Search Engine: an effective tool for exploring the Internet

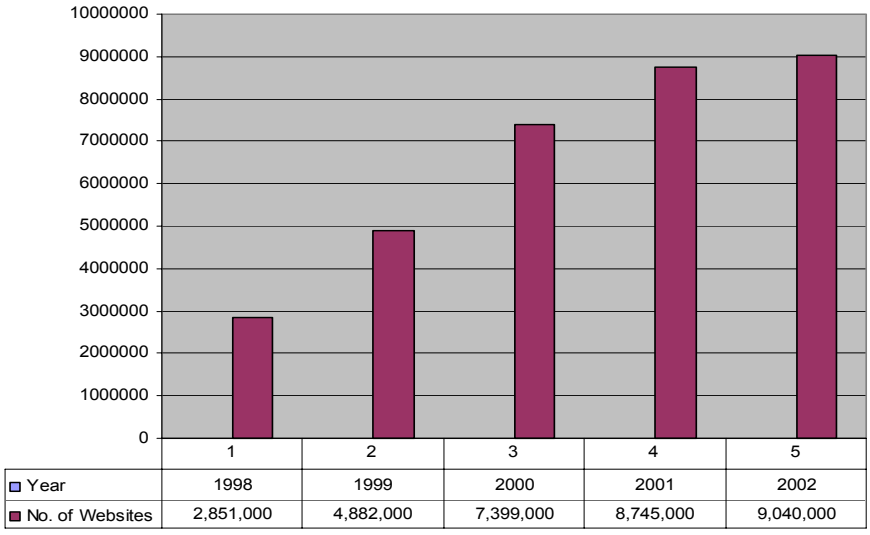
W.M. Tharanga D. Ranasinghe
Assistant Librarian/Visiting Lecturer, University of Kelaniya

Introduction

The Internet has become the largest source of information. Today, millions of Websites exist and this number continuous to grow.

In their annual review (2002) of the World Wide Web, researchers at OCLC have determined that the Web now contains about 9 million unique sites. “The Web continues to grow at a substantial rate”, said Ed O'Neill, manager of the OCLC Web Characterization Project.

Figure 1: Number of Websites



	1	2	3	4	5
Year	1998	1999	2000	2001	2002
No. of Websites	2,851,000	4,882,000	7,399,000	8,745,000	9,040,000

(Source : OCLC)

Table 1: Growth rate of Websites

	1998 - 1999	1999 - 2000	2000 - 2001	2001 - 2002	1998 - 2002
Sites	71%	52%	18%	3%	217%

(Source: OCLC)

According to another survey done by **Netcraft** in August 2006, there are **92,615,362** sites (**92 million**) and this is an increase of 4.4 million sites (4.8%) from the July, 2006 survey.

The end user uses the Internet heavily in accessing information in their day to day needs since the Internet is the biggest repository of knowledge in the history of mankind. Thus the Internet has become the world's widely accessed information source and Websites are added, updated, and obsolete daily.

Therefore, "Finding the right information at the right time" is the challenge in the Internet age. Hence there is a need for a more effective way of retrieving information on the Internet.

There are three ways of finding information on the Internet.

1. Guessing the URL (Uniform Resource Locater)
2. Using a Search engine
3. Using a Subject Directory or a Subject Gateways

If the URL (Uniform Resource Locater) is known, it can be typed in the Internet browsers' (Example: Internet Explorer, Netscape Navigator, etc.) address bar and can be accessed the Website soon.

Otherwise use directories (Yahoo!, dmoz, LookSmart) for broad, general topics. Subject directories select and classify resources into subject categories and subcategories. Some include reviews and/or ratings which can be accessed by keyword search or by browsing the classifications.

A subject gateway can be defined as a facility that allows easier access to network-based resources in a defined subject area. The simplest types of subject gateways are sets of Web pages containing lists of links to resources. The resources accessible through these gateways are reviewed, selected, evaluated and catalogued by information professionals or subject experts.

What is a search engine?

A search engine is a searchable database which collects information on web pages from the Internet, and indexes the information and then stores the result in a huge database where it can be quickly searched. The search engine then provides an interface to search the database.

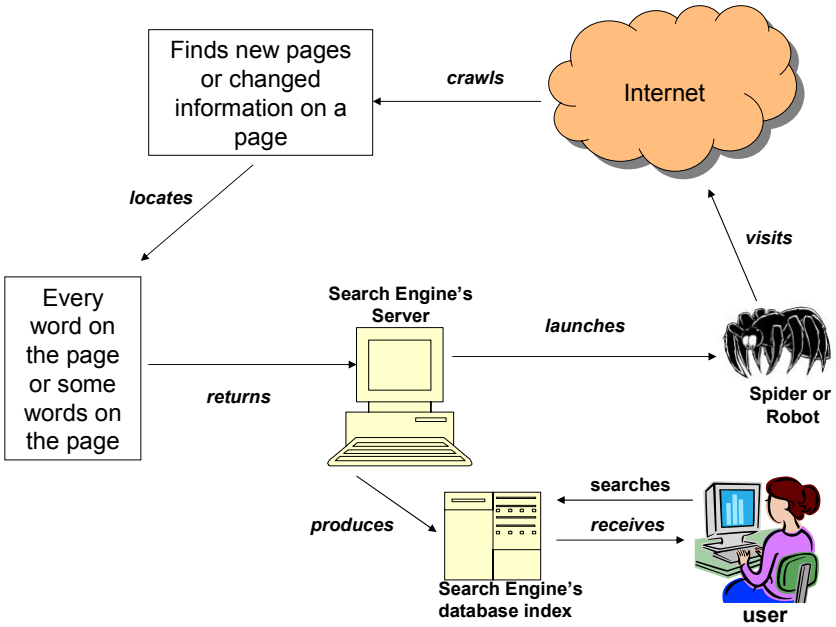
Examples : Google, Alta Vista, Exite

A Search engine has **three** parts.

- **Spider:** Deploys a robot program called a *spider* or *robot* designed to track down web pages. It follows the links these pages contain, and add information to search engines' database. Example: **Googlebot** (Google's robot program)
- **Index:** Database containing a copy of each Web page gathered by the spider.
- **Search engine software :** Technology that enables users to query the index and that returns results in a schematic order.

How does a search engine work?

Figure 2: How does a search engine work?



Types of search engines

In broad sense, search engines can be divided into two categories.

1. **Individual search engines**

   An individual search engine uses a spider to collect its information regarding websites for own searchable index. There are two types of individual search engines.

   i . General search engines
   Examples: Google, AltaVista, HotBot, Lycos

   ii. Subject specific search engines
   Examples: MetaPhys, Chritech, ReligionExplorer, Chordie, ChemFinder

2. **Meta search engines**

   A Meta search engine searches multiple individual engines simultaneously. It does not have its own index, but uses the indexes collected by the spiders of other search engines.

   Example: metacrawler, Ixquick, mamma

Advantages of using search engines

Search engines are best at finding unique keywords, phrases, quotes, and information buried in the full-text of web pages since they normally index WWW documents word by word.

Search engines allow the user to enter keywords, and then they are searched against its database. Users can use advanced search techniques such as phrase searching, truncation/wildcard searching, as well as for Boolean operators (AND, OR, NOT combinations).

With comparison to web directories, search engines are huge databases and contain a large amount of materials. Also, the database is updated at a variable rate.

Conclusion

Search engine is searchable database which allows locating the information on the Internet by submitting the keywords. It is a very useful tool for quickly and easily search for the information Online. It is important to formulate the search statement using advanced searching techniques to filter the most relevant information out of search engines huge database more efficiently and effectively.

References:

1. August 2006 web server survey/ Netcraft. http://news.netcraft.com. Accessed on: 14th August, 2006

2. Muller, Jeanne Froidevaux. **A Librarians' guide to the Internet: searching and evaluating information**/ Jeanne Froidevaux Muller. Oxford: Chandos publishing, 2003

3. Size and growth statistics/ OCLC. http://www.oclc.org/research/projects/archive/wcp/stats/size.htm. Accessed on: 14th August, 2006