

ENTERPRISE INFORMATION RETRIEVAL: A SURVEY

Hamid Turab Mirza

*Department of Computer Engineering, College of Electrical and Mechanical Engineering
National University of Science and Technology, Rawalpindi, Pakistan*

Keywords: Information Extraction, Corporate Search, Knowledge Management, Document Classification.

Abstract: Efficient retrieval of the relevant information is a critical success factor for many enterprises. Despite of all the advancement in the web search technology, enterprise searching is still faced with many challenges and problems. Boundaries of the enterprise search are broad and expectations of the users are quite high, in addition to many challenges faced one of the major problems is the difference between the nature of web and enterprise searching. Many solutions have been proposed and techniques have been devised to improve the enterprise search, but still effective enterprise searching is a challenge for the researchers and the commercial companies, however it is realized that the solution for which will deliver enormous benefits.

1 INTRODUCTION

Since nineteen fifties research efforts have lead to effective ways for retrieval of relevant documents from homogeneous collection of text, such as newspaper archives, scientific abstracts, and CD ROM encyclopedias, however latter on in nineteen nineties there was a major paradigm shift and efforts were made to deal with the issues posed by enormous scale of data, great heterogeneity, unfettered interlinking, democratic publishing, the presence of adversaries and most of all the diversity of purpose for which web search may be used (Hawking, 2004; Baeza-Yates and Ribeiro-Neto, 1999).

Over the time web search technology has developed enormously, today search engines can quickly return results on single-word queries of a 15-terabyte corpus, furthermore new techniques from natural language processing (NLP) such as information extraction, automatic identification of named entities, machine translation, taxonomy generation and classification have been combined with classic search methods and have shown significant benefits. Broder et al. (2004) states that this revolution has exposed hundreds of millions of people to the experience of searching and taxonomy browsing and has reshaped their expectations of the knowledge retrieval process, and this is not only while browsing the web, but more importantly, while at work when they are performing their jobs.

2 ENTERPRISE INFORMATION RETRIEVAL: THE BUSINESS PROBLEM

An October 2000 study by the University of California, Berkley concludes that, at the time of publication of the study, there were 550 billion documents on the Internet, Intranets and Extranets; a number that increases by 7.5 million each day. Delphi Group (2002) named this information explosion a “digital sprawl” and points out that its consequences can be just as disastrous to smooth functioning of the Enterprise as the snarls of the rush hour traffic are to the overloaded expressways. Moreover, challenges associated with managing this kind of information have often been identified using the metaphor of the “needle and haystack”.

The current problem of overabundance of information is affecting the efficiency of the enterprises in general but specifically of those whose operations increasingly involve effective access and manipulation of information as their key operating competence. Raghavan (2001) states that it is estimated that about a third of the time of a typical knowledge worker is spent searching for the information. The IDC report entitled “The High Cost of Not Finding Information” by Feldman and Sherman (2003) quantifies the significant economic penalties both in form of lost opportunities and through lost productivity.

Further it is noticed that on the contrary to searching on the web, expectations have not been met at the enterprise level. Knowledge management in the enterprise setting and even simple document search functions often produces disappointing results. The same thing has been observed by Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia; that poor enterprise search is considered a normal thing, it is further pointed out that if the employees desire it or the customers complain about it, even then the organization as a whole typically fails either to recognize the seriousness of the situation or the possibility of doing it in a better way (Broder et al., 2004; Hawking, 2004).

Although now there is much awareness that effective enterprise searching will bring massive economic benefit and many research efforts are going on, but so far still there are a lot of unsolved problems in this area. At present yet again researchers are faced with the problem of the same magnitude and dimension, as they were with the word wide web i.e. how to bring highly effective search to the complex information space within the enterprise. Hawking (2004) explains that the present work on enterprise searching is still in infancy and so far the research just hints at its economic magnitude, states some of the unsolved question in the enterprise search domain and enterprise search test collection has been proposed.

3 BOUNDARIES OF ENTERPRISE SEARCH

Majority of the information in an enterprise is unstructured which means that it does not resides in the databases. All this unstructured information resides in the form of HTML pages, documents in proprietary formats and forms (e.g. paper and media objects) together with information in relational and proprietary databases, all these documents constitute the enterprise information ecosystem (Mukherjee and Mao, 2004).

Hawking (2004) interprets the term Enterprise Search to include:

- Any organization with the text content in electronic form;
- Search of the organizations external website;
- Search of the organizations internal web site (Intranet);

- Search of the non-textual and continuous media held with in the organization
- Search of other electronic text held by the organization in the form of email, database records, documents on fileshares and like.

4 REQUIRED CHARACTERISTICS: THE EXPECTATIONS

The ultimate goal of the enterprise search system is to respond to a request by searching all the documents that may possibly contain a useful answer and then to present a search result in the order, which is of “maximal utility to the searcher” Hawking (2004). Hence Broder’s (2002) “answering the need behind the query” is also very much applicable to the enterprise search. (Abrol et al., 2001; Raghavan, 2001) have identified that any enterprise search engine should have following characteristics:

- The need to access information in diverse repositories including file systems, HTTP web servers, Lotus Notes, Microsoft Exchange, content management systems such as Documentum, as well as relational databases.
- The need to respect fine-grained individual access-control rights, typically at the document level; thus two users issuing the same search/navigation request may see differing sets of documents due to the differences in their privileges.
- The need to index and search a large variety of document types (formats), such as PDF, Microsoft Word and PowerPoint files, etc. and different languages (such as, English, European and Asian languages).
- The need to seamlessly and scalably combine structured (e.g. relational) as well as unstructured information in a document for search, as well as for organizational purposes (clustering, classification, etc.) and for personalization.
- The need to combine search results from internal as well as external sources of information.

Stenmark (1999) has compiled a comprehensive list of criteria’s to check the suitability of an enterprise information retrieval system regarding supported platforms, document formats, real time

updating of the indexes, installation and maintenance of the product. However, Hawking (2004) has criticized that none of the 81 criteria's in Stenmark tables relate to the quality of the search result and further points out that the required characteristics identified by (Abrol et al., 2001; Raghavan, 2001) do not completely represent the complexity of the situation, as nowadays many enterprises are building systems in which documents are synthesized from the paragraphs stored in the databases; moreover any standard product should also deal with the security and accessibility issues, like which candidate paragraphs are to be presented depends on the searchers interest profile and his access rights.

5 PROBLEMS & CHALLENGES FACED: THE DIFFERENCE FROM THE WEB

With all the technological advancements in the field of information retrieval, it seems that search in enterprise should be improving and indeed be easier than the searching on the web, in the same way employees also seek web-like experience in the enterprise, but the internet and enterprise domain differ fundamentally in the nature of the content, user behavior and economic motivations. It is observed that "good" answer to the query on the Internet is the most relevant or the best matched document, whereas on the contrary the notion of "good" answer on the intranet is often defined as the "right" answer. Since enterprise users might know or have previously seen the specific document(s) that they are looking for therefore unlike Internet the correct answer is not necessarily the most "popular" document; moreover it is emphasized that finding the right answer is often more difficult than finding the best answer (Fagin et al., 2003; Mukherjee and Mao, 2004; Raghavan, 2001; Hawking, 2004).

It is expected that enterprise information delivery must clearly meet the performances that users have come to expect on the Internet. Although some techniques for scaling and performance developed on the web can be adapted to the enterprise, many techniques for searching, organizing and mining information on the web are less applicable to the enterprise. Despite the fact that the enterprise corpora are smaller; they lack the highly hyperlinked nature of the web, hence the most successful techniques of the web, based on link analysis do not apply in the enterprise. This results in lower

relevancy of retrieval documents. Other factors are of security, reliability and performance issues that complicate the problem (Broder et al., 2004; Mukherjee and Mao, 2004).

One of the many challenges faced is that content from heterogeneous repositories e.g. email system and content management system typically do not cross-reference each other via hyperlinks, where as on the Internet the strongly connected component accounts for roughly 30% of crawled pages but on corporate intranets this is much smaller e.g. 10% on IBM's intranet. Also it is observed that popular PageRank and HITS algorithms that work extremely well on Internet are not as good on the intranets; therefore there is a need to employ other techniques to improve the search relevance on an intranet (Kleinberg, 1999; Brin and Page, 1998). Moreover deployment environment for these domains also differs, economic and time constraints in enterprises prevent quick up-gradation to new technologies, all these factors lead to the dissatisfaction of the end user due to the poor quality of search result.

Hawking (2004) has summarized the research problems that are arising in the area of enterprise searching;

- Defining an appropriate enterprise search test collection.
- Effective ranking over heterogeneous collections characteristic of enterprises.
- Building an employee portal - A distributed IR problem.
- Effective search over collections of e-mail.
- Estimating document importance for documents, which are not part of a web.
- Exploiting search context within enterprise searches.
- Providing effective search over foreseeable future enterprise collections of interlinked continuous media.

So far the problem of enterprise searching is a challenge for the researchers and the commercial companies, but the solution for which will deliver enormous benefits. Although many enterprise search engines are available but only few are able to work with the range of databases, content management systems, emails formats, document formats, operational and security requirements of a typical medium scale enterprise. The reason being that due to the complex nature of typical enterprise information space a highly performing text retrieval algorithm developed in the laboratory cannot be applied directly to the organization. Therefore it is

very difficult to measure the quality of the search results obtained and very hard to approach the effectiveness level achieved by the state of the art whole of web search engines (Hawking, 2004).

6 PROPOSED SOLUTIONS

One of the obvious solutions could be conversion of all the non-web data of an enterprise into the web format. Emails can be converted to html documents using available converters like hypermail, in the same way documents can also be converted to HTML or XML formats. However (Hawking, 2004) argues that this simple transformation will not solve the problem, unless the converted documents are interlinked and organized in the same way as in the normal web. Delphi Group (2002) recommends that new and enhanced capabilities like text analytics, classification, profiling, search and improved delivery components should be combined with basic keyword search to provide enterprises with ways to organize, find and leverage their information assets into improved decision making and increased productivity.

Different researchers have worked on the enterprise search problem and have revealed critical findings; Craswell et al (2001a) show dramatic benefits from the use of anchor text on Australian National University (ANU) web, Hawking et al. (2004) conclude that link evidence from the external web is unnecessary for good performance in navigational search tasks on enterprise web sites. Upstill et al. (2003) have investigated the value of query independent evidence such as indegree, two variants of Page Rank and URL-type in homepage finding tasks on three different test collections and the ANU intranet data, they found that PageRank gives no discernable benefit over indegree for collections up to 18.5 million pages; Furthermore considering the above mentioned findings it is suggested that the following specific techniques can improve the enterprise search.

6.1 Spidering and Indexing

Data must be spidered and indexed before it can be searched, in the commercial world most of the enterprise applications do not expose information about what has changed; hence Mukherjee and Mao (2004) suggest that adoption of search standards by application vendors can help solve this problem.

6.2 Data Filtering and Search Relevance

It is proposed that clean data will result into better search relevance; it also helps automatic classification, feature extraction and clustering. Especially for those enterprises that are going to index external content it is important to use techniques such as link-density analysis and entity-extraction to filter the data. Mukherjee and Mao (2004) argue that since intranets are essentially spam free therefore it's not appropriate to imply web strategies like hyperlink analysis; on the other hand rank-aggregation approach proposed by Fagin et al. (2003) is suitable.

6.3 Classification and Taxonomy Navigation

Searching provides an efficient way for users to find relevant information in business portals but only if they know what to search for; research has shown that presenting results in categories provides better usability, hence there is a different need for browsing and navigating information (Dumais et al. 2001; Raghavan, 2001).

Taxonomies are the most popular way of organizing documents into a navigable structure, with taxonomy users can easily navigate through the category hierarchy to find relevant information. Further it is suggested that scoped searching within a category typically returns more relevant results than un-scoped search. Well-known examples of taxonomies include the directory structure of Yahoo! and the Open Directory Project. Currently there is need to do manual work in this area, although it is a fact that manual taxonomy creation is time consuming and expensive, but automation in this area is still in its infancy and hence is not reliable (Dumais et al. 2001; Raghavan, 2001).

6.4 Information Extraction and Text Mining

For structuring, accessing and maintaining large amounts of heterogeneous information, appropriate meta-level descriptions are needed, which specify the structure, context and potential usage of object level knowledge. Liao et al. (1999) suggests an ontology based approach for meta modeling and retrieval of heterogeneous data, formal knowledge and documents, further they identified information ontology, domain ontology and enterprise ontology

as main contributors to a vocabulary for comprehensive information meta modeling.

Metadata in semi-structured documents greatly improves content search and organization, but Liao et al. (1999) points out that research at the enterprise data revealed that important metadata in documents e.g. author is often incorrect and is set to some default value. It is suggested that correct metadata should be enforced to improve the quality of data; domain experts can be hired to tag or annotate the documents manually, however manual approach is not adequate for large volumes of information; hence automation is necessary.

Moreover information extraction and text mining are useful tools for reducing tagging costs, however effectiveness of information extraction and text mining depends on the document quality and homogeneity of the target information entities. Mukherjee and Mao (2004) suggested that removal of redundant/obsolete data would benefit relevance. Other techniques such as duplicate detection and near duplicate detection can ensure that irrelevant data is eliminated from active corpora.

6.5 Federation

In a typical organization, all the information is not accessible to every one and there are cases where it cannot be indexed or is forbidden from being indexed due to legal or security constraints. Choo et al. (2002) suggest that in such cases federated search is the only way to provide a single point of access to data from enterprise repositories and applications.

6.6 Use Base Relevance

Since enterprise users have their specific identity, Mukherjee and Mao (2004) proposed that user profiles could enhance the input context to provide personalization and targeted search, historical patterns of access can also be useful, moreover it will enable the users to participate in taxonomy building process.

It has been observed that people employ variety of strategies when searching through emails, files or web bookmarks for a specific item. Ringel et al. (2003) has explored the effects of providing important events as context to support searching through the contents. On the other hand also there is need to have a proper implementation strategy for the enterprise searching system, Delphi Group (2002) stresses that while an effective IR strategy improves the working environment in the long run, in the short term it requires people to make changes

in the way they work, without the effective change-management strategy, any IR implementation is unlikely to meet expectations.

7 AVAILABLE TOOLS & PRODUCTS

So far many commercial enterprise search products have been developed and launched by different vendors; but still they are not very widely adopted by the market, according to survey by Delphi Group (2002) less than 25% of organizations have actually deployed classification software, in that Verity is the leading application deployed followed by Inktomi. Others include Microsoft, Alta Vista, IBM/Lotus, Memex and Autonomy. Following is a brief overview of the techniques used by leading enterprise-searching products:

7.1 Verity's K2 Enterprise

Raghavan (2001) states that for each type of repository be accessed, Verity provides a gateway, which allows spider to access the content in the repository together with the associated security information regarding which user can access what documents. Moreover it also provides automatic classification, personalization, combined text and structured querying.

7.2 Google Enterprise Solution

It is suggested by Google (2005) that while making a selection of enterprise search solution a major consideration should be to select such a system that can index documents without adding overhead, either for document creators or for the administrators. Further it is emphasized that Google enterprise search solution intelligently integrates usability and power; hence will boost the productivity and will put the intellectual capital to work.

7.3 Panoptic Expert

In large and growing organizations it can be difficult to keep track of the employee expertise, Craswell et al. (2001b) explains that Panoptic Expert is a web based system which automatically identifies experts in an area, based on the documents already published in an organizations intranet. Moreover to simplify the things a web like interface has been

provided to the system but instead of returning documents it returns a list of experts.

7.4 Microsoft's Stuff I've Seen (SIS)

Microsoft has developed a system called Stuff I've Seen (SIS) that makes it easy for people to find information they have seen before. Dumais et al. (2003) explains that the two key aspects of the SIS are provision of unified index of information, regardless of the format and as the user might have seen the information before, therefore rich contextual are used to present information.

7.5 IBM's UIMA SDK

On the contrary to other researchers, Broder et al. (2004) makes a point and emphasized that the major factor is the integration of the technologies rather than the difference between the nature of the web and the enterprise. Moreover they point out that the current advanced technologies do not work together and typically each of them has completely different view of the world, represent the underlying documents in different ways and are concerned with performance in different areas. Hence it is stressed that the "missing part" is the architecture that enables the integration of the technologies with search and retrieval. Broder et al. (2004) states that such architecture has been developed within the IBM research namely, the Unstructured Information Management Architecture (UIMA).

7.6 Search-Derivative Applications (SDAs)

Since nowadays searching is inherent in most major business applications, ranging from all type of enterprise content management (ECM) solutions, via supply-chain systems, to enterprise resource planning (ERP) and customer relationship management (CRM). Hence Lervik (2004) points out that companies that use an enterprise search platform at the core of any business solution have the flexibility to not only solve one problem, but any task that can make their life easier and increase the efficiency and productivity. Moreover Lervik (2004) states that SDAs are the future of the enterprise application development.

8 CONCLUSIONS

With the abundance of the electronic information on the Internet, Intranets and Extranets, searching the relevant information through the piles of documents is a growing problem. Although on the World Wide Web the problem of searching has been solved to a great extent and queries do get relevant response in an efficient way but the enterprises are still faced with the problem of effectively searching their intranets and webs; however efficient retrieval of the relevant information is a critical success factor for many enterprises and due to lack of it businesses do suffer from significant economic penalties both in form of lost opportunities and through lost productivity.

Since the web searching is quite effective, therefore organizational employees have the same expectations from the enterprise searching, apparently it seems that since search technology has advanced a lot so enterprise searching should not be a problem, but on the contrary a lot of problems and challenges are faced while the searching the relatively much smaller corpora of the enterprise. Enterprises are very complex space, and searching must be done on different format of documents, emails, relational databases, internal web and the company's web site. In addition to many other challenges faced, the major problem is the difference in the searching on web and an enterprise; and the well-proven web information retrieval techniques cannot be used on the enterprise data. Moreover there are security, reliability and performance issues that complicate the problem.

So far many solutions have been proposed and different techniques have been devised to improve the searching in the enterprise, it is suggested that already in place information systems should expose the changes in the data to the search application so that it can be spidered and indexed effectively. Moreover filtering, removal of duplicate data and adding meaningful meta data to the documents will improve the searching. Other important factor could be creation of taxonomies and using the profile information to identify the accessibility of the user and to provide personalized and targeted search.

Although many commercial enterprise search products have been launched in the market but it has been observed that very few organizations have adopted them. Now there is much awareness that effective enterprise searching will bring massive economic benefit and research efforts are going on, but still there are a lot of unsolved problems in the area of enterprise information retrieval.

REFERENCES

- Abrol, Mani., Latache, Neil., Mahadevan, Uma., Mao, Jianchang., Mukherjee, Rajat., Raghavan, Prabhakar., Tourn, Michel., Wang, John. and Zhang, Grace. (2001). "Navigating large-scale semistructured data in business portals". In Proceedings of the 27th VLDB Conference, pages 663-666, Roma, Italy, 2001. www.vldb.org/conf/2001/P663.pdf [Accessed 07 May 2005].
- Brin, Sergey. and Page, Lawrence. (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine". *Computer Networks* 30(1-7): 107-117 (1998). <http://www-db.stanford.edu/pub/papers/google.pdf> [Accessed 08 May 2005].
- Baeza-Yates, Ricardo. and Ribeiro-Neto, Berthier. (1999). "Searching the Web." *Modern Information Retrieval*. New York: ACM Press/Addison-Wesley, 1999. 367-395.
- Broder, Andrei. (2002). "Taxonomy of Web Search". *SIGIR Forum* 36(2): 3-10 (2002). <http://www.sigir.org/forum/F2002/broder.pdf> [Accessed 07 May 2005].
- Broder, Andrei., Ciccolo, Arthur. (2004). "Towards the next generation of enterprise search technology". *IBM Systems Journal* 43(3): 451-454 (2004) www.research.ibm.com/journal/sj/433/broder.pdf [Accessed 08 May 2005].
- Craswell, Nick., Hawking, David., Robertson, Stephen. (2001a). "Effective Site Finding Using Link Anchor Information". *SIGIR 2001*: 250-257. http://research.microsoft.com/users/nickcr/pubs/craswell_sigir01.pdf [Accessed 08 May 2005].
- Craswell, Nick., Hawking, David., Vercoustre, Anne-Marie, and Wilkins, Peter. (2001b). "Panoptic Expert: Searching for experts not just for documents". *Ausweb Poster Proceedings*, 2001. http://research.microsoft.com/users/nickcr/pubs/craswell_ausweb01.pdf [Accessed 09 May 2005].
- Choo, K., Mukherjee, R., Smair, R., and Zhang, W. (2002). "The Verity federated infrastructure". *Proceedings of the Conference on Information and Knowledge Management, McLean, Virginia* (2002) 621. <http://portal.acm.org/citation.cfm?id=584792.584897> [Accessed 08 May 2005].
- Dumais, S. T., Cutrell, E., and Chen, H. (2001). "Optimizing search by showing results in context". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, Washington* (March 2001) 277-284. <http://research.microsoft.com/~sdumais/chi2001.pdf> [Accessed 08 May 2005].
- Delphi Group. (2002). "Perspectives on Information Retrieval". http://www.delphigroup.com/research/ir_perspectives_summary.pdf [Accessed 08 May 2005].
- Dumais, Susan., Cutrell, Edward., Cadiz, Jonathan., Jancke, Gavin., Sarin, Raman., Robbins, Daniel. (2003). "Stuff I've seen: a system for personal information retrieval and re-use". *SIGIR 2003*: 72-79 <http://research.microsoft.com/~sdumais/SISCore-SIGIR2003-Final.pdf> [Accessed 09 May 2005].
- Feldman, Susan. and Sherman, Chris. (2001). "The High Cost of Not Finding Information: An IDC White Paper." IDC, 2001. <http://www.knowledge-wave.com/scripts-include/en-us/downloads/idcinfo2996.pdf> [Accessed 07 May 2005].
- Fagin, Ronald., Kumar, Ravi., McCurley, Kevin., Novak, Jasmine., Sivakumar, D., Tomlin, John., Williamson, David. (2003). "Searching the Workplace Web". *WWW 2003*: 366-375 <http://www.almaden.ibm.com/cs/people/siva/papers/isearch.pdf> [Accessed 07 May 2005].
- Google, Inc. (2005). "Simplicity and Enterprise Search A New Model for Managing Your Enterprise Information". [Online]. January 2005. http://www.google.co.uk/enterprise/pdf/google_simplicity_enterprise_wp.pdf [Accessed 09 May 2005].
- Hawking, David. (2004). "Challenges in Enterprise Search". *Proceedings of the Australasian Database Conference, ADC2004*. http://es.csiro.au/pubs/hawking_adc04keynote.pdf [Accessed 07 May 2005].
- Hawking, David., Crimmins, Francis., Craswell, Nick., Upstill, Trystan. (2004). "How Valuable is External Link Evidence When Searching Enterprise Webs?". *ADC 2004*: 77-84. http://research.microsoft.com/users/nickcr/pubs/hawking_adc04.pdf [Accessed 08 May 2005].
- Kleinberg, Jon. (1999). "Authoritative Sources in a Hyperlinked Environment". *ACM* 46(5): 604-632 (1999). <http://www.cs.cornell.edu/home/kleinber/auth.pdf> [Accessed 08 May 2005].
- Liao, M., Abecker, A., Bernardi, A., Hinkelmann, K. and Sintek, M. (1999). "Ontologies for Knowledge Retrieval in Organizational Memories", In *Proceedings of the Learning Software Organizations (LSO'99) workshop, Kaiserslauten, Germany*, ed., F. Bomarius, pp. 19--26, (June 1999). <http://citeseer.ist.psu.edu/liao99ontologies.html> [Accessed 08 May 2005].
- Lervik, John. (2004). "The new perspective on enterprise search". *Proceedings of Online Searching 2004. Fast Search & Transfer ASA, Norway*. [Online]. 2004. http://www.online-information.co.uk/2004proceedings/thursam/lervik_j.pdf [Accessed 03 May 2005].
- Mukherjee, Rajat. and Mao, Jianchang. (2004). "Enterprise Search: Tough Stuff". *ACM Queue* vol. 2, no. 2 - April 2004 [Online]. <http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=140> [Accessed 08 May 2005].
- Raghavan, Prabhakar. (2001). "Structured and Unstructured Search in Enterprises". *IEEE Data Eng. Bull.* 24(4): 15-18 (2001).

- <http://sites.computer.org/debull/A01dec/verity.ps>
[Accessed 07 May 2005].
- Ringel, Meredith., Cutrell, Edward., Dumais, Susan., Horvitz, Eric. (2003). "Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores". Proceedings of Interact 2003, p. 184-191.
<http://research.microsoft.com/~horvitz/landmark.pdf>
[Accessed 09 May 2005].
- Stenmark, Dick. (1999). "A Method for Intranet Search Engine Evaluations", In Käkölä, T. (ed.), Proceedings of IRIS22, August 7-10, Department of CS/IS, University of Jyväskylä, Finland.
<http://w3.informatik.gu.se/~dixi/publ/method.pdf>
[Accessed 07 May 2005].
- Upstill, Trystan., Craswell, Nick., Hawking, David.(2003). "Query-Independent Evidence in Home Page Finding". ACM Trans. Inf. Syst. 21(3): 286-313 (2003).
http://research.microsoft.com/users/nickcr/pubs/upstill_tois03.pdf [Accessed 08 May 2005].



SciTeP
Science and Technology Publications