Anatomy and Working of Search Engines

Almost daily we use search engines and get the desired result but have you ever wondered what actually happens behind the scene? How is the data stored? Where is the data stored?  Although it is a complex and tedious process, we are going to explain some of the important terminologies used in the search engines and the basic process involved.  Data is stored in a number of files on the internet, as ASCII files, binary files or in the databases. Search engines may vary on the way the data is stored. If the data is stored in the database the same can be queried easily to create search engines. For HTML files, graphics & PDFs search engine is an additional program.

 A search engine that does not have a given content, searches it elsewhere. This data comes from a program that crawls many pages and reads the contents. Such a program is called a **ROBOT** or a **SPIDER**. It crawls the URLs specified by the search engine and marks when a new one is found. When user searches, he is not actually searching the contents. Instead, he is searching an index of the content, the spider has found.

The processes invloved are:

### 1. Simple data queries:

### 2. Complex data queries:

### 3. Boolean searching

### 4. Pre-processed data

### **5. Indexing content:**

### 7. Noise words:

### 6. Document index

When searching a website, chances are that you may not be actually searching the content, but rather a pre-formatted copy of the content.

## Major Data Structures used to store data by Search engines:

1.Big files

2.repository

3.lexicon

4.Document index

5.Hit list

6.forward index

7.inverted index

This paper explains how much we are dependent on search engines, the role it plays in our day today life and how search engine eases the work of people.