REPORT:

'THE MERITS OF INTELLIGENT AGENTS FOR RETRIEVING

ELECTRONIC DOCUMENTS FROM THE INTERNET.'

When the Internet was introduced to the general public, a wealth of information was also opened to them. For anyone using the Internet, finding the exact information that they required was a lot harder to do than it is now. The main reason for this is the use of search engines that look through the electronic documentation that is on the Internet and return to the user documents that contain the keywords that the user had previously inputted into the search engine. Intelligent information agents carry out these searches.

Intelligent agents can be defined as:

"An autonomous, computational software entity that has access to one or more, heterogeneous and geographically distributed information sources, and which pro -actively acquires, mediates, and maintains relevant information on behalf of users or other agents"

In short agents are tools that do jobs on the Internet that have been specified by the programmer or the user. These agents do jobs such as reading your e-mail, speed up your downloads, filter web sites, organise your shopping or even be your assistant on the desktop. The most frequently used agents are the ones used for search engines.

Search engines like Google or AltaVista use basic Information Retrieval (IR) to bring their information together to display to the user to choose from. This method of searching for information is quick and simple to use. It also has the ability to search more than one information database, for example MetaCrawler. They also have the capacity to display their results in "relevance order". This relevance is calculated by working out how often the keywords queried by the user appears in a document, and then shows those documents with the most matches at the top of the list of results. Using the advanced commands that the search engine provides can also search for phrases. This can make the searches quicker and more efficient because searching for the separate words of a phrase can bring up irrelevant information because the whole phrase does not appear in the document.

Even though agents cannot display the exact document you are looking for every time, they can considerably narrow the search for the user. Some find the documents by having a database of documents with information of how many times a word occurs in the document. Whenever a word is searched for, the agent returns an ordered list of documents that have the queried word with the highest occurrence.

The problem with most search agents is that they cannot get over obstacles such as syntax, context or the semantics of the search query. This is apparent when sites deliberately make spelling mistakes to avoid their site being blocked by a filtering agent. The same filters are applied to some search agents and the same problems can arise.

One way that search engines can get ahead of the rest is to be personalized to certain users. The more that a user searches for a topic, the more likely those topics will turn up in general searches, thus making it even more efficient. This would require a type of personal database that could be loaded up to the search engine. Either the database would have to be kept online or the engine would have to be available to download for personal use. Most agents are free for download, so the likeliness of getting revenue from downloading the engine is not high.

Most agents are programmed in Java, JavaScript or Visual Basic, mainly because both languages can be easily integrated into the Windows web browser (Internet Explorer). These are regarded as more basic languages than languages like C or C++. There are more people who are proficient in Java, JavaScript or Visual Basic.

The costs of keeping a search engine running have to be taken into account. With the economy in a "slowdown" (Johnson 2002), ways of keeping the search engine running will have to be explored. Possibilities could be to charge for each search, or charging companies to advertise on the search engine. Owners of documents in the search engine database could also pay to have their site at the top of the list when their topic is searched for.

Agentland.com (2001). *Learn all about Intelligent Agents*. Retrieved on 11th May 2003 from the World Wide Web: http://www.agentland.com/

Geek.com (2002). *PlayStation 3 ... not a "console"?* Retrieved on 13th May 2003 from the World Wide Web:

http://www.geek.com/newsbeta2/geeknews/2002Sep/gee20020910016266.htm

Hermans, B. (1997). *Intelligent Software Agents on the Internet*. Retrieved on 3rd May 2003 from the World Wide Web: http://www.firstmonday.dk/issues/issue2_3/ch_123/

Huhns, M. N. (1997). *AGENTS ON THE WEB*. IEEE INTERNET COMPUTING, pp.80-82.

Jen-Yuan, Y. (2001*). Information agent technology for the Internet: A survey.*

Johnson, D. (2000). *The Future of the Search Engines*. Retrieved on 13th May 2003 from the World Wide Web: http://www.searchengineguide.com/detlev/2002/0213_dj1.html

Lander, E. (2003). *The Future of Search Engines: Are All Eyes on Google?* Retrieved on 10th May 2003 from the World Wide Web:

http://www.topsitelistings.com/searchengines/future_of_google.htm

Ltd, F. i. S. (2003). *Search Engine Services*. Retrieved on 10th May 2003 from the World Wide Web: http://www.futureinsites.co.uk/marketing.jsp

Wang, C.-C. (2000*). XML-Based Dynamic Search Agent for the Internet Information Retrieval.*