A
Seminar
on

"Search Engine"

Guided By
Asst. Prof. N.L.Sharma
MBM Engg. College

By
Govind Salvi
Information Technology
Roll no 2471

Overview

What is Search?

What does it means by Search Engine?

Search Engine Basics

How Search Engine Works?

Conclusions

References

What is Search and Search Engine?

Search basically means find some thing.
Example: when we are trying to find some word in a dictionary we are actually searching for that particular word in the dictionary.

The one who performs this searching function is known as Search Engine. On this particular case Search Engine is Human.

When we talk about System terminology the machine (Program) which Perform this searching operation is called Search Engine.

The two basic type of Search Engines are..
a) Database search Engine
b) Web Search Engine

Database Search Engine

Database search engine perform the search process by Pattern Matching when the user gives a particular query.

Example:-

Query   SELECT name,class,roll_no FROM student WHERE roll_no > 100;

When we give above query to Database search Engine it perform the pattern matching with the algorithm that selects name class roll_no form table student that satisfy condition roll_no >100
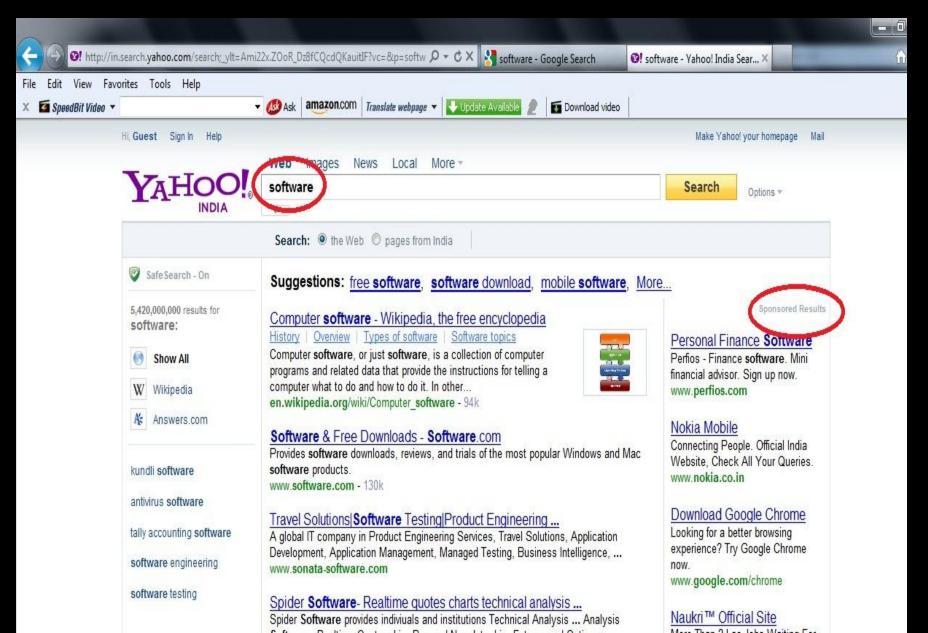
Web Search Engine

A software program that helps users find information stored on a personal computer, or a network of computers, such as the Internet. A user enters search terms and the search engine retrieves a list of World Wide Web sites, personal computer files, or documents.

Types of Web Search Engine

a. Crawler Based Search Engine

b. Human Powered Search Engine

c. Hybrid Search Engine

Visualization

Search Engine Basics

Meta Tag:-A special HTML tag that provides information about a Web page. they provide information such as who created the page, how often it is updated, what the page is about, and which keywords represent the page's content.

Spider (crawlers):- A program that automatically fetches Web pages. Spiders are used to feed pages to search engines. It's called a spider because it *crawls* over the Web. Another term for these programs is WebCrawler.

How Search Engine Works ?

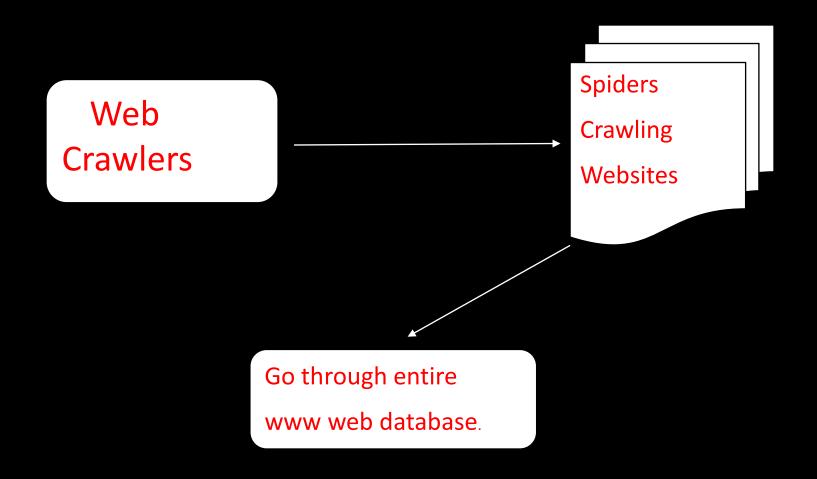The following steps are a part of the Google Search Framework:
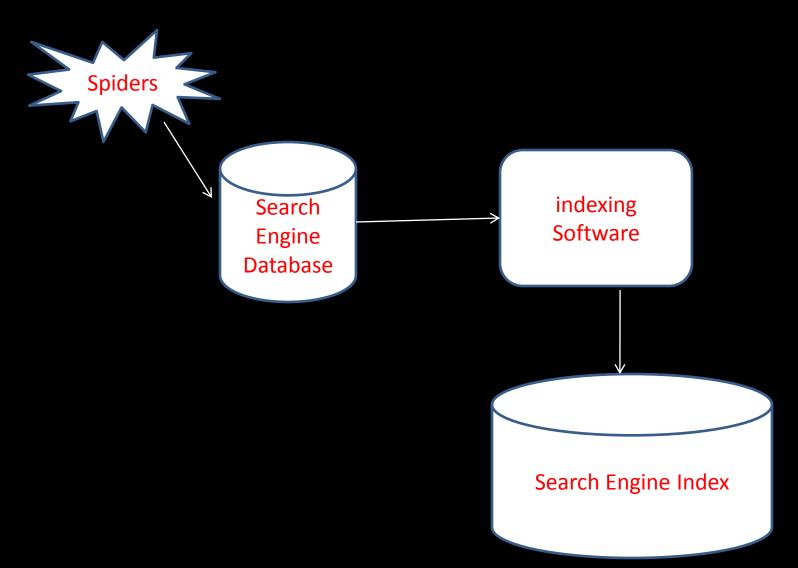
a.   Crawling

b. Indexing
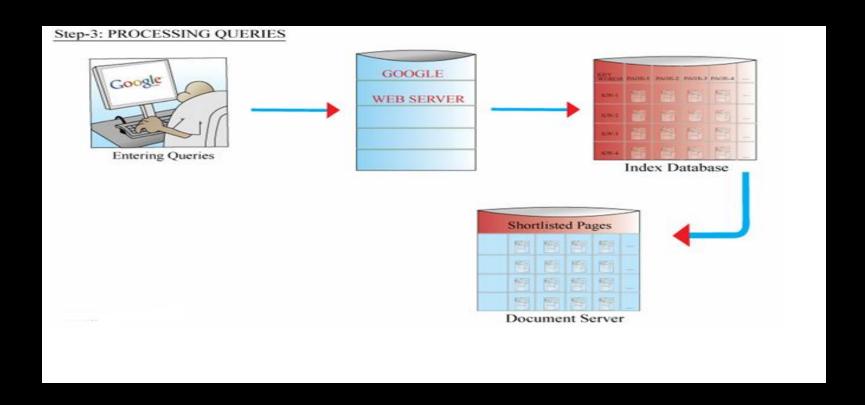
c. Query Processing

d. Ranking

Crawling: Data of the new & updated content from all the web pages on the web. Crawlers / Spiders / Robots /Bots are Search Engine specific software that collects data from the Web.
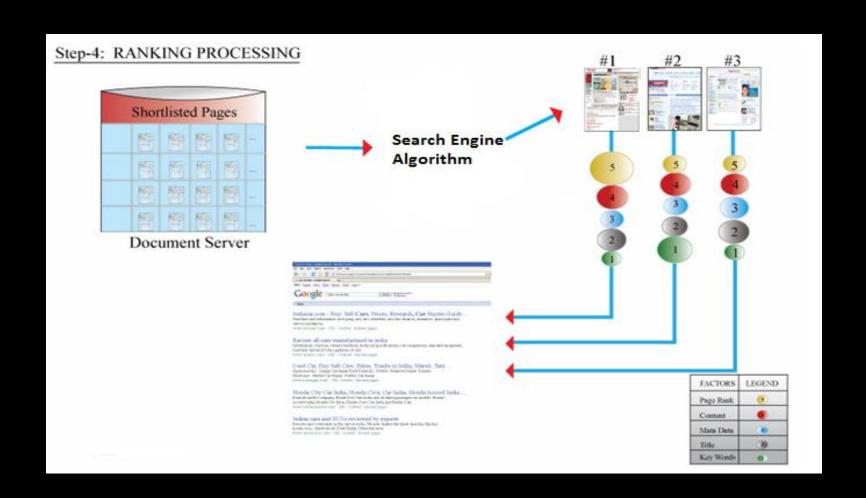
Web Crawlers → Spiders Crawling Websites

Go through entire www web database.

Indexing: This step is required for converting the unstructured page data to a structured format so that it can be used efficiently for Query processing.

Spiders

Search Engine Database

indexing Software

Search Engine Index

Query Processing: This step takes the User query (Keyword) as its input, which the user enters into the SEARCH BOX. The outcome of this step is the short listed pages from trillions of pages and which are relevant for this query.

Step-3: PROCESSING QUERIES

Entering Queries → GOOGLE WEB SERVER → Index Database → Shortlisted Pages / Document Server

Ranking: This is the final step of the search process. Google algorithm comes into picture here

TF-IDF Algorithm

TF Term Frequency

How many times a word appears in a document

IDF Inverse Document Frequency

Logarithm of (Number of documents /
                number of documents containing the term)

TF-IDF of a keyword in a page = TF * IDF

Example

100 web pages. Keyword: *Mbm Engg. College*
#1 has 8 mentions. TF = 8.
#2, 17, 19, 76 have 4 mentions. TF = 4.
20 pages have 1 mention. TF = 1.
IDF = $\log_2 (100 / 25) = 2$

TF-IDF of *Mbm Engg. College* in #1= 8 * 2=16High!

TF-IDF of *Mbm Engg. College* in #2, 17, 19, 76 = 4 * 2 = 8   Not so high

TF-IDF of *Mbm Engg. College* in 20 others = 1 * 2 = 2   Small

TF-IDF of *Mbm Engg. College* in all the rest = 0 * 2 = 0   Irrelevant

Conclusions

Search engine plays important role in accessing the content over the internet, it fetches the pages requested by the user

It made the internet and accessing the information just a click away.

The need for better search engines only increases

References

Web Links

a. Wikipedia.   Till seminar
b. www.telezent.com   1-may-2011
c. http://searchenginewatch.com/   24-april-2011
d. http://homeforprofits.com/search-engines/working-of-a-search-engine/
15-april-2011
a. http://computer.howstuffworks.com/internet/basics/   15-april-2011

Books

Computer Networking Top down Approach by Kurose and Ross. 3 edition

Thank You!