

Architecture of a Specialized Search Engine Using Mobile Agents

Aditya Kumar Mishra

M.Tech. Student
Shobhit University, Meerut, India
qwsaditya@gmail.com

Dr. Niraj Singhal

Associate Professor
Shobhit University, Meerut, India
sonia_niraj@yahoo.com

Abstract

Today with enormous growth of the Internet and the web, finding the right information has become a very challenging task. There is a need for a more effective way of retrieving information from the web. Search engines works as bridge between web users and web documents. Without search engines, this is a difficult to surf this vast amount of information. The rapid growth and mixture of objects available on the web has posed incomparable challenges for the designers of search engines. This paper presents need and general architecture of a search engine, several types of search engines and design issues. It also presents architecture of a specialized search engine using mobile crawling agents.

Keywords: Internet, World Wide Web, Specialized Search Engines, Mobile Agents

1. INTRODUCTION

The Internet [3,9] has modernized the world of computer and communications like never before. It has provided a globalized dimension to the world to share the source of information. Within a span of few years, the Internet [9] has changed the way we do business and communicate. The web [4,9] is a hyperlinked repository of hypertext documents lying in different websites distributed over far-end distant geographical locations. Web is messy and poorly structured where much poor quality information is placed together with useful information and, finding the right information from it at the right time is a very challenging task. Hence, there is a need for a more effective way of retrieving information on the Internet.

To help this sheer volume of data available on the web, programs that run off of special websites called search engines; are being designed. The search engines that contain information of some specific field are called specialized search engines. Specialized search engines index pages for particular topics only and very often there are many pages that cannot be found in general-purpose search engines but only in specialized ones.

A mobile software agent [6] is able to migrate from host to host to work in a heterogeneous environment. The mobile agent environment, which is a software system distributed over a network of heterogeneous computers and its primary task is to provide an environment in which mobile agents can run. A mobile agent transports itself along with its state. When it reaches the new host, the agent should be able to perform appropriately in the new environment. This paper presents an architecture of a specialized search engine using mobile agents that helps to reduce the network load significantly.

2. SEARCH ENGINES

A web search engine [4,9] is a coordinated set of programs that is able to read every searchable page on the web, create an index of the information it finds, compare that information to a user's search request (i.e. query), and finally return the results back to the user. A general web search engine (figure. 1) has three parts i.e. *Crawler*, *Indexer* and *Query engine*.

The web crawlers are small programs that peruse the web on the search engine's behalf, and follow links to reach different pages. Starting with a set of seed URLs, crawlers extract URLs appearing in the retrieved pages, and store pages in a repository database. The indexer extracts all the uncommon words from each page and records the URL where each word has occurred. The result is stored in a large table containing URLs; pointing to pages in the repository where a given word occurs. The query engine is responsible for receiving and filling search requests from users. It relies on the indexes and on the repository.

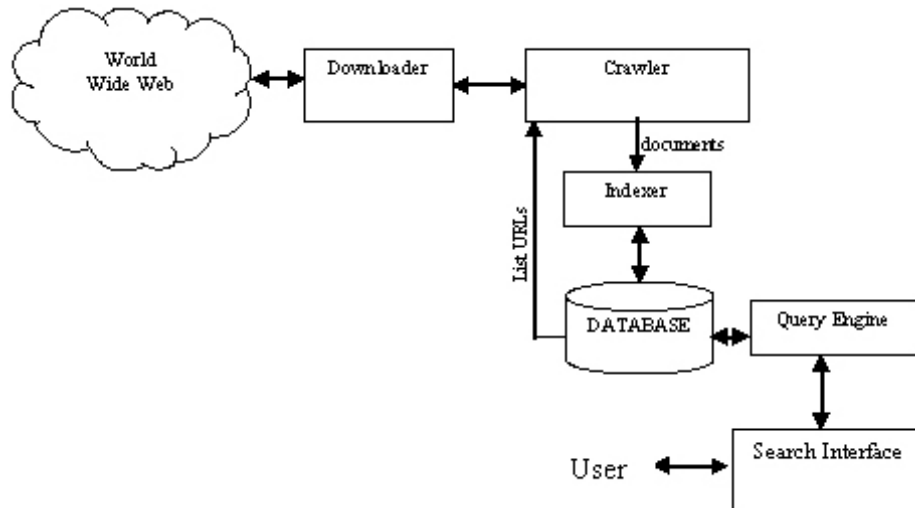


Figure 1. Architecture of a typical web search engine

2.1 Types of Search Engines

Search engines are good at finding unique keywords, phrases, quotes, and information contained in the full text of web pages. Search engines allow user to enter keywords and then search for them in its table followed by database. Various types [9] of search engines available are, crawler based search engines, human powered directories, meta search engines, hybrid search engines and specialized search engines.

Crawler based search engines create their listings automatically with the help of web crawlers. Such search engines are huge and often retrieve a lot of information. Human powered directories depend on humans to create repository. They are organized into subject categories and classification of pages is done by subjects. Hybrid search engines typically favor one type of listing over the other. Many search engines today combine a crawler based search engine with a directory service. Meta search engines accumulate search and screen the results of multiple primary search engines. Meta crawlers allow searches to be sent to several search engines all at once. The results are then blended together onto one page.

Specialized search engines search specialized databases, allow users to enter search terms in a particularly easy way, look for low prices on items they are interested in purchasing, and even give users access to real, live human beings to answer questions. Specialized search engines index pages for particular topics only and very often there are many pages that cannot be found in general-purpose search engines but only in specialized ones. Some of the specialized search engines are huge sites that actually host the resources they link to, or used to be search directories but have evolved to include links not only to sites that were submitted to them. There are many specialized search engines for every imaginable topic and it is always wise to be aware of the specialized search engines for your niche. Some of the specialized search engines are Pandia Powersearch, Webquest, Virtual Search Engines, The Search Engines Directory, etc.

2.2 Design issues

Three main characteristics to improve search engines are improving algorithms to search the web, using filters towards the user's results; and improving the user interface for query input. Factors that determine the quality of a search engine are freshness of contents, index quality, search features, retrieval system and user behaviour.

Web users are very heterogeneous and search engines are used by professionals as well as by laymen. The search engine needs to be enough smart to serve for both types of users. Moreover, there is a tendency that users often only look at the results set that can be seen without scrolling, and the results which are not on first page are nearly invisible for the general user. Search engines find problems in keeping its database up-to-date with the entire web because of its enormous size and the different update cycles of individual websites. Other characteristics [9] that a large search engine is expected to have are scalability, high performance, politeness, continuity, extensibility and portability.

On the web, documents are written in several different languages. The documents need to be indexed in a way which allows it to search for documents written in diverse languages with just one query. Search engines should be able to index documents written in multiple formats, as each file format provides certain difficulties for the search engines. Data is distributed widely in the web world, and is located at different sites and platforms. Documents can be added or removed easily in the web. Changes to these documents remain unnoticed by others.

3. MOBILE AGENTS

A software mobile agent [6] is a movable software which is able to migrate from host to host to work in a heterogeneous network environment. Initially this approach was known as remote programming that views computer-to-computer communication as one computer not only to call procedures in another, but also to supply the procedures to be performed. Each message that goes through the network comprises a procedure that the receiving computer is to perform and data that are its arguments. The procedure and its state are termed a mobile agent as they represent the sending computer even while they are in the receiving computer.

The mobile agent approach do not consume much network bandwidth. They only consume bandwidth when they move. They continue to execute after they move, even if they loose network connectivity with their creators. Therefore, if a client requires extensive communications with a particular server somewhere on the network, then implementing such a system using mobile agents is attractive. An agent can move closer to the remote server, reducing the network traffic, performs all tasks and comes back. During that period the client machine does not have to be switched on. It has to be switched on only when it is time to welcome back the agent.

3.1 Applications of Mobile Agent Technology

Several fields in which mobile agent technology is of immense use are [1,5], web information retrieval, data processing, mobile computing system, network management, electronic commerce, energy efficiency and metering, wireless multimedia sensors, grid computing and grid services, distributed data mining, multimedia, human tracking, security, intrusion detection, affective computing, climate environment and weather, e-learning, location, recommendation and semantic web services.

They consume fewer network resources since they move the computation to the data instead of the data to the computation. Mobile agent systems are computer and network independent. The construction of distributed systems can be made easier with mobile agents. An attractive area to use mobile agents is in processing data over distributed networks. In this paradigm, the agent travels to the nodes on the network, process the information on those nodes and then return home. Instead of spending a huge amount of time going through on-line bookstores to find the best deal on a book, firing up an agent to do this task would save us a considerable amount of time. The agent would be programmed to visit a number of bookstores and find the best deals on books we need.

4. PROPOSED WORK

The Internet and the web have given access to a large amount of information, but the search for a particular topic has become very difficult and use a lot of time and bandwidth. An agent-based solution provides one or more agents that visit web servers searching for interesting pages. It saves bandwidth because the only needed communications are those to send and receive the agent [2,7].

Using mobile agents approach [8,10], agents are useful when very large volumes of data are stored at remote hosts, these data should be processed in the locality of the data rather that transferred over the network. The main concern is, move the computations to the data rather than the data to the computations. By migrating to the location of the resource, a mobile agent can interact with the resource much faster than from across the network that reduces network traffic also. The design approach of the specialized search engine using mobile agents is shown in Figure 3.

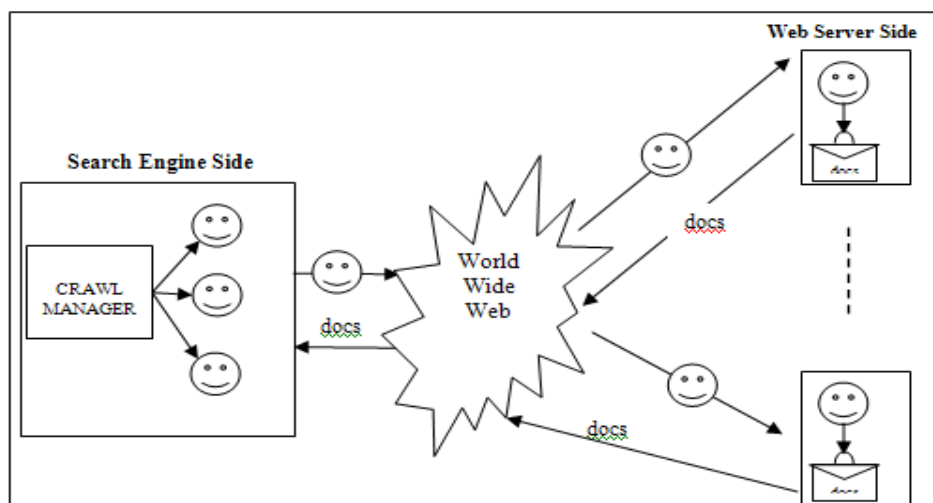


Figure 2. Specialized Search Engine using Mobile agents

As shown in the figure given above, it uses a crawler manager at the search engine side that deputes migrating crawlers to the web servers with a list of URLs of respective web servers. The migrating crawler, on reaching a server crawls the pages, select the best of the pages for its collection and comes back to the search engine with the collection. It reduces unnecessary overhead of being the unnecessary pages to the search engine site.

For example,

Let the number of documents on a remote site=100

Average size of each document=10KB

The network traffic with a traditional crawler= $100 \times 10 = 1000\text{KB}$

With mobile agents,

Let size of a mobile agent is 10 KB and 50% of the documents on the webserver are of use, then,

The network traffic with migrating crawler= $10 + (10 \times 50) + 10 = 520\text{KB}$

It shows that using a migrating crawler, the network traffic can be reduced a lot. While searching for a special field, all the pages on a website are of no use and it is of no use to bring all pages to the search engine side. If 50% of the pages found on a webserver are of use, then the network traffic for a traditional crawler is 1000KB whereas with a migrating crawler it can be reduced to 520 KB only. It decreases the network traffic very much.

5. CONCLUSION

Size of the digital information on the web is increasing at a very fast rate. Efficient search engines are needed to collect this huge source of information, and designing such a quality search engine is not an easy task. Considering wide variety of objects available on the web, it is desirable that a search engine should not only be able to surf them but also be able to select high ranked web data for the user. Specialized search engines search specialized databases and allow users to search information they are interested in. Specialized search engines index pages for particular topics only and very often there are many pages that cannot be found in general-purpose search engines but only in specialized ones. Mobile agent are able to migrate from one computer to another autonomously and continue its execution on the destination computer. This paper presented use mobile agent technology in designing specialized web search engines that help in reducing network load significantly.

6. REFERENCES

- [1] Abdelkader Outtagarts, "Mobile Agent-based Applications: a Survey", International Journal of Computer Science and Network Security, Vol. 9 No. 11, pp. 331-339, November 2009.
- [2] Ashutosh Dixit, Jitendra Kumar Seth and A. K. Sharma, "A migrating parallel exponential crawling approach to search engine" International Journal of Computer Science & Emerging Technologies, Vol. 1, Issue 2, August 2010.
- [3] Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard, Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, Stephen Wolff, "A Brief History of the Internet", www.isoc.org/internet/history.
- [4] Dirk Lewandowski, "Web searching, search engines and Information Retrieval, Information Services & Use", 25 (2005) 137-147, IOS Press, 2005.
- [5] Eid Mohamad, Artail Hassan, Kayssi Ayman and Chehab Ali, "Trends in Mobile Agent Applications" Journal of Research and Practice in Information Technology, Vol. 37, No. 4, pp. 323-351, November 2005.
- [6] H. S. Nwana, "Software Agents: An Overview," Knowledge Engineering Review, Cambridge University Press, Vol. 11, No. 3, pp.1-40, 1996.
- [7] Jingsheng Zhao, Jiapeng Zhang, and Haiping Fang, "A Study of Mobile Agent-based Information Searching System", Proceedings of 2009 International Workshop on Information Security and Application, Qingdao, China, November 21-22, 2009, ISBN: 978-952-5726-06-0, 2009.
- [8] Moemeng, C.; Gorodetsky, V.; Zuo, Z.; Yang, Y. and Zhang, C., "Agent-Based Distributed Data Mining: A Survey", Data Mining and Multiagent Integration, Springer Science Business Media, LLC 2009.
- [9] Niraj Singhal and Ashutosh Dixit, "Web Crawling Techniques: A Review", Proceedings of National Seminar on Information Security: Emerging Threats and Innovations in 21st Century (ISBN: 978-81-89547-55-4), India, pp. 34-43, April 2009.
- [10] Niraj Singhal, R. P. Agarwal, Ashutosh Dixit, A. K. Sharma, "Information Retrieval from the Web and Application of Migrating Crawler", published in IEEEExplore 978-0-7695-4587-5/11 © 2011 IEEE, DOI 10.1109/CICN.2011.99. pp. 476-480, 2011.