E-Content of

INTERNET TECHNOLOGY AND WEB DESIGN

Chapter: 8.10 Search Engine and Meta Search Engine

Topic: 8.10.1 Search Engines

Search Engines

- The **search engine** is a web program that enables the users to enter words and phrases to search, and scan the vast information on the web to locate sites that matches the words or phrase.
- A **web search engine** is a software system that is designed to search for information on the World Wide Web.
- The User can locate useful or interesting web sites by using a search engine.
- This is an organization with a web site containing a huge database of web site addresses.
- Search engines also maintain real-time information by running an algorithm on a web-crawler.
- When a query is typed in form of subject or a name that describes what user are seeking, and the search engine provides with a list or selection of web site addresses that fit the enquiry.
- The User can then simply click on an address to jump to that website.

Search Engine Optimization (SEO)

- Search Engine Optimization is the process of increasing the visibility of a website or a web page in a search engine's natural search results.
- It the process of optimizing the web page contents to increase its relevance to specific keywords and to remove barriers to the indexing activities of search engines.
- SEO makes the website more efficient and visible top its users during the search.

Web Search Engines

- A web search engine is an interactive tool that helps locate documents on the internet containing terms being searched by the user.
- Search engines use the bottom up approach for finding your way around the web.
- You give a search engine, a list of keywords of phrases (called a query) and it returns to you a list of web pages that contain those words or phrases.
- In other words, search engines are actually databases that contain references to thousands of resources.
- Users are able to interact with the database, submitting queries that “ask” the database if it contains resources that match specific criteria.
- Some search engines search only the title of web pages, some by uniform resource locator (URL), some by words in each document in a web site and some by combination of these.
- Each search engine has its own way deciding which of the web pages on its list is most likely to be one that you are looking for.
- Some allow more complicated queries than others, where keywords can be combined with Boolean (logical) operations, such as AND, OR and NOT, to produce rather complicated queries.
- The rules combining these operations are called the syntax of the search engine.
- There is an art in designing queries that result in the search engine returning a useful list to you.
- Given the vast number of web pages, a query that is too general may field literally millions of web pages, most of them useless to you.
- A query that is too specific may miss many web pages that you would have liked to see.
- Each engine is also defined as an on line utility that quickly searches thousands of web documents for a word or phrase being searched.

List of Search Engines

- Google
- Yahoo.
- Ask Jeeves.
- Duck Duck Go.
- Kosmix.
- Yebol.
- Bing
- Msn

FIG 8.6 : Search Engines

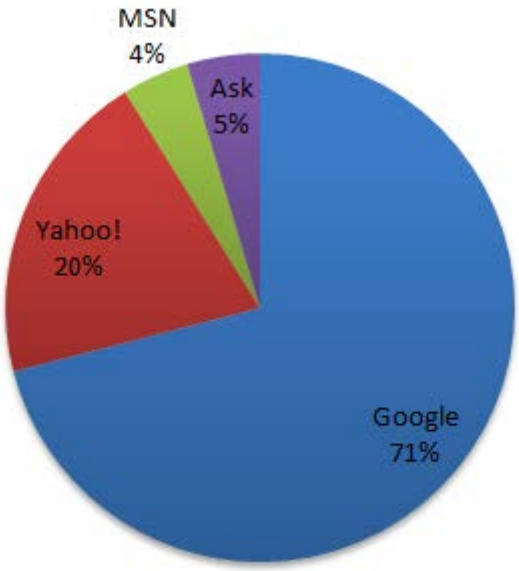

FIG 8.7: Search Engine Usage

Working of Search Engines

- Search engines are information retrieval system designed to help find information stored on a computer system.
- There are many types of search engines but we would limit our discussion to web search engines which search for information on World Wide Web.
- Search engine were started with an idea to ease the process of finding information on the internet.
- A search engine consists of three parts.
- **First part**
  - First part is the spider which is also called the crawler or bot.
  - This spider part visits a web page, reads it and then follow links to other pages within the site.
  - This process is often referred to as 'Crawling' or 'Spidering'.
  - Crawling of a website is done on the very regular basis. The spider visits a website, following links from other website or website submission, it received.

- The content that spider finds is sent to its database or index as it is popularly known.
- This index is like a huge book that contains a copy of web page or cache that the spider finds out.

- **Second part**
  - This constitutes the second part of a search engine. It also stores the structure and the way pages are linked to each other.
  - This information would be updated every time there is a change in content or linking.

- **Third part**
  - Third part of a search engine is search engine software that works behind the interface when we use a search engine.
  - This software will shift through the trillions on indexed pages to match the search query that user has asked.
  - The pages are ranked by search engines and the search results are based on this ranking and relevance to search term.
  - This is how a search engine determines what order shall be listed for a particular search.

Chapter : 8.10 Search and Meta Search Engines

Topic : 8.10.2 Meta Search Engines

Meta Search Engines

- Meta Search Engine is a search tool that sends user requests to several other search engines and/or databases and aggregates the results into a single list or displays them according to their source.

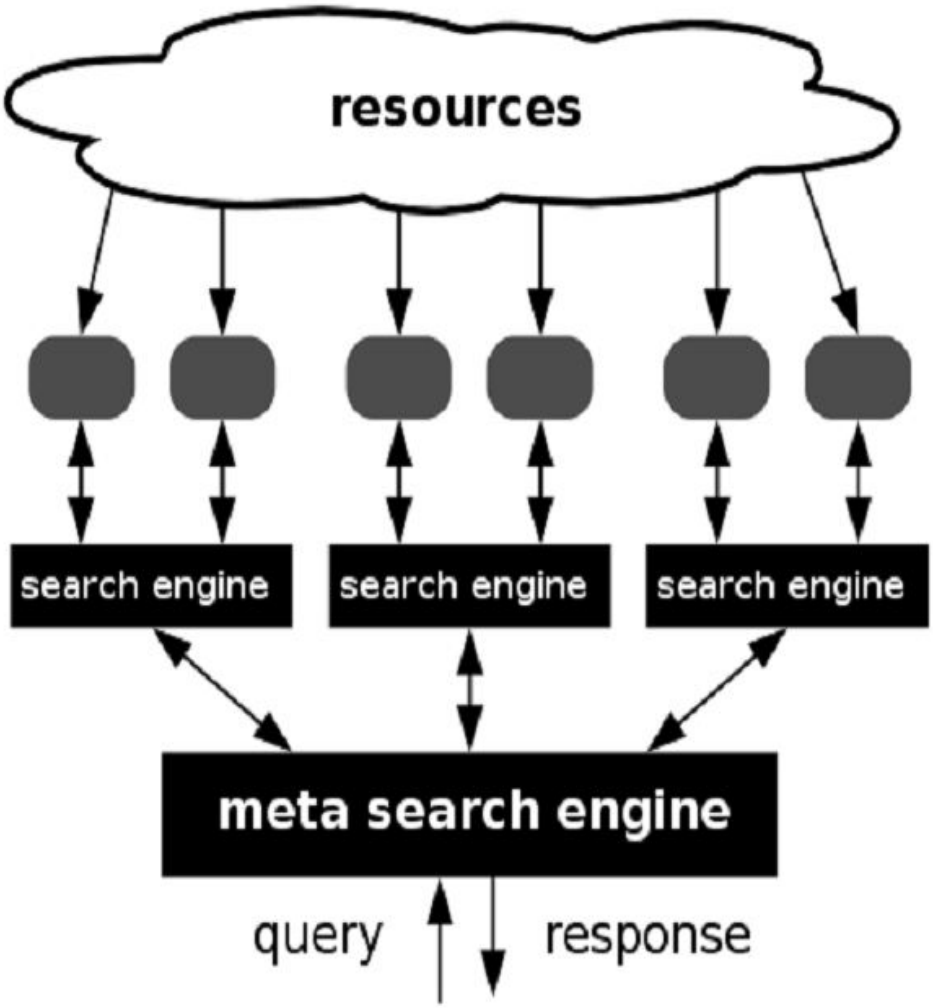

FIG 8.8 : Meta Search Engine

- Meta search engines enable users to enter search criteria once and access several search engines simultaneously.
- Search engines provide fast retrieval of information of interest from the web.
- The problem of knowing where search engines are how to use them poses some difficulties.
- Furthermore, empirical results indicate that only 45% of relevant results will likely be returned by a single search engine that is, each has a recall rate of 45%.
- This limitation is compounded by the fact that the coverage of a typical search engine is between only 5% - 30% of the web.
- Meta search engines are designed to mitigate such problems by accessing multiple individual search engines.
- The System architecture of a meta search engine that contains the following components
  - Query Interface Module – Responsible for getting user's query input.
  - Dispatch Module – Responsible for determining to which search engines a specific query is sent.
  - Knowledge-based Module – Used by the Dispatch module to perform decision-making (optional).
  - Interface Agents Module – Responsible for interacting with different search engines using different query formats.
  - Evaluation Module – Responsible for ranking results according to some predefined evaluation methods (optional).
  - Display Module – Responsible for displaying results.

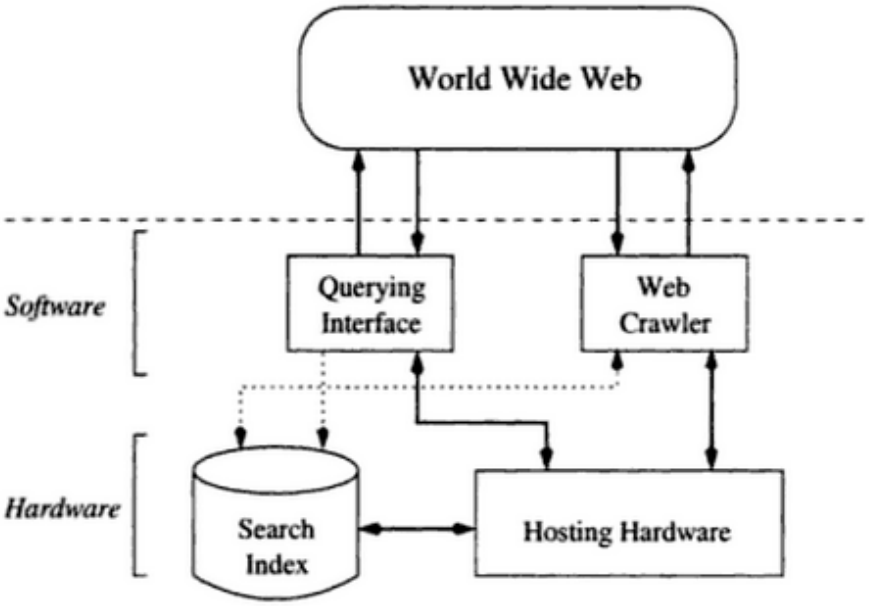

FIG 8.9 : The Architecture of Meta Search Engine

- Meta search engines operate on the premise that the web is too large for any one search engine to index it all and that more comprehensive search results can be obtained by combining the results from several search engines.
- This also may save the user from having to use multiple search engines separately. The process of fusion also improves the search result.

Web Meta Search Engines

- A Web Meta Searcher also known as mega indexes is a tool that helps users locate information available via the World Wide Web.
- Meta Searches do not have their own database. Instead they have access to other primary search engines.
- Web Meta searchers provide a single interface that enable users to search many different search engines, indexes and databases simultaneously.
- There are a number of web searchers available.
- The content of search engines, indexes and databases vary, the same query typed into several search engines is likely to produce different results.

- When searching a topic, users often want to see results from various sources.
- One way to compare the results of several search engines is to type and retype a query into individual search engines one at a time. This can be very time consuming.
- A Meta searcher helps to make this task more efficient by providing a single interface where the query is typed in one and results can be obtained from multiple search engines.
- Meta searchers are different from other search engines and indexes in the following respects.
- Single search engine and indexes provide a collection or database of resources that can be queried.
- Meta searchers do not provide a database i.e., these search engines do not collect web pages, do not accept URL and do not classify or review web sites.
- Instead they provide service to search the database of several other search engines at the same time to locate web pages that matches query given by the user.
- Example of web meta searchers include
  - All in one search
  - clnet
  - Cyperlands Web
  - CUSI
  - Electric Library
  - Eureka
  - Savvy Search
  - ProFusion
  - MetaCrawler