SEARCH ENGINE

Search Engine..**Ok.**

But what is a ..

Prepared: -

Nisarg Patel

(090010107016)

INDEX

SEARCH ENGINE

Abstract:-

❖ *Definition :-*

- Basically, a search engine is a software program that searches for sites based on the words that you designate as search terms. Search engines look through their own databases of information in order to find what it is that you are looking for.

- A Search Engine is a program that allows you to search the Internet for information. There are many search engines on the World Wide Web.

- You might have heard of search engines like *Google*, *Yahoo!*, or *MSN*. These are the most popular search engines. There are lots of other excellent search engines on the Internet that you may never have heard of!

- Let's look at the different types of search engines available on the Internet.Allows a user to enter words which characterise a required page.Returns links to pages which match the query. A page on the web connected to a backend program

WHAT IS SEARCH ENGINE??

- A search engine is an Internet tool that locates web pages and sorts them according to specified keywords. A computer program that can contact other network resources on the Internet, search for specific information by keywords ,and report the results ;example is google.
- Search engines use automated software programs (spider, crawler, robot) to crawl the WWW by following links. These software programs download web pages into the search engine's index (a database). Often every word on a web page will be indexed. Search Engine results are organized in a way that is determined by a special algorithm to rank the results so that the the "best" results listed first. Search engines and directories are not the same thing; although the term "search engine" often is used interchangeably.Search engines automatically create web site listings by using spiders that "crawl" web pages, index their information, and optimally follows that site's links to other pages.

Web Search Engine:-

- Web search engine is a tool designed to search for information on the World Wide Web. The search results are usually presented in a list and are commonly called *hits*. The information may consist of web pages, images, information and other types of files.
- Eg:- Google.com, Yahoo!, Altavista.com, Excite.com

- Although search engine is really a general class of programs, the term is often used to specifically describe systems like Google, Alta Vista and Excite that enable users to search for documents on the World Wide Web and USENET newsgroups. Some search engines also mine data available in databases or open directories. Unlike Web directories, which are maintained by human editors, search engines operate algorithmically or are a mixture of algorithmic and human input. It is a program Web Search Engine that searches documents for specified keywords and returns a list of the documents where the keywords were found.

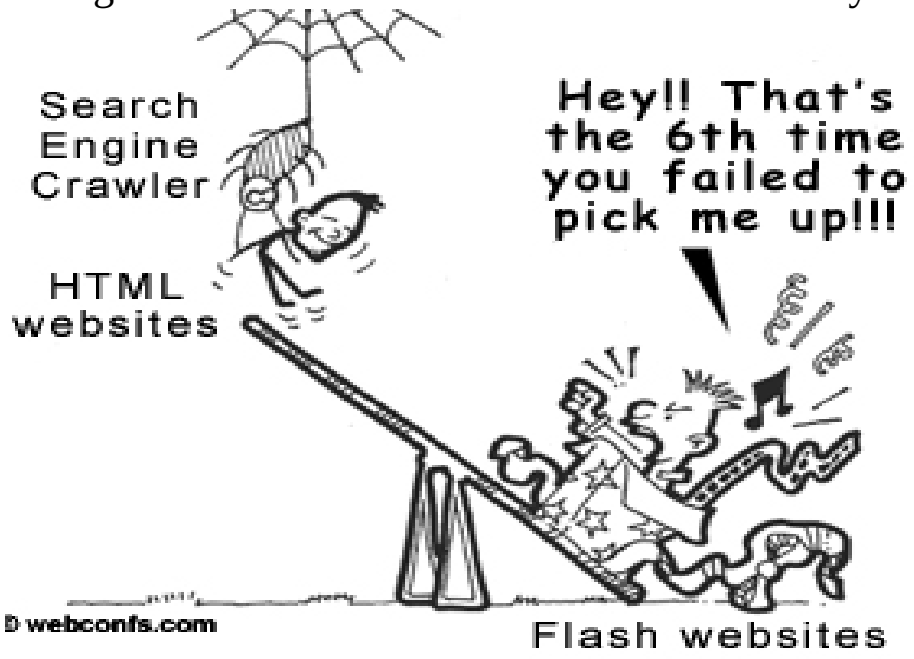
SEO (Search Engine Optimization):-

> Used by companies to get a higher result in search engines

- Black hat:- Using illegal techniques to trick the search engine, like paying sites to link to you.
- White hat:- Using legitimate techniques
- Whenever you enter a query in a search engine and hit 'enter' you get a list of web results that contain that query term. Users normally tend to visit websites that are at the top of this list as they perceive those to be more relevant to the query. If you have ever wondered why some of these websites rank better than the others then you must know that it is because of a powerful web marketing technique called Search Engine Optimization (SEO).
- SEO is a technique which helps search engines find and rank your site higher than the millions of other sites in response to a search query. SEO thus helps you get traffic from search engines.

How it works?

- Typically, a search engine works by sending out a *spider* to fetch as many documents as possible. Another program, called an *indexer,* then reads these documents and creates an index based on the words contained in each document.
- Each search engine uses a proprietary algorithm to create its indices such that, ideally, only meaningful results are returned for each *query*. They include incredibly detailed processes and methodologies, and are updated all the time. This is a bare bones look at how search engines work to retrieve your search results. All search engines go by this basic process when conducting search processes, but because there are differences in search engines, there are bound to be different results depending on which engine you use.
- The searcher types a query into a search engine. Search engine software quickly sorts through literally millions of pages in its database to find matches to this query. The search engine's results are ranked in order of relevancy.



- The first basic truth you need to know to learn SEO is that search engines are not humans. While this might be obvious for everybody, the differences between how humans and search engines view web pages aren't. Unlike humans, search engines are text-driven. Although technology advances rapidly, search engines are far from intelligent creatures that can feel the beauty of a cool design or enjoy the sounds and movement in movies. Instead, search engines crawl the Web, looking at particular site items (mainly text) to get an idea what a site is about. This brief explanation is not the most precise because as we will see next, search engines perform several activities in order to deliver search results – *crawling*, *indexing*, *processing*, *calculating relevancy*, and *retrieving*.
- First, search engines crawl he Web to see what is there. This task is performed by a piece of software, called a *crawler* or a *spider* (or Googlebot, as is the case with Google). Spiders follow links from one page to another and index everything they find on their way. Having in mind the number of pages on the Web (over 20 billion), it is impossible for a spider to visit a site daily just to see if a new page has appeared or if an existing page has been modified, sometimes crawlers may not end up visiting your site for a month or two.
- What you can do is to check what a crawler sees from your site. As already mentioned, crawlers are not humans and they do not see images, Flash movies, JavaScript, frames, password-protected pages and directories, so if you have tons of these on your site, you'd better run the Spider Simulator below to see if these goodies are viewable by the spider. If they are not viewable, they will not be spidered, not indexed, not processed, etc. - in a word they will be non-existent for search engines.

Ranking Results:-

- Now that the best matches for your search query have been decided, the search engine has to rank them in the most appropriate order. It uses a complex algorithm to run calculations on every match to decide which is the most relevant. They are then sorted out for you in order of most relevant to least relevant.

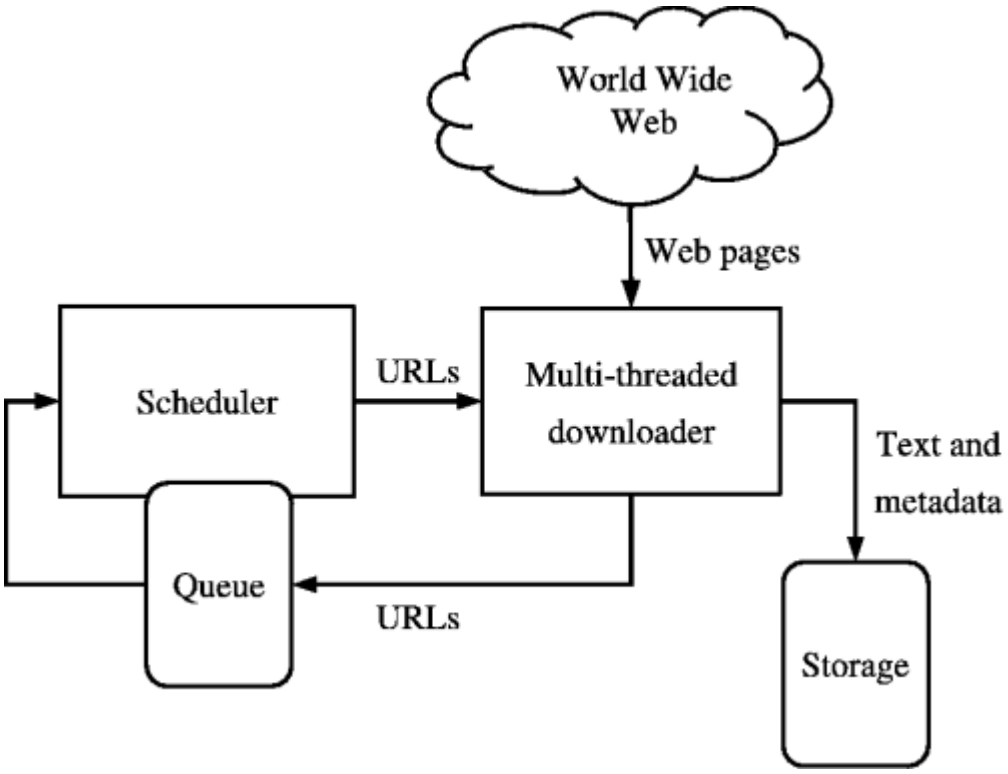
Example of Search Engines:-

- Google is always a safe bet for most search queries, and most of the time your search will be successful on the very first page of search results.
- Yahoo is also a great choice, and finds a lot of stuff that Google does not necessarily pick up. There are quite a few search engines that will help you do this with clustered results or search suggestions. Some of these include Clusty, WiseNut, AOL Search and Teoma**,** in addition to Gigablast ,AllTheWeb and SurfWax. Images on the Web are easy to find, especially with targeted image search engines such as Picsearch, Ditto, and of course, Google has some fantastic image search capabilities. You can also check out my list of Image Search Engines-Directories-Collections, or Clip Art-Buttons-Graphics- Icons-Images on the Web.
- There's so much multimedia on the Web that your main problem will be finding enough time to look at it all. Here are a few places you can use to search for sounds, movies, and music on the Web: Loomia, Torrent Typhoon, The Internet Movie Database, SingingFish, and Podscope. For even more multimedia search engines and sites

Operation of Search Engine:-

- A search engine operates, in the following order
  - Web crawling
  - Indexing
  - Searching

Web Crawling:-



- A Web crawler is a computer program that browses the World Wide Web in a methodical, automated manner. Other terms for Web crawlers are *ants*, *automatic indexers*, *bots*, and *worms* or *Web spider*, *Web robot*, or especially in the FOAF (an acronym of Friend of a friend) community *Web scutter*.This process is called *Web crawling* or *spidering*.

- Many sites, in particular search engines, use spidering as a means of providing up-to-date data.
- Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches. Crawlers can also be used for automating maintenance tasks on a Web site, such as checking links or validating HTML code. crawlers can be used to gather specific types of information from Web pages, such as harvesting e-mail addresses (usually for spam).
- A Web crawler is one type of bot, or software agent. In general, it starts with a list of URLs to visit, called the *seeds*. As the crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the list of URLs to visit, called the *crawl frontier*. URLs from the frontier are recursively visited according to a set of policies.

Indexing:-

- Search engine indexing collects, parses, and stores data to facilitate fast and accurate information retrieval. Index design incorporates interdisciplinary concepts from linguistics, cognitive psychology, mathematics, informatics, physics and computer science. An alternate name for the process in the context of search engines designed to find web pages on the Internet is Web indexing.
- Popular engines focus on the full-text indexing of online, natural language documents. Media types such as video and audio and graphics are also searchable.Meta search engines reuse the indices of other services and do not store a local index, whereas cache-based search engines permanently store the index along with the corpus.

Search:-

- A web search query is a query that a user enters into web search engine to satisfy his or her information needs. Web search queries are distinctive in that they are unstructured and often ambiguous; they vary greatly from standard query languages which are governed by strict syntax rule

Use search engine to your advantage:-

- Search engines can help you to identify sources that will provide serious information, products or services, or entertainment. Pick the right search engine for your research needs.
- Yahoo and AltaVista will help you to distinguish between different categories of web sites. refine your search whenever possible.

Types of search engines:-

- Conventional(librarycatalog).
  Search by keyword, title, author, etc.
- Text-based(Lexis-Nexis,Google,Yahoo!).
  Search by keywords. Limited search using queries in natural language.
- Multimedia(QBIC,WebSeek,SaFe)
  Search by visual appearance (shapes, colors,…).
- Question answering systems (Ask,NSIR,Answerbus)
  Search in (restricted) natural language
- Clustering systems (Vivísimo, Clusty)
- Research systems (Lemur, Nutch)

Advantages:-

- Search engine provide some popular ways of finding information.
- Search engine do have the ability to have refined or more precise results.
- Search engine within a websites allow you to search information only on that website.

Disadvantages:-

- Search engine such as google allow users to allow users to remove URLS after the websites owner has removed the offending information, which deletes it from index.
- Sometimes Privacy problem about information.
- Increasing time taken(meta find search engine)

CONCLUSION:-

- Search engines are full of useful data, an analysis of which can yield surprising and valuable insights. Each time analyze a paper or a patent, or understand a new term,we beg in to see how a search engine can create new parameters and detection algorithms to combat webspam. Some search engines have already began to offer fee-based premium search services that contain no advertising. If this is the trend, it may eventually change people's view of Internet search engines as a free resource for fair information. Use site look for all pages. Follow standards for search forms.

REFERNCES:-

- www.wikipedia.org
- www.pdfcloud.net
- www.searchengineguide.com
- http:/ / www.ask.com
- S. Brin, “The Anatomy of a Large-Scale Hypertextual Web Search Engine” (1998).
- http://www.webopedia.com/TERM/S/search_engine