy uses query vectors formed from training queries. Topic areas are represented as centroids of query clusters. Training phase assigns to a cluster a weight, which measures effectiveness of queries in cluster on that collection. For new queries, cluster whose centroid vector is most similar to query vector is selected and its weight returned. Set of weights returned by all collections is used to apportion retrieved set. $(w_i / \sum_{i=1}^{C} w_i) * N$ documents are retrieved from collection *i*, where $w_i$ is weight returned by $i^{th}$ collection.

In collection fusion results from disjoint sets or independent search indexes are merged. In data fusion[2], results produced by different formulations of query in the context of a single search index are merged. Some of the data fusion techniques could be applied to improve collection fusion[20]. Post-processing like duplicate removal or other advanced processing is done with the results. Simple or advanced re-ranking based on a variety of methods is done and result is presented to user like any SE. Result presentation may be as in the form of a traditional SE or with multiple views, clustering of similar documents etc. Attitude of a typical SE towards an MSE is generally of three types: i) Cooperative, when an SE may provide special interface or even private access to resources such as their index database or at least provide notifications of interface changes; ii) Non-cooperative but not hostile, when an SE neither performs any blocking nor provides any special access mechanisms for MSE; and iii) Hostile, when an SE forbids access by MSEs usually through technical methods and in extreme cases through legal remedies.

***Architecture and Underlying Technology of Meta-Search Engines***

Technology evolution of MSEs has been traced by means of a time-line (Fig. 2). MetaCrawler[21,22] (University of Washington), probably one of the first MSEs, provided a single interface to allow users to search simultaneously, from different SEs and used a relatively simple mechanism to combine results from different SEs, eliminating duplicate URLs and merging results in an interleaving fashion. MetaCrawler[23] used a linear combination based scheme called Normalize-Distribute-Sum algorithm. Mamma MSE[#4] (Carleton University) uses rSort ranking, which considers each duplicate result as a vote and ranks the result set accordingly.

Search[#5] MSE uses SavvySearch metasearch technology (Colorado State University) to give results. SavvySearch[24-26], employs a meta-index for selecting relevant SEs based on the terms in user query. Query processing in SavvySearch proposes a search plan of target SEs as different ordered steps that user can either select or disregard. SavvySearch uses links visited by users to improve its meta-index. Dogpile[#6] (Thunderstone Software Inc.) is another popular MSE.

ProFusion[27] (University of Kansas) supports both manual and automatic query dispatch and analyses incoming queries, categorizes, and automatically picks best SEs for query, based on apriori knowledge (confidence factors). It uses these confidence factors to merge search results into a re-weight list of returned documents, removes duplicates, optionally removes broken links, and presents final rank-ordered list to user. Later modifications[28] include autonomously adaptive

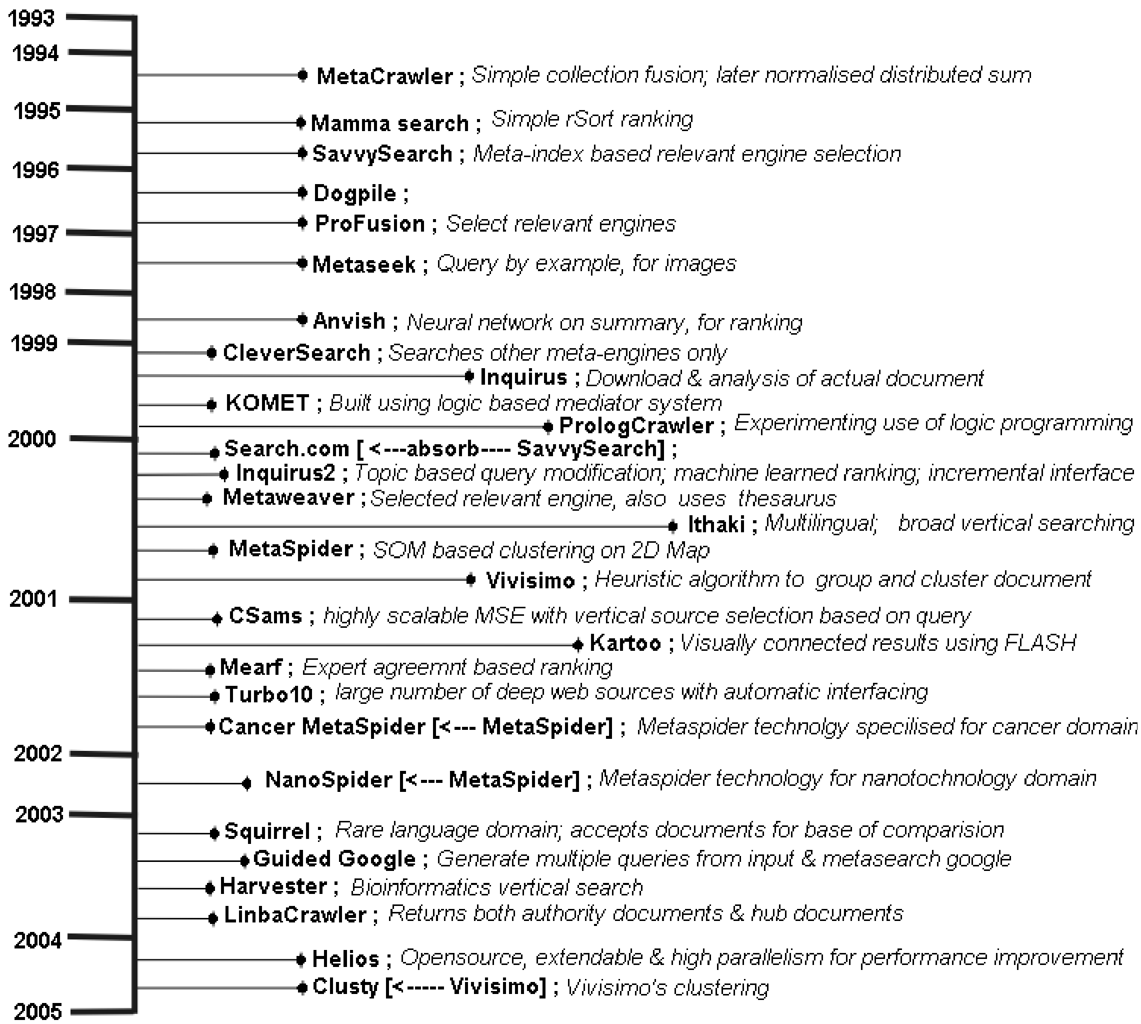


Fig. 2 — Technology evolution of meta-search engines

multi-agent implementation consisting of different types of agents for query dispatch, search and learning.

Metaseek[29,30] (Columbia University) is a content-based MSE used for finding images on web, based on visual information. Metaseek targets query by example as opposed to keyword search. It intelligently selects interfaces of multiple on-line image SEs by ranking their performance for different classes of user queries. Anvish[31,32] (Louisiana State University) uses a CC4 (corner classification) neural network[33,34] based document classification algorithm. It uses title and short summary that are commonly supplied for each document by Internet SEs. It assumes that top two documents returned by each search are relevant. Based on keywords in titles and summaries of relevant documents, it looks for similar documents to mark as relevant.

KOMET MSE[35] (University of Karlsruhe) was built using KOMET (Karlsruhe Open MEdiator Technology) system, a logic-based mediator shell that uses declarative language KAMEL (KArlsruhe MEdiator Language), which is based on annotated logic. It provides a framework for easy construction of mediators, and enables reuse of existing code. PrologCrawler[36] (University of Pisa), developed in Prolog, was aimed at investigating and experimenting usability of logic programming for developing web applications. CleverSearch[#7], a Real Meta-Metasearch Engine, claimed to gather data from major MSEs like MetaCrawler, MetaFind, Mamma, Surfy, Dogpile, SavvySearch, HuskySearch, AskJeeves, CNN and ProFusion. It sorts results on the basis of number of occurrences for given SEs. In 2004, CleverSearch was discontinued.

Inquirus[37] (NEC Research Institute) took a different approach in addition to getting a list of URLs and summaries returned by other SEs; it fetches and analyses actual documents in result set, instead of relying on engine's ranking mechanism. Inquirus also performs result clustering, and query expansion, offering better ranking at the cost of scalability. This approach can filter pages, which are no longer available (dead links) or do not contain search terms. Image metasearch[37] functions were later reported where text analysis of document is used to find relevance of image. Later additions to Inquirus[38-40] (Inquirus2) include topic-geared query modification, machine learned classifiers and ranking incorporated in an incremental user interface.

Personal Search Assistant[41] (PSA) is a client side MSE (University of Hyderabad). PSA can run as a background process in user's machine. Agent running in background should help users to reduce the amount of time spent on a search. In this case, profile is supplied to an agent that can automatically gather information on behalf of the user. A user can view results retrieved and manually select the ones to be stored in database. Metaweaver[42] (Tokyo Institute of Technology), like ProFusion and SavvySearch, uses an adaptive system from previous search experience to select most relevant SEs for current query of user. It evaluates specialties of each SE by analyzing hit list for sample queries, and to utilize them for selecting adequate SEs for user queries. Unlike other similar MSEs, it effectively utilises thesaurus. Ithaki[#8], named in honour of Greek poet's poem Ithaki, can search in 14 languages besides English. It covers major SEs and is capable of topic specific meta-vertical searching on news, images, mp3 audios etc. Vivisimo[#9] MSE (Carnegie Mellon University) uses a specially developed heuristic algorithm to group or cluster textual documents.

MetaSpider[43] (University of Arizona) is a client side MSE integrated with textual clustering. It connects to 6 general-purpose SEs, and combined search results are validated. After verifying content of returned web pages, using Arizona Noun Phraser, all noun phrases are extracted from each document based on part-of-speech tagging and linguistic rules. A self-organising map is used to automatically cluster web pages in real-time into different regions on a 2-D map to give user a graphical overview of whole document set. Vertical MSEs like MedSpider specialized in medical domain, NanoSpider[44] in nano technology domain, –Cancer Meta spider[45] for cancer resources on Web were developed from MetaSpider by applying domain specific modifications.

CSams[46] (Computer Science Academic Metasearch engine) is a MSE (Binghamton University) prototype implementation aimed at creating a highly scalable MSE capable of selectively searching multiple domain-specific SEs based on user query. Mearf[47] (University of Minnesota) uses its own ranking method based on expert agreement on particular query. Mearf is implemented based on answers for which different experts agree on are more likely to be relevant than answers for which there is little agreement among experts.

Kartoo[#10], a visual MSE, helps to see search results visually interconnected by keywords, accomplished through the use of Macromedia Flash. Turbo10[48] (Fleetfoot Internet Solutions) is capable of searching a large number of deep-web sources. It automates process of creating and maintaining software adapters that connect to, search, and extract results from different SEs. Precise image metasearch engine[49] is a client side MSE. It uses fuzzy logic rules and a neural network in order to provide an additional search service for human photos on Web. Squirrel[50] (Uppsala University) is a client side MSE, which collects text material in lesser-used languages. It produces adequate and up-to-date learning materials for learners. In a later upgrade, Squirrel-2[#11] (server side version with rudimentary web-interface) user can indicate a model document to serve as origin of comparisons, and when there are no documents, documents get ordered by increasing difficulty.

Guided Google[51] (University of Melbourne) is a client-side MSE built using Google web-services. The functionalities provided in this engine are based on combinational keyword searching. Harvester[52] (European Molecular Biology Laboratory) is a bioinformatics MSE for genes and protein-associated information. Harvester crosslinks many popular bioinformatics resources and allows cross searches. A ranking system similar to Google page-rank sorts search results and displays more relevant information. ProThes[53,54] is a client-side MSE. It has a graphical user interface for query specification, and a thesaurus-based query customisation system. It employs domain-specific knowledge, represented by a conceptual thesaurus. It also provides simple heuristics for result merging and partial re-ranking. LinbaCrawler[55] (Texas A&M University) returns not only authority documents, which contain relevant information of a user query, but also hub documents, which have links pointing to many relevant documents and may cover different sub-topics of query. Helios[56] (University of Pisa) is an open source MSE used to design more sophisticated Web IR tools. It uses high parallelism to reduce document-collecting time. MSE, Snaket (snaket.di.unipi.it), which uses Helios

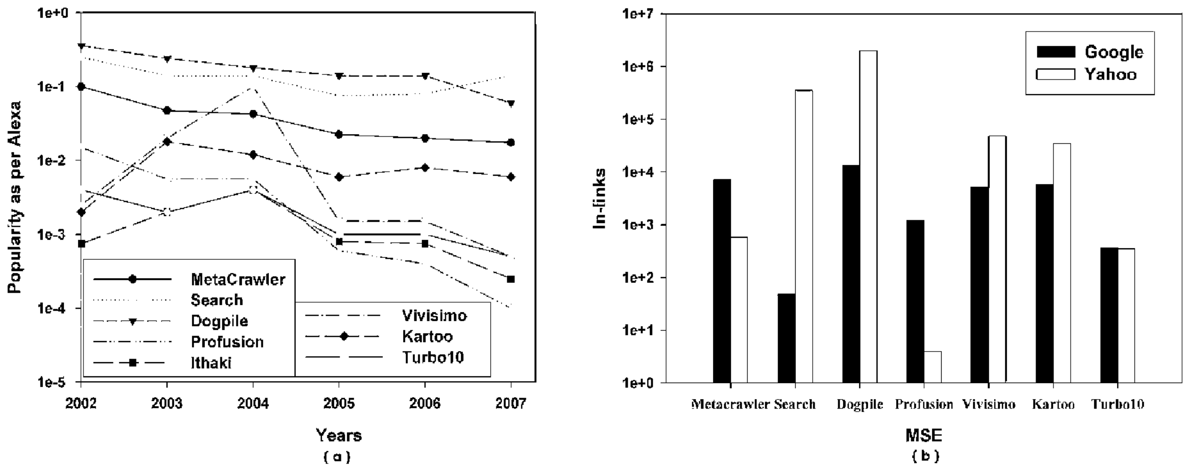


Fig. 3 — Popularity statistics: a) From Alexa; b) In-links to MSEs as found by Google and Yahoo

to search web, offers two complementary views on its returned results - the classical ranked list and a hierarchical organization of results into labelled folders created on-the-fly at query time.

In 2001, it was estimated that there were more than 1200 MSEs available on web[40] but for many of them, technical details are not publicly available[47]. Some of the functional MSEs[#12] available for use today are Dogpile, Search, Mamma, Kartoo and Vivisimo.

***Comparison of Meta-Search Engines***

*Performance Reports*

Standard criteria[23] for evaluating IR systems are precision and recall. Precision is the ratio of number of correct documents identified by SE to total number of documents identified by SE. Recall is the ratio of number of correct documents identified by SE to number of all correct documents in collection. Manual pick version of ProFusion[27] outperformed MetaCrawler (0.35) and SavvySearch (0.47) with a precision of 0.56 when evaluated by top 20 retrieved documents from 12 independent queries. In a pairwise t-test[43], average precision rate of MetaSpider (0.816) ranked highest followed by MetaCrawler (0.697) and Northern Light (0.561). MetaCrawler gave highest average recall rate (0.331) in comparison with MetaSpider (0.308) and NorthernLight (0.203).

It is difficult to identify all relevant documents from the entire collection. Some MSEs[26,29,30] use an alternate evaluation of visit and no-result. A visit event occurs when a user follows one of the links suggested by a SE. A no-result event occurs when one of the queried SEs finds no relevant documents. To study effects of learning in SavvySearch, a minimal meta-index was allowed to accumulate experience over a 28-day period. SavvySearch improved performance on both visits and no-results; visits averaged 0.36 in first 5 days and 0.42 in last 5 days. No-results for same days averaged 0.142 and 0.135 respectively. After study, most used term (5000 usages) had visits of 0.76 and no-results of 0.08.

Downloading sequentially[56], 100 results from 6 SEs (Altavista, Gigablast, Google, Looksmart, Ask/Teoma and Yahoo) as test bed engines, a total of 600 search results required 12.4s. This test was done using wget, the http retrieval tool. Using Helios MSE, same results were retrieved in parallel in less than 4.6s. Many other studies[36,47,48,51,53] reported query modification, result improvement, evaluation of various fusion algorithms etc. Re-conducting experiments or conducting new experiments on all functional MSEs available may produce results different from those reported earlier, due to rapid growth of Internet and dependency of MSEs on SEs.

*Popularity Statistics*

To study popularity of different MSEs, Alexa[#14] Internet web-service, a subsidiary of Amazon.com, which provides information on web traffic to web sites, has been used. Alexa ranks sites based on visits from users of its Alexa Toolbar for various web browsers. Alexa report is based on statistical sampling (random sample) of Internet users. Only those MSEs with second level domain name (search.com, Ithaki.net) have been evaluated. It is observed that Dogpile followed by Search and MetaCrawler are the most popular MSEs (Fig. 3a). Number of pages that link to an MSE shows its popularity (in-link rank). In-link ranks of MSEs reported by Google and Yahoo as in January 2008 (Fig. 3b) can be obtained by querying 'link:*domain-name-of MSE*' for Google and site-explorer[#15] for Yahoo. Yahoo ranks Dogpile at the top followed by Search but Google finds MetaCrawler, Vivisimo and Kartoo as close seconds to Dogpile. The discrepancies in results of Google and Yahoo may be due to less overlap between their indexes. Popularity and link data are given in logarithmic (base 10) scale. These statistics are based on experiments carried out at a particular time and using random sampling. Due to dynamic nature of web and that values are being statistically estimated, these figures reflect only general trend of usage of different MSEs.

## Conclusions

Analysis and implementation of collection fusion in MSEs reported, seems to be promising in improving IR. Research into MSEs has been mainly focused on source engine selection, re-ranking and integration of multiple SE's results. Apart from design and evaluation of full MSEs, research in distributed search, database selection and result merging combination techniques have a direct bearing on the domain of meta-searching. Recent smarter meta-search technology includes clustering and linguistic analysis that attempts to show themes within results, textual analysis and display that can help to dig deeply into a set of results. Rapid growth and evolution of web is posing new challenges to MSEs. Emergence of Web 2.0 services[57] like social-networking, weblogs, RSS feeds, and increase in non textual information like podcasing, online-videos[#13] and convergence of non-conventional forms of communication such as mobile phones with Internet will require easy IR in these areas.

## References


1 Gilchrist A, Text retrieval: an overview, *Learned Publishing*, **16** (2003) 61-69.

2 Montague M, *Metasearch: Data fusion for document retrieval*, Ph D thesis, Dartmouth College, Hanover, NH, 2002.

3 Barker J, Finding Information on the Internet: A Tutorial (Types of Search Tools), in *UC Berkeley -Teaching Library Internet Workshops* (University of California, Berkeley) 2004; http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/ToolsTables.html.

4 Bergman M K, The deep web: surfacing hidden value, *J Electron Publish*, **7** (2001); http://www.press.umich.edu/jep/07-01/bergman.html.

5 Ru Y & Horowitz E, Indexing the invisible web: a survey, *Online Inform Rev*, **29** (2005), 249-265.

6 Chau M, Spidering and filtering web pages for vertical search engines, in *Proc Americas Conf on Information Systems* (AMCIS 2002 Doctoral Consortium, Dallas, Texas) 2002.

7 Underwood L, A Brief History of Search Engines, 2004; http://webreference.com/authoring/search_history.

8 Pinkerton B, *WebCrawler: Finding what people want*, Ph D Thesis, University of Washington, 2000.

9 Sullivan D, Nielsen NetRatings: search engine ratings, August 22, 2006; http://searchenginewatch.com/showPage.html?page=2156451.

10 Brin S & Page L, The anatomy of a large scale hypertextual web search engine, in *Proc 7th World Wide Web Conf* (Elsevier Science, Amsterdam) 1998, 107-117.

11 Tsaparas P, *Link analysis ranking*, Ph D thesis, University of Toronto, 2004.

12 Hosts on the Internet, ISC Internet Domain Survey, Network Wizards, Internet Systems Consortium, 2005; http://www.isc.org/ops/ds/reports/2005-01/report.html.

13 Lawrence S & Giles C L, Searching World Wide Web, *Science*, **280** (1998) 98-100.

14 Lawrence S & Giles C L, Accessibility of information on the web, *Nature*, **400** (1999) 107-109.

15 Bharat K & Broder A, A technique for measuring relative size and overlap of public Web search engines, in *Proc 7th Int Conf on World Wide Web 7* (Brisbane, Australia) 1998, 379-388.

16 Sander-Beuermann W & Schomburg M, Internet Information Retrieval: The Further Development of Meta-Searchengine Technology, in *Proc 1998 Internet Summit, Internet Society* (Genf. Institute for Computer Networks and Distributed Systems, University of Hannover, Germany) 22-24 July 1998.

17 Gulli A & Signorini A, The indexable Web is more than 11.5 billion pages, in *Poster Proc 14th Int Conf on World Wide Web* (ACM Press, Chiba, Japan) 2005, 902-903.

18 Shaw J A & Fox E A, Combination of Multiple Searches, in *Proc. 3rd Text Retrieval Conf (TREC-3)* (Gaithersbirg, Maryland) 1994, 105-108.

19 Voorhees E M, Gupta N K & Johnson-Laird B, The collection fusion problem, in *Proc. 3rd Text Retrieval Conf (TREC-3)* (Gaithersbirg, Maryland) 1994, 95-104.

20 Ng K B & Kantor P P, An investigation of the preconditions for effective data fusion in IR: a pilot study, in *Proc of the 61 Meeting of the American Society for Information Science*, 1998.

21 Selberg E & Etzioni O, Multi-Service Search and Comparison using the Metacrawler, in *Proc 4th World Wide Web Conf Boston* (MA, USA) 1995, 195-208.

22 Selberg E & Etzioni O, The MetaCrawler architecture for resource aggregation on Web, *IEEE Expert*, **12** (1997), 11-14.

23 Selberg E W, *Towards comprehensive web search*, Ph D thesis, University of Washington, 1999.

24 Dreilinger D E, *Description and evaluation of a meta-search agen*t, Master's thesis, Computer Science Dept., Colorado State University, 1996.

25 Dreilinger D & Howe A, Experiences with selecting search engines using meta-search, *ACM Transact Informa Syst*, **15** (1997) 195-222.

26 Howe A E & Dreilinger D, SavvySearch: A meta-search engine that learns which search engines to query, *AI Mag*, **18** (1997) 19-25.

27 Gauch S, Wang G & Gomez M, Profusion: Intelligent fusion from multiple, distributed search engines, *J Univ Compu Sci*, **2** (1996), 637-649.

28 Fan Y & Gauch S, Adaptive agents for information gathering from multiple distributed information sources, in *Proc 1999 AAAI Spring Symp on Intelligent Agents in Cyberspace* (AAAI Press, Stanford University) 1999.

29 Benitez A B, Beigi M & Chang S F, A content-based image meta-search engine using relevance feedback, *IEEE Internet Computing*, **2** (1998) 59-69.

30 Beigi M, Benitez A B & Chang S F, Metaseek: A content-based Meta search engine for images, in *Proc IS&T/SPIE Storage and Retrieval for Image and Video Databases* (San Jose, CA) 1998, 118-128.

31 Kak S, Better web searches and prediction with instantaneously trained neural networks, *IEEE Intell Syst*, **14** (1999) 78-81.

32 Kak S & Shu B, A neural network-based intelligent metasearch engine, *Inform Sci*, **120** (1999) 1-11.

33 Enhong C, Shangfei W, Zhenya Z & Xufa W, Document classification with CC4 neural network, in *Proc of ICONIP 2001* (Sanghai, China) 2001.

34 Enhong C, Zhenya Z, Kazuyuki A & Xufa W, An Extended corner classification neural network based document classification approach, *J Software*, **13** (2002) 871-878.

35 Calmet J & Kullmann P, Meta web search with KOMET, in *Proc of IJCAI-99 Workshop for Intelligent Information Integration* (Stockholm, Schweden) 1999.

36 Bolognesi E & Brogi A, A Prolog Meta-search engine for the World-Wide Web, in *Proc Inter Workshop on Distributed and Internet Programming with Logic and Constraint Languages* (Las Cruces, USA) 1999.

37 Lawrence S & Giles C L, Inquirus, the NECI Meta search engine, in *Proc 7th Int Conf on WWW* (Brisbane, Australia) 1998, 95-105.

38 Glover E J, Lawrence S, Birmingham W P & Giles C L, Architecture of a Metasearch engine that supports user information needs, in *8th Int Conf on Information and Knowledge Management (CIKM'99)* (Kansas City, Missouri, United States) 1999, 210-216.

39 Glover E J, Lawrence S, Gordon M D, Birmingham W P, Giles CL, Recommending web documents based on user preferences, in *ACM SIGIR 99 Workshop on Recommender Systems* (ACM Press, Berkeley, CA) 1999.

40 Glover E J, *Using extra-topical user preferences to improve web-based Meta search*, Ph D thesis, University of Michigan, 2001.

41 Kaushik P R & Murthy K N, Personal Search Assistant: a configurable personal Meta search engine, in *5th Australian World Wide Web Conf* (Southern Cross University, Lismore, Australia) 1999.

42 Mori M & Yamada S, Adjusting to specialties of search engines using metaweaver, in *WebNet 2000 World Conf on WWW and Internet* (San Antonio, TX) 2000, 408-412.

43 Chen H, Fan H, Chau M & Zeng D, MetaSpider: Meta-searching and categorization on the Web, *J Am Soc Inform Sci Technol*, **52** (2001) 1134-1147.

44 Chau M, Chen H, Qin J, Zhou Y, Qin Y, Sung W-K, & McDonald D, Comparison of two approaches to building a vertical search tool: A case study in the nanotechnology domain, in *Proc JCDL 2002* (Portland, OR, USA) 2002, 135-144.

45 Chen H, Fan H, Chau M & Zeng D, Testing a cancer meta spider, *Int J Human-Compu Stud*, **59** (2003), 755-776.

46 Zonghuan Wu, Weiyi M, Clement Yu & Zhuogang Li, Towards a highly-scalable and effective Meta search engine, in *Proc 10th Int Conf on World Wide Web* (Hong Kong) 2001, 386-395.

47 Oztekin B U, Karypis G & Kumar V, Expert agreement and content based reranking in a Meta search environment using Mearf, in *Proc 11th Int World Wide Web Conf* (Honolulu, Hawaii, USA) 7-11 May 2002.

48 Hamilton N, Mechanics of a deep net Metasearch engine, in *Proc 12th Int World Wide Web Conf* (Budapest, Hungary) 2003.

49 Anagnostopoulos I, Anagnostopoulos C, Psoroulas I, Loumos V & Kayafas E., An hybrid Meta search for precise image search in web, in *25th Int Conf on Inform Technol Interfaces (ITI) 2003* (Cavtat/Dubrovnik, Croatia) 2003, 375-381.

50 Nilsson K, *A meta search approach to locating and classifying reading material for learners of Nordic languages*, Masters Thesis, Computational Linguistics, Dept. of Linguistics, Uppsala University, 2003.

51 Hoong D C & Buyya R, Guided Google: A Meta search engine and its implementation using the Google distributed web services, *Int J Compu Applications*, **26** (2004) 181-187.

52 Liebel U, Pepperkok R & Kindler B, Harvester, a fast meta search engine of human protein resources, *Bioinformatics*, **20** (2004) 1962-1963.

53 Braslavski P, Alshanski G & Shishkin A, ProThes: thesaurus-based meta-search engine for a specific application domain, in *Proc 13th Inter World Wide Web Conf on Alternate track papers & posters* (New York, NY, USA.) 19-21 May 2004.

54 Braslavski P, Shishkin A & Alshanski G, 3 in 1: Meta-search, Thesaurus, and GUI for focused web information retrieval, in *Proc 6th Nat Russian Res Conf on Digital Lib (RCDL2004)*, (Pushchino, Russia) 29 Sept-1 Oct 2004, 135-140.

55 Li L, Zhang W, Shang Y & Liu Y, LinbaCrawler: A linkage-based meta search engine, in *Inter Conf on Intelligent Agents, Web Technology and Internet Commerce - IAWTIC'2004*, (Gold Coast – Australia) 12-14 July 2004.

56 Gulli A & Signorini A, Building an Open Source Meta Search Engine, *Proc of 14th Inter World Wide Web Conf*, Chiba, Japan, 2005, 1004–1005.

57 Barsky E, Introducing Web 2.0: weblogs and podcasting for health librarians, *J Can Health Libr Assoc*, **27** (2006) 33-34.

## Web URLs

#1 http://www.answers.com/topic/information-retrieval
#2 http://laisha.com/zine/odphistory.html
#3 http://docs.yahoo.com/info/misc/history.html
#4 http://www.mamma.com
#5 http://www.seach.com (SavvySearch.com)
#6 http://dogpile.com
#7 http://cleversearch.hypermart.net
#8 http://ithaki.net
#9 http://vivisimo.com
#10 http://www.kartoo.com
#11 http://phon.joensuu.fi/nodalida/abstracts/28.shtml (Squirrel-2)
#12 http://dmoz.org/Computers/Internet/Searching/Metasearch/
#13 http://video.google.com
#14 http://www.alexa.com/
#15 http://siteexplorer.search.yahoo.com