International Journal of Scientific Research Engineering & Technology (IJSRET)
Volume 1 Issue11 pp 016-021 February 2013 www.ijsret.org ISSN 2278 - 0882

Design of Human Powered Directories using Mobile Agents

Sandeep Rana[1], Niraj Singhal[2]
[1]Shanti Institute Of Technology, Meerut, India
[2]Shobhit University, Meerut, India
1sandy_rana33@yahoo.com, 2sonia_niraj@yahoo.com

ABSTRACT
The Internet is a worldwide mechanism for information dissemination, a medium for collaboration and communication between individuals and their computers from local to global scope. Web is a system of interlinked hypertext documents accessed via Internet. Web contains billions of visible pages and it is not easy for a user to search for a specific web page. Search Engines help users to search for specific web page out of huge collection of pages. Human powered directories depend on humans to create a repository. In this paper we present use of mobile agents in designing Human powered directories.

Key words: Internet, world wide web, Human powered search directories, mobile agents.

1. INTRODUCTION
The Internet is a global system of interconnected computer networks that use the standard Internet protocol suite (often called TCP/IP) to serve billions of users worldwide. It is a network of networks that consists of millions of private, public, academic, business, and government networks. The Internet carries an extensive range of information resources and services, such as the inter-linked hypertext documents of the World Wide Web and the infrastructure to support email.

The World Wide Web (WWW or Web), is a system of interlinked hypertext documents accessed via the Internet. The Web [9] is a global, large repository of text documents, images, multimedia and many other items of information, referred to as information resources. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia, and navigate between them via hyperlinks documents It is estimated that web contains more than 2000 billion visible pages and five times more lying in the hidden web. Due to the extremely large number of pages present on Web, the search engine depends upon crawlers for the collection of required pages.

Human powered directories are built by human selection i.e. they depend on humans to create repository. They are organized into subject categories and classification of pages is done by subjects. Such directories do not contain full text of the web page they link to, and are smaller than most search engines [6].

A mobile software agent [10] 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.

2. RELATED WORK
A search engine [2] 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. The invention of this technology has granted users quick and easy access to the knowledge they seek; by essentially categorizing web pages according to their relevancy in regards to a request, or query. Search engines operate as a link between web users and web documents. Without search engines, this vast source of information in web pages remain hidden from us. It is a searchable database which collects information from web pages on the Internet, indexes the information and then stores the result in a huge database where from it can be searched quickly. A general web search engine (figure 1) has

three parts [1,2,9] 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 them to the list of URLs to be crawled. 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. Methods used in web database creation are full text indexing, keyword indexing and human indexing. Full text indexing is where every word on the page is put into a database for searching. It helps user to find every example in response to a specific name or terminology. A general topic search will not be very useful in the database and one has to dig through a lot of false drops. In keyword indexing only important words or phrases are put into a database. In human indexing a person examines the page and determines a very few key phrases that describes it. This allows the user to find a good start of works on a topic, assuming that the topic was picked by the human as something that describes the page.

The query engine is responsible for receiving and filling search requests from users. It relies on the indexes and on the repository. Because of the web's size, and the fact that users typically only enter one or two keywords, result sets are usually very large.

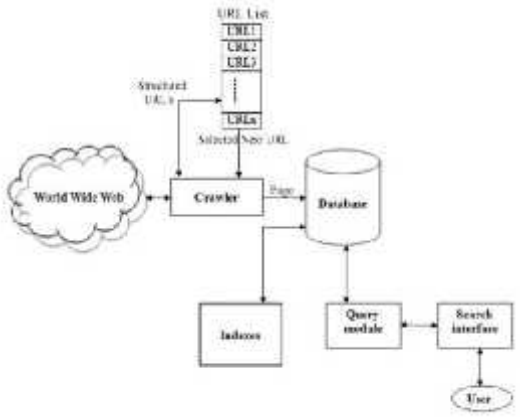

Figure 1: Architecture of a typical web search engine

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 users to enter keywords and search for them in database table and indexes. Various types [6] of search engines available are Crawler based search engines, Human powered directories, Meta search engines and Hybrid search engines.

Crawler based search engines create their listings automatically with the help of web crawlers. It uses a computer algorithm to rank all pages retrieved. Such search engines are huge and often retrieve a lot of information. It allows to search within the results of previous search and enables us to refine search results. Human powered directories are built by human selection i.e. they depend on humans to create repository.

They are organized into subject categories and classification of pages is done by subjects. Such directories do not contain full text of the web page they link to, and are smaller than most search engines. Hybrid search engines differ from traditional text oriented and directory based search engines. These search engines typically favor one type of listing over the other. Meta search engines accumulate search and screen the results of multiple primary search engines. Unlike search engines, meta crawlers don't crawl the web themselves to build listings. Instead, they allow searches to be sent to several search engines all at once. The results are then blended together onto one page.

2.2 Human powered directories

Human powered directories are built by human selection i.e. they depend on humans to create repository. They are organized into subject categories and classification of pages is done by subjects. Such directories do not contain full text of the web page they link to, and are smaller than most search engines[6].

Human-powered directories, such as the Yahoo directory, Open Directory and LookSmart, depend on human editors to create their listings. Typically, webmasters submit a short description to the directory for their websites, or editors write one for the sites they review, and these manually edited descriptions will form the search base. Therefore, changes made to individual

web pages will have no effect on how these pages get listed in the search results.

Human-powered directories are good when a user is interested in a general topic of search. In this situation, a directory can guide and help us narrow our search and get refined results. Therefore, search results found in a human-powered directory are usually more relevant to the search topic and more accurate. However, this is not an efficient way to find information when a specific search topic is in mind.

2.2.1 Uses of Human powered directories

Because the directories are compiled by humans and for humans the relevancy of directory results is very high. Search engines know that almost any search engine spider starts its regular crawl at a directory. If one is listed there, the search engines will find it during their next crawl, even without its submission.

Additionally, directories themselves have high Page Rank values. When one is listed in a directory, the directory is linking to one's site. This link is considered high-quality and is very good for one's overall link popularity, which is an influential factor when it comes to ranking. For the search engines that are able to analyze the context of one's link (such as Google), the directory listing will be important because the link is supplemented by a relevant description.

2.3 Mobile Agents

An agent [8] is an entity that acts on behalf of others in an autonomous fashion and performs its actions in some level of proactivity and reactivity, exhibits some levels of the key attributes of learning, co-operation, and mobility. A software agent [10] is a mobile software agent if it is able to migrate from host to host to work in a heterogeneous network environment. Approach that is forming a new paradigm for distributed computing is one that employs mobile agents. Initially this approach was known as Remote Programming [5]. The Remote Programming approach 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.

This approach is attractive since the reliability of the network is not crucial as they do not consume much network bandwidth. They only consume bandwidth when they move. They continue to execute after they move, even if they lose 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. This is due to the fact that 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 will have to be switched on only when it is time to welcome back the agent.

2.3.1 Advantages and Applications of Mobile Agents

Mobile agents seem to be useful for many different applications. Despite the fact that there are not many distributed computing problems that cannot be solved without mobile agents, nevertheless mobile agents make certain applications easier to develop and may improve reliability and efficiency. Various advantages of mobile agents are [10]:

- **Efficiency:** Mobile agents consume fewer network resources since they move the computation to the data instead of the data to the computation.
- **Less Bandwidth:** Most communication protocols involve several interactions which cause a lot of network traffic. Mobile agents consume bandwidth only when they move.
- **Robustness and Fault Tolerance**: The ability of mobile agents to react dynamically to adverse situations makes it easier to build fault tolerance behavior in complex distributed systems.
- **Support for Heterogeneous Environments**: Mobile agent systems are computer and network independent.
- **Support for Electronic Commerce**: Mobile agents are being used to build electronic markets since they embody the intentions, desires and resources of the participants in the market.

- **Easier Development Paradigm:** The construction of distributed systems can be made easier with mobile agents.

An attractive area to use mobile agents is in processing data over unreliable networks. In such networks, the low-reliability network can be used to transfer agents, rather than a chunk of data, from place to place. 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.

When a user fires a query to the search engine and is not able to fetch the relevance information or links of related search at one single directory of the network this means a user checked all links present on the network and the user cannot find useful information of our search, by doing this process a lot of time is wasted and relevance information is not been search yet. To overcome this type of problem a user maintains the special type of directories i.e. Human powered directories. The main advantage of maintaining these type of directories is that a user manage all the related or relevance information under single directory which can be updated or managed by using mobile agents to collect the relevance information from our home network as well as to the other web servers also.

3. PROPOSED WORK

Human-powered directories depend on human editors to create their listings. Typically, webmasters submit a short description to the directory for their websites, or editors write one for the sites they review, and these manually edited descriptions will form the search base. The changes made to individual web pages will have no effect on how these pages get listed in the search results. Therefore, search results found in a human-powered directory are usually more relevant to the search topic and more accurate.

In this paper a query based directories (Human powered directories) is created on the home network by collecting the relevance information (URL's) from the other web servers by using mobile agents and stored the fetch links in the repository of home network, by using mobile agents user only fetch relevance information and also avoid the congestion in the network.

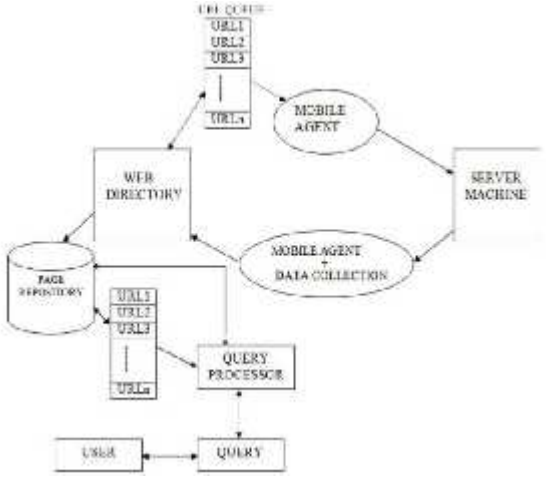

Figure 2: Architecture of Human Powered Directories

Algorithm

```
{
scan the query;
for each word of the query
{
extract a matched url u;
place u in a linked list l;
}
store l on page repository;
create a queue 'q' as query; If q != null
{
create queue(q);
}
else
{
search(q);
}
append l to q;
}
```

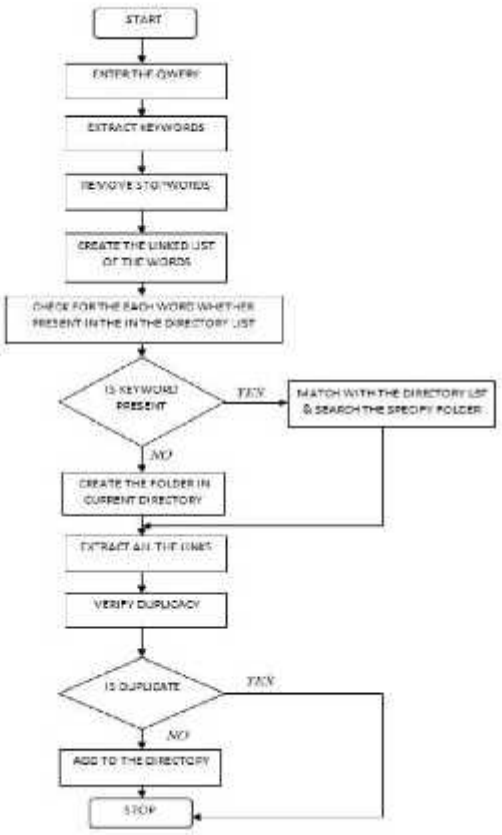

Figure 3: Flowchart to create Human Powered directories

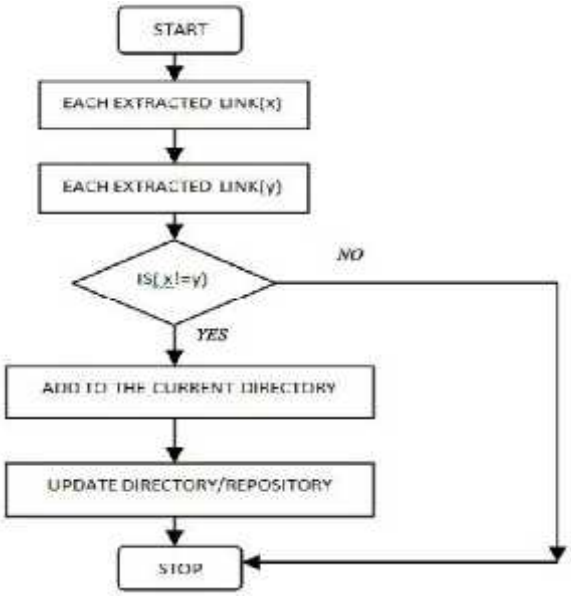

Figure 4: Flowchart to verify the duplicacy of the extracted link in the collection

Human powered directories are popular simply because of the higher quality of links submitted and the caliber of the sites handpicked to be included in the index. The Open Directory has been around since 1999, and is a human-edited directory also known as DMOZ the Open Directory Project purports to be the largest on the Web, constructed and maintained by a vast global community of volunteer editors.

In terms of traffic, these directories are also of little value. The percentage of Web surfers that would prefer browsing categories in a directory over a regular search is quite small; few people tend to dive into the depth of the branchy vertical hierarchy of categories. Thus, your category listing won't bring you much traffic on its own. Because the directories are compiled by humans and for humans the relevancy of directory results is very high.

4. CONCLUSION

Human powered directories are special types of search engines, organized into subject categories. The directories are smaller than most search engines and do not contain full text of the webpage they link to. Such directories are good when a user is interested in a general topic of search. A directory guides and helps user to narrow our search and get refined results. In this paper we present design of a human powered directory using mobile agents, that reduces network traffic substantially.

REFERENCES

1. Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke and Sriram Raghavan, "Searching the Web.", ACM Transactions on Internet Technology, 1(1): August 2001.
2. B.Grossan, “Search Engines : What they are, how they work, and practical suggestions for getting the most out of them”,1997.
3. D. Chess, C. Harrison and A. Kreshenbaum, “ Mobile Agents: Are they a good idea”, Springer Verlag Lecture Notes in Computer Science, pp. 25-47, 1996.
4. Hyacinth S. Nwana, “Software Agents: An Overview”, Knowledge Engineering Review, Vol. 11, No 3, pp.1-40, Sept 1996.

5. J. White, " Mobile Agents" , AAAI Press/MIT Press, pp. 437-472, 1997.
6. Niraj Singhal, Ashutosh Dixit, R. P. Agarwal and A. K. Sharma, "Information Retrieval from the Web and Application of Migrating Crawler", presented in IEEE sponsored International Conference on Computational Intelligence and Communication Systems (CICN-2011) organised by MIR Labs Gwalior India, and published in IEEE Xplore 978-0-7695-4587-5/11 © 2011 IEEE, pp. 476-480, 2011.
7. Niraj Singhal, R. P. Agarwal, Ashutosh Dixit and A. K. Sharma, "Reducing Network Traffic and Managing Volatile Web Contents Using Migrating Crawlers with Table of Variable Information", published in World Applied Sciences Journal, ISSN 1818-4952, Vol.19, No. 5, pp. 666-673, IDOSI Publications, 2012.
8. S. Green, L. Hurst, B. Nangle, P. Cunningham, F. Somers and R. Evans, " Software Agents: A review" . Technical Report, Trinty College Dublin, Ireland, May 1996.
9. Sergey Brin and Lawrence Page. "The anatomy of a large-scale hyper textual Web search engine". Proceedings of the Seventh International World Wide Web Conference, pp. 107—117, April 1998.
10. Qusay H. Mahmoud, "Software Agents: Characteristics and Classification", School of Computing Science, Simon Fraser University Burnaby, BC. Canada V5A 1S6.