

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

The screenshot shows a web browser window with multiple tabs. The active tab is 'http://in.search.yahoo.com/search;_ylt=Ami22x.ZOoR_Dz8fCQcdQKauitIF?vc=&p=softw'. The browser's address bar also shows 'software - Google Search' and 'software - Yahoo! India Search...'. The browser's menu bar includes 'File', 'Edit', 'View', 'Favorites', 'Tools', and 'Help'. The toolbar contains icons for 'SpeedBit Video', 'Ask', 'amazon.com', 'Translate webpage', 'Update Available', and 'Download video'. The Yahoo! India logo is visible on the left, and the search bar contains the word 'software'. The search results are displayed in a grid layout. The left sidebar shows 'Safe Search - On' and '5,420,000,000 results for software:'. The main content area shows 'Suggestions: free software, software download, mobile software, More...'. The first suggestion is 'Computer software - Wikipedia, the free encyclopedia', which is circled in red. The second suggestion is 'Software & Free Downloads - Software.com'. The third suggestion is 'Travel Solutions|Software Testing|Product Engineering ...'. The fourth suggestion is 'Spider Software- Realtime quotes charts technical analysis ...'. The right sidebar shows 'Sponsored Results' (circled in red) with links to 'Personal Finance Software', 'Nokia Mobile', 'Download Google Chrome', and 'Naukri™ Official Site'.

File Edit View Favorites Tools Help

SpeedBit Video Ask amazon.com Translate webpage Update Available Download video

Hi Guest Sign In Help Make Yahoo! your homepage Mail

Web Images News Local More

YAHOO! INDIA software Search Options

Search: the Web pages from India

Safe Search - On

5,420,000,000 results for software:

Show All

Wikipedia

Answers.com

kundli software

antivirus software

tally accounting software

software engineering

software testing

Suggestions: free software, software download, mobile software, More...

Computer software - Wikipedia, the free encyclopedia

History Overview Types of software Software topics

Computer software, or just software, is a collection of computer programs and related data that provide the instructions for telling a computer what to do and how to do it. In other...

en.wikipedia.org/wiki/Computer_software - 94k

Software & Free Downloads - Software.com

Provides software downloads, reviews, and trials of the most popular Windows and Mac software products.

www.software.com - 130k

Travel Solutions|Software Testing|Product Engineering ...

A global IT company in Product Engineering Services, Travel Solutions, Application Development, Application Management, Managed Testing, Business Intelligence, ...

www.sonata-software.com

Spider Software- Realtime quotes charts technical analysis ...

Spider Software provides individuals and institutions Technical Analysis ... Analysis Software, Realtime Quotes, Business Intelligence, Financial Data, ...

Sponsored Results

Personal Finance Software

Perfios - Finance software. Mini financial advisor. Sign up now.

www.perfios.com

Nokia Mobile

Connecting People. Official India Website, Check All Your Queries.

www.nokia.co.in

Download Google Chrome

Looking for a better browsing experience? Try Google Chrome now.

www.google.com/chrome

Naukri™ Official Site

More Than 2 Lacs Jobs Waiting For You

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.

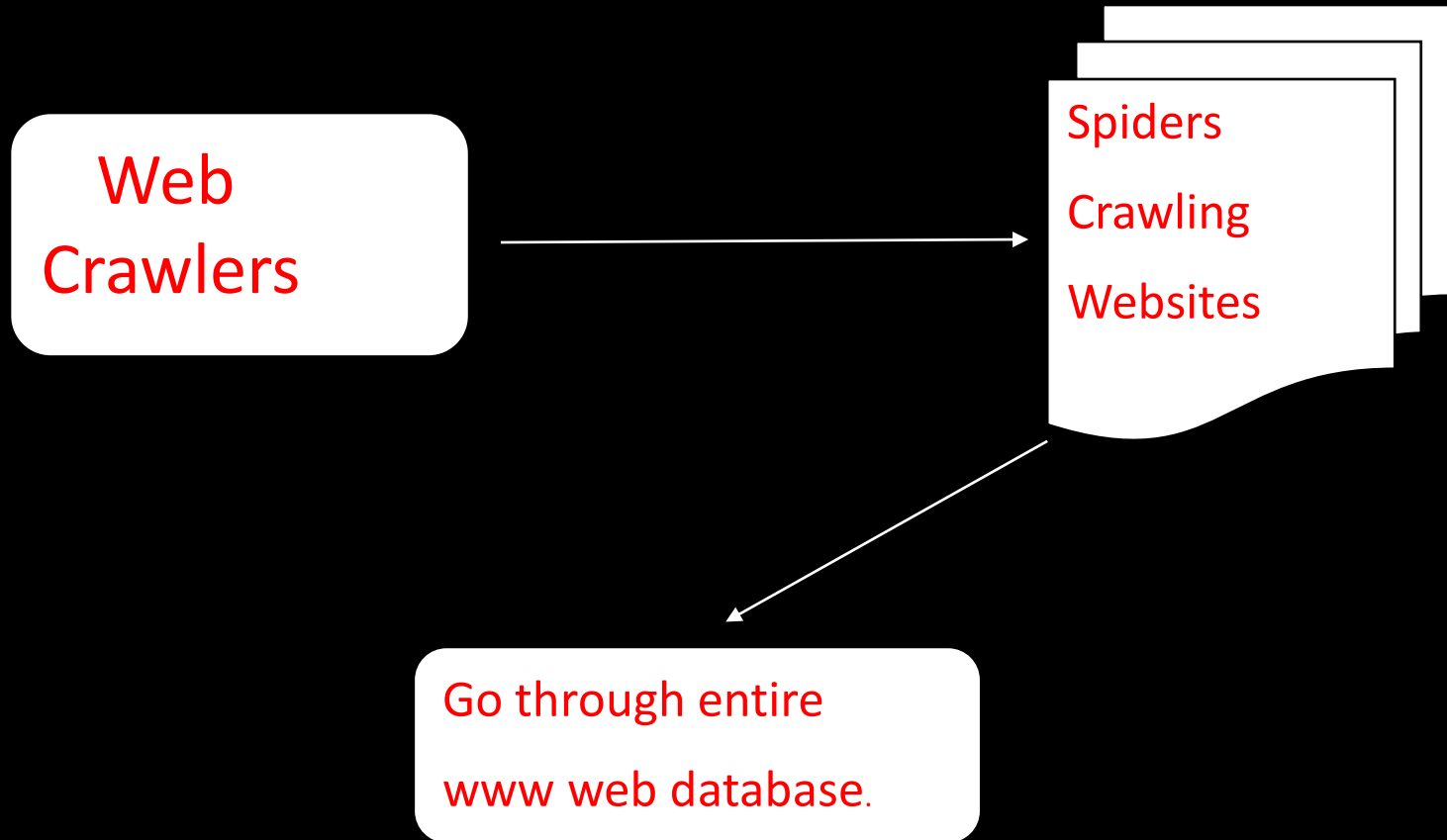
Spider (crawlers):- A program that automatically fetches Web pages. Spiders are used to feed pages to search engines. It's called a spider because it *crawls* over the Web. Another term for these programs is WebCrawler.

How Search Engine Works ?

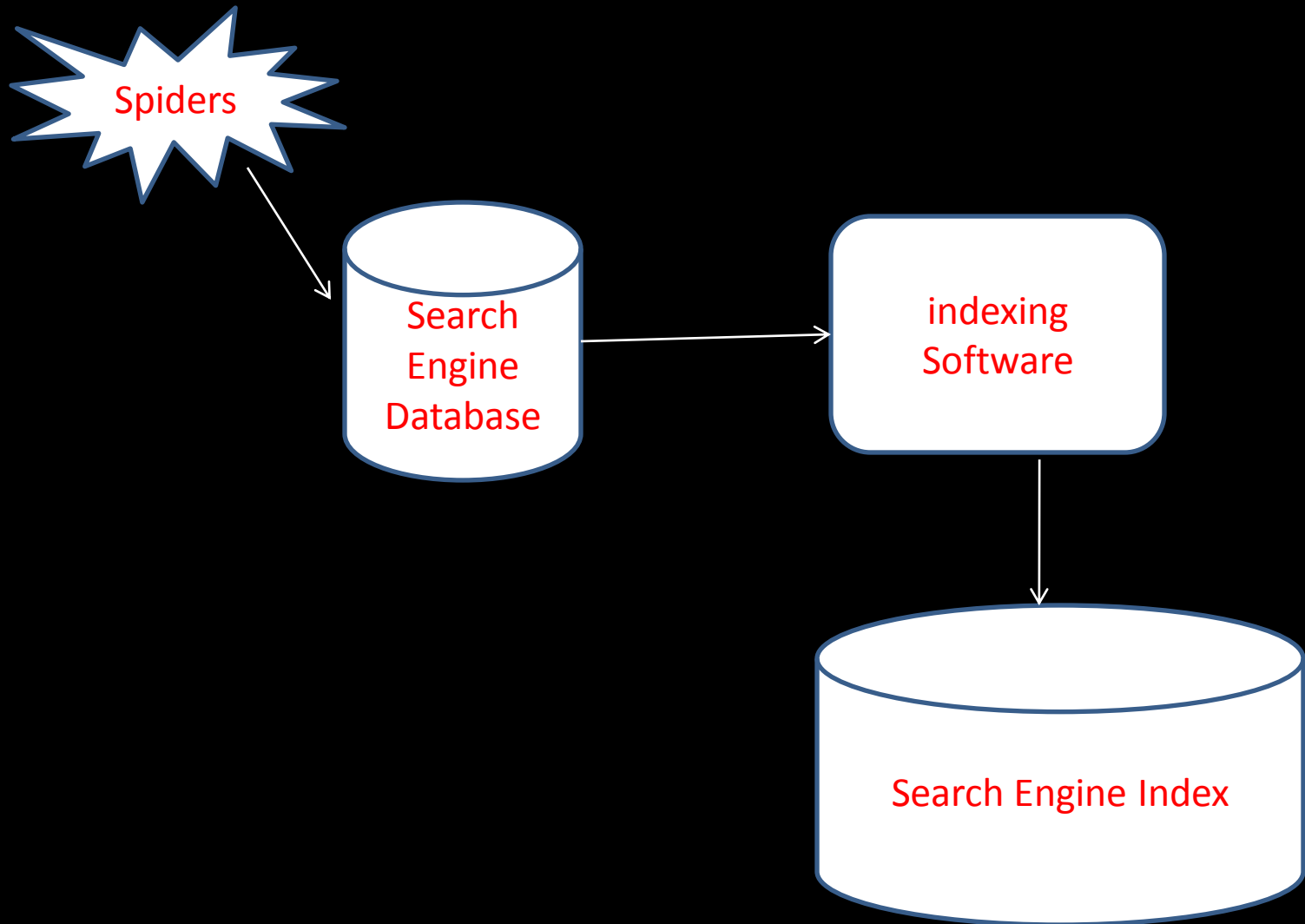
The following steps are a part of the Google Search Framework:

- a. Crawling**
- b. Indexing**
- c. Query Processing**
- d. Ranking**

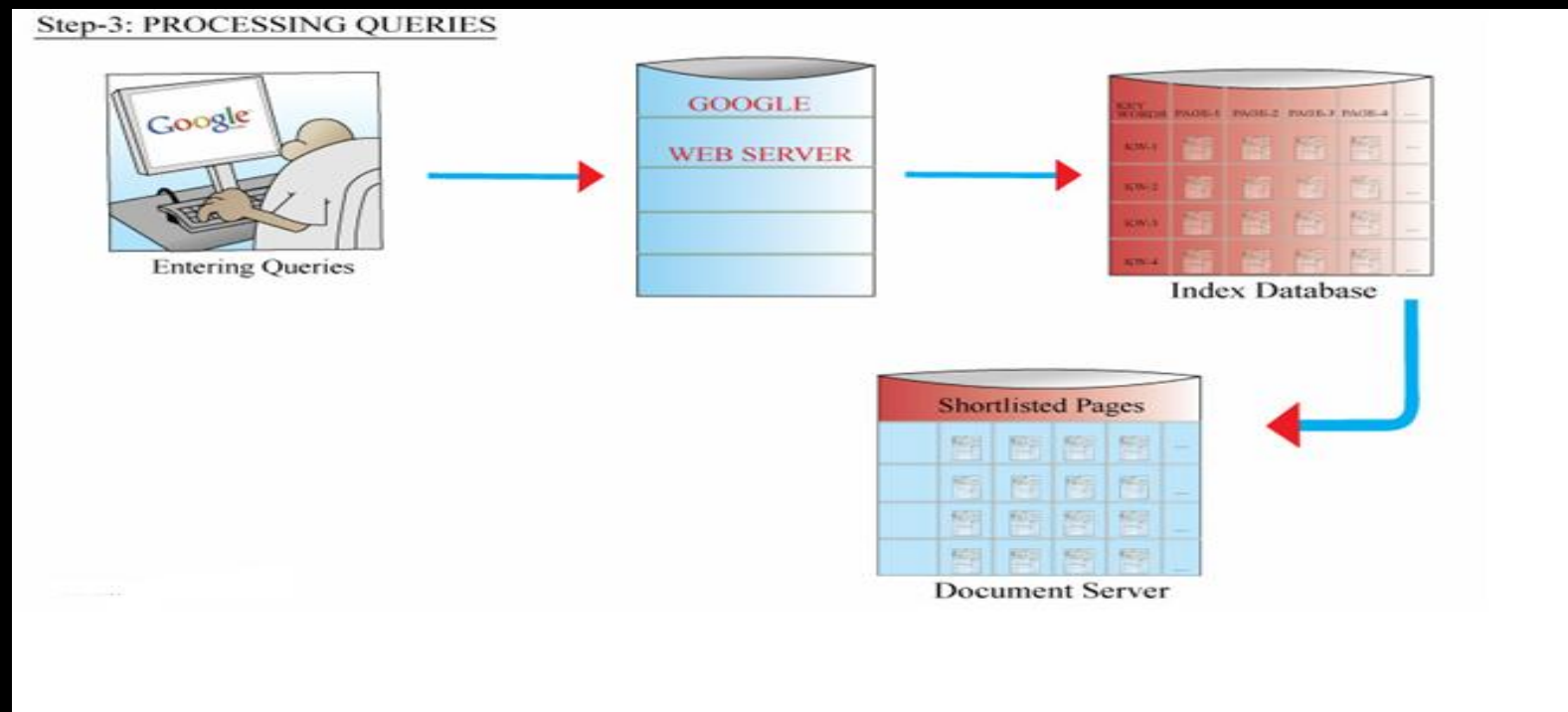
Crawling: Data of the new & updated content from all the web pages on the web. Crawlers / Spiders / Robots /Bots are Search Engine specific software that collects data from the Web.



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.



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.

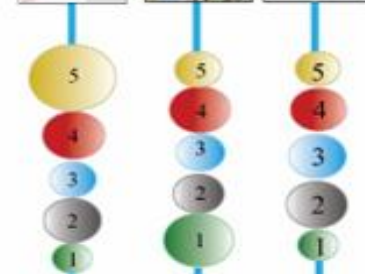


Ranking: This is the final step of the search process. Google algorithm comes into picture here

Step-4: RANKING PROCESSING



Search Engine Algorithm



FACTORS	LEGEND
Page Rank	5
Content	4
Meta Data	3
Title	2
Key Words	1

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 = $8 * 2 = 16$ High!

$TF-IDF$ 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!**