ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
September 2013,
Vol. 6, No. (3):
Pgs. 341-349

Search Engine: A Review

S.K.VIJAY[1], MANISH MAHESHWARI[2] and ROOHI ALI[2]

[1]Department of Computer Science, Barkatullah University, Bhopal, India.
[2]Department of Computer Application,Makhanlal Chaturvedi, National University, Bhopal.

(Received: August 10, 2013; Accepted: September 05, 2013)

ABSTRACT

Searching online has become part of the everyday lives of most people. Whether to look for information about the latest gadget to getting directions to a popular trend, most people have made search engines part of their daily routine. Beyond trivial applications, search engines are increasingly becoming the sole or primary source directing people to essential information. For this reason, search engines occupy "a prominent position in the online world"; they have made it easier for people to find information among the billions of web pages on the Internet. Due to the large number of websites, search engines have the complex task of sorting through the billions of pages and displaying only the most relevant pages in the search engine results page (SERP) for the submitted search query. With the continued growth of the Internet and the amount of websites available, it has become increasingly difficult for sites looking for an audience to achieve visibility. There are millions of new websites appearing on the Internet every month. As a result of this continued growth, it has made it increasingly difficult for websites to stay visible among all the other competing sites.

Key words: Searching information, application.

INTRODUCTION

During the mid-nineties the Web started experiencing a tremendous growth in both, number of users and number of websites. With the continued exponential growth of the Internet, it became apparent the need for classification of the content of the Internet. As result, search engines and Web directories started to appear in the early 1990s to organize pages and to make it easy for people to find information online. Right from Archie, Veronica, Excite, Lycos, AltaVista, Infoseek, and Yahoo to Google all are search engines. They index millions of sites on the WWW, so that Web surfers can easily find Websites with the information needed. By making indexes, or large databases of Web sites (based on titles, keywords, and the text in the pages), search engines can locate relevance to Web sites when users enter search terms or phrases.

This paper describes the application of selected Search Engine Optimization (SEO) techniques for a website and analyzes its effectiveness. It covers the various aspects of search engine and the effect its techniques have on the number of users who visit the site. Search engine architecture, working, approaches, types, classification, rankings are also analyzed.

What is Search Engine?

A search engine is a coordinated set of routines that includes:

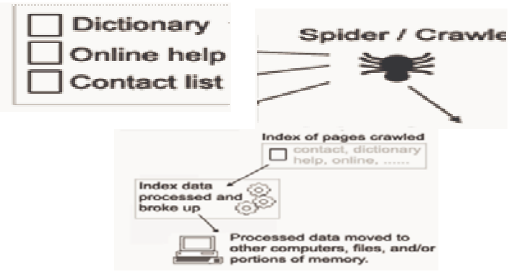
"A spider ("crawler" or a "bot") according to the inbuilt algorithm ,method goes to every page or representative pages on every site that is to be searched and reads it, using hypertext links on each page to discover and read website's other pages. It can also view as a program that creates a bulky index ("catalog") from the pages that have been searched. It will receive search request, compares it to all the entries in the index, and returns results to you."

Search Engine History

A software program or script available through the Internet that searches documents and files for keywords and returns the results of any files containing those keywords. Today, there are millions of various search engines available on the Internet, each with their own abilities and features. The very first search engine ever developed is considered Archie, which was used to search for FTP files only and the first text-based search engine is considered as Veronica. Today, the most popularly used and well known search engine is Google.

Architectural Framework

As large search engines have millions and sometimes billions of pages, so many search engines not only search the pages but also display the results according to their importance. This importance is basically determined by using various algorithms and methods.


□ Dictionary
□ Online help
□ Contact list
Spider / Crawle
Index of pages crawled
contact, dictionary help, online, ...
Index data processed and broke up
Processed data moved to other computers, files, and/or portions of memory.

Above figure shows, the architecture of how a search engine works. In the figure, the starting point of all search engines is a spider or crawler (indicated by a symbol), which visits the pages that will be included in the search phrase and grabs the contents of each of those pages needed.

Once a page has been crawled in the data contained within the page is processed, this sometimes involves stripping out stop words, grabbing the location of each of the words in the page, the frequency of their occurrences, links to other pages, images/figures, etc. This data is used to rank the page and is the primary method used by search engine to determine if a page is shown or not and in what order it will come.

Finally, when the data has been processed it is often broken up into one or more files, moved to different destinations (computers or servers), or loaded into memory where it can be accessed when users perform a search for information.

Some Popular Search Engines
Third-party Description
Bing

Microsoft search engine offers many of the features as other search engines, while adding new features such as going more in-depth with product searches, flight information, and image searches.

Google

The most popular and well known search engine on the Internet.

Yahoo

Another well known and popular search engine.

WolframAlpha

A search engine that enables a user to get more intelligent results and statistical information.

Different Search Engine Approaches

´ Major search engines index the content of a large portion of the WWW and provide results to the user. E.g.: Google, Yahoo (which uses Google), AltaVista, and Lycos.

´ Specialized content search engines uses a selective criterion about what part of the Web is crawled and indexed. They selectively index only the related sites about these products and provide a shorter but more focused result. E.g.: CRM applications.

´ On the other hand some others provide a general search of the WWW but allow entering a search request in natural language, such as a normal English sentence. E.g.: Ask Jeeves.

´ Specific tools and some major sites a number of search engines at the same time and compile results in a single list. E.g.: Yahoo.

´ Larger corporate sites or individual sites may use a search engine to index the content of their own sites diplomatically. It is now a practice to license or sell search engines for use by major search engine companies.

Search Engine Terminology

Different types of search

Boolean search

A search that allows the inclusion or exclusion of documents containing certain words through the use of Boolean operators AND, NOT and OR.

Concept search

A search for documents related conceptually to a word, rather than specifically containing the word itself.

Fuzzy search

A search that will find matches even when words are only partially spelled or misspelled.

Keyword search

A search for documents containing one or more words that are specified by a user.

Phrase search

A search for documents containing exact sentence or phrase specified by a user.

Proximity search

A search where users to specify that documents returned should have the words near each other.

Other terminology of search engine

Search Engine Optimization (SEO)

"SEO is the science of customizing elements of website to achieve the best possible search engine ranking" when a web user searches on a keyword.

Search Engine Results Page (SERP)

The page that displays a list web page based on the user's search query. "The results normally include a list of web pages with titles, a link to the page, and a short description showing where the keywords have matched content within the page. A SERP may refer to a single page of links returned, or to the set of all links returned for a search query".

Ranking

The position of the webpage within the search engine results page (SERP).

PageRank

The proprietary search ranking algorithm used by Google Search "that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of 'measuring' its relative importance within the set";this numerical weight (its PageRank value) indicates the importance or authority of the web page, and it's also a determining factor of a page's ranking on the search results.

Pay-per-click (PPC)

it's an advertising model where search users are sent to the advertiser's page via paid listings. Each time a user clicks on any of the paid listings, the advertiser pays a certain amount for each click. An example of a PPC program is Google Adwords.

Web crawler / bot

a program that is "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" for users searching for information online.

Indexed pages

Search engine crawlers collect, parse and store web page data in the index database for use by the search engine to display on the search results. Once a web page data gets stored in the search engine index, the page has been indexed.

Keyphrase / Keyword / Search query

These terms are used interchangeably; it is the word or set of words that a web user enters into the search engine text box for searching.

Inbound links / Backlinks / External links

These terms are used interchangeably; these are links from other sites that point (or link) to your website.

Long-tail keywords

These search queries that contain three or more words; a very specific search for which there is less competition. For example, search queries such as "*roses"* and "*red roses for mother's day"*; the latter would be considered a *long-tail keyword* because it's more specific.

Web-based Content Management System (CMS)

it's "a bundled or standalone application used to create, manage, store, and deploy content on Web pages" such as video, text, images, etc... Examples of web-based CMS platforms are Drupal, Joomla and Wordpress.

Search Engine Classification

Search engines uses different indexing strategies and therefore are inherently different.

Crawler-based search engines

Create their listings automatically by using a piece of software to "crawl" or "spider" the web and then index what it finds to build the search base.
E.g.: Google, AltaVista.

Crawler-based search engines are good for a specific search topic and can be very efficient in finding relevant information in this situation. However, when the search topic is general, crawler-base search engines may return hundreds of thousands of irrelevant responses to simple search requests, including lengthy documents in which your keyword appears only once.

Crawler-based search engines have **three** major components.

The crawler

Also called the spider. The spider visits a web page, reads it, and then follows links to other pages within the site. The spider will return to the site on a regular basis, such as every month or every fifteen days, to look for changes.

The index

Everything the spider finds goes into the second part of the search engine, the index. The index will contain a copy of every web page that the spider finds. If a web page changes, then the index is updated with new information.

The search engine software

This is the software program that accepts the user-entered query, interprets it, and sifts through the millions of pages recorded in the index to find matches and ranks them in order of what it believes is most relevant and presents them in a customizable manner to the user.

Human-powered directories

Depend on human editors to create their listings. Typically, webmasters submit a short description to the directory for their websites, or editors write one for the sites they review, and these manually edited descriptions will form the search base. Therefore, changes made to individual web pages will have no effect on how these pages get listed in the search results.

E.g. Yahoo directory, Open Directory and LookSmart, Human-powered directories are good for general topic of search. In this situation, a directory can guide and help you narrow your search and get refined results. Therefore, search results found in a human-powered directory are usually more relevant to the search topic and more accurate. However, this is not an efficient way to find information when a specific search topic is in mind.

Meta-search engines

Transmit user-supplied keywords

simultaneously to several individual search engines to actually carry out the search. Search results returned from all the search engines can be integrated, duplicates can be eliminated and additional features such as clustering by subjects within the search results can be implemented by meta-search engines.
E.g. Dogpile, Mamma, and Metacrawler,

Meta-search engines are good for saving time by searching only in one place and sparing the need to use and learn several separate search engines. “But since meta-search engines do not allow for input of many search variables, their best use is to find hits on obscure items or to see if something can be found using the Internet.”

Different types of the major search engines falling in different categoriesSearch Engines

Types	
Google	Crawler-based search engine
AllTheWeb	Crawler-based search engine
Teoma	Crawler-based search engine
Inktomi	Crawler-based search engine
AltaVista	Crawler-based search engine
LookSmart	Human-Powered Directory
Open Directory	Human-Powered Directory
Yahoo	Human-Powered Directory, also provide crawler-based search results powered by Google
MSN Search	Human-Powered Directory powered by LookSmart, also provide crawler-based search results powered by Inktomi
AOL Search	Provide crawler-based search results powered by Google
AskJeeves	Provide crawler-based search results powered by Teoma
HotBot	Provide crawler-based search results powered by AllTheWeb, Google, Inktomi and Teoma, "4-in-1" search engine
Lycos	Provide crawler-based search results powered by AllTheWeb
Netscape Search	Provide crawler-based search results powered by Google

The Limits of Search Engine Technology

Basically all of the main search engines operate on the same principles automatic searching and crawling the WWW, follow hyperlinks and index content in the databases. Artificial intelligence is used for this. Technical limitations and problems are always there in both inclusion and rankings. Here are the most common of them:

Spidering and Indexing Problems

´ Search engines aren’t good at completing online forms (login), and thus any content contained behind them may remain hidden.

´ Websites using a CMS (Content Management System) often create duplicate versions of the same page - a major problem for search engines looking for completely original content.

´ Errors in a website’s crawling directives (robots.txt) may lead to blocking search engines entirely.

ˇ Poor link structures lead to search engines failing to reach all of a website’s content.

Interpreting Non-Text Content

´ Although the engines are getting better at reading non-HTML text, content in rich media format is traditionally difficult for search engines to parse.

´ This includes text in Flash files, images, photos, and video, audio & plug-in content.

Content to Query Matching

´ Text that is not written in common terms that people use to search.

´ Language and internationalization subtleties.

´ Location targeting.

´ Mixed contextual signals.

The “Tree Falls in a Forest”

The “tree falls in a forest” postulates translates perfectly to search engines and web content i.e. if no one links to your content, the search engines may choose to ignore it.

The engines by themselves have no inherent gauge of quality and no potential way to discover fantastic pieces of content on the web. Only humans have this power - to discover, react, comment and link to. Thus, great content cannot simply be created - it must be shared and talked about. Search engines already do a great job of promoting high quality content on websites that have become popular, but they cannot *generate* this popularity - this is a task that demands talented Internet marketers i.e. SEO.

SEO isn't just about getting the technical details of search-engine friendly web development correct. It's also about marketing. This is perhaps the most important concept to grasp about the functionality of search engines. You can build a perfect website, but its content can remain invisible to search engines unless you promote it. This is due to the nature of search technology, which relies on the metrics of relevance and importance to display results.

Search Engine Optimization

In mid 1990's When search marketing began ,manual submission, the meta keywords tag and keyword stuffing were all are commonly used and necessary to rank well. In 2004, link bombing with anchor text, buying hordes of links from automated blog comment spam injectors and the construction of inter-linking farms of websites could all be leveraged for traffic. In 2011, social media marketing and vertical search inclusion are mainstream methods for conducting search engine optimization.

The future is uncertain, but in the world of search, change is a constant. For this reason, search marketing will remain a steadfast need for those who wish to remain competitive on the web. As websites compete for attention and placement in the search engines and those with the best knowledge and experience with these rankings receive the benefits of increased traffic and visibility that is what SEO and SEM is used for.

What is Search Engine Optimization (SEO)?

SEO is the practice of improving and promoting a web site in order to increase the number of visitors the site receives from search engines. There are many aspects to SEO, from the words on your page to the way other sites link to you on the web. Sometimes SEO is simply a matter of making sure your site is structured in a way that search engines understand.

Search Engine Optimization isn't just about "engines." It's about making site better for people too.

Why does website need SEO?

The majority of web traffic is driven by the major commercial search engines - Google, Bing and Yahoo!. Although social media and other types of traffic can generate visits to any website, search engines are the primary method of navigation for most Internet users. This is true whether site provides content, services, products, information or just about anything else.

Search engines are unique in that they provided targeted traffic - people looking for what is offered. Search engines are the roadways that makes this happen. If site cannot be found by search engines or content cannot be put into their databases, it will miss out on incredible opportunities available to websites provided via search.

Search queries, the words that users type into the search box, carry extraordinary value.

Experience has shown that search engine traffic can make (or break) an organization's success. Targeted visitors to a website can provide publicity, revenue, and exposure like no other channel of marketing. Investing in SEO, whether through time or finances, can have an exceptional rate of return compared to other types of marketing and promotion.

How SEO Operates?

Search engines have two major functions - crawling & building an index, and providing answers by calculating relevancy & serving results.

Crawling and Indexing

Each stop is its own unique document (usually a web page, but sometimes a PDF, JPG or other file). The search engines need a way to "crawl" the entire city and find all the stops along the way, so they use the best path available – links.

Crawling and Indexing

Crawling and indexing the billions of documents, pages, files, news, videos and media on the World Wide Web.

Providing Answers

Providing answers to user queries, most frequently through lists of relevant pages, through retrieval and rankings.

The link structure of the web serves to bind all of the pages together." Through links, search engines' automated robots, called "crawlers" or "spiders" can reach the many billions of interconnected documents.

Once the engines find these pages, they next decipher the code from them and store selected pieces in massive hard drives, to be recalled later when needed for a search query. To accomplish the monumental task of holding billions of pages that can be accessed in a fraction of a second, the search engines have constructed datacenters all over the world.

These monstrous storage facilities hold thousands of machines processing large quantities of information. After all, when a person performs a search at any of the major engines, they demand results instantaneously – even a 1 or 2 second delay can cause dissatisfaction, so the engines work hard to provide answers as fast as possible.

Providing Answers

Search engines are answer machines. When a person looks for something online, it requires the search engines to scour their corpus of billions of documents and do two things – first, return only those results that are relevant or useful to the searcher's query, and second, rank those results in order of perceived usefulness. It is both "relevance" and "importance" that the process of SEO is meant to influence.

To a search engine, relevance means more than simply finding a page with the right words. In the early days of the web, search engines didn't go much further than this simplistic step, and their results suffered as a consequence. Thus, through evolution, smart engineers at the engines devised better ways to find valuable results that searchers would appreciate and enjoy. Today, 100s of factors influence relevance, many of which we'll discuss throughout this guide.

How Do Search Engines Determine Importance?

Currently, *the major engines typically interpret importance as popularit*y – the more popular a site, page or document, the more valuable the information contained therein must be. This assumption has proven fairly successful in practice, as the engines have continued to increase users' satisfaction by using metrics that interpret popularity.

Popularity and relevance aren't determined manually. Instead, the engines craft careful, mathematical equations – algorithms – to sort the wheat from the chaff and to then rank the wheat in order of tastiness (or however it is that farmers determine wheat's value).

These algorithms are often comprised of hundreds of components. In the search marketing field, we often refer to them as "ranking factors"

Advantages and Disadvantages of SEO

Advantages

It's cost-effective

If well ranked, company has greater chance at becoming visible around the world. Internet is an effective marketing tool. However, there is no guarantee that business will boom immediately but it sure will, little by little.

1. Once website gets a good place. This implies that it is properly valued in the SEO community. In reality, which ranked first in their field are perceived to be really good.
2. Need the money for search engine optimization is established, regardless of the number of hits. This saves money. Plus, no confusion with the accounting part.

Disadvantages

1. Your position in the search engine is unpredictable. Everything depends on the algorithm.
2. This is time consuming. It takes a long time

to improve a notch.

3. Participants can make use of the black hat tactics. These unfair practices hinder the tree of your company. You control not about other people's heads. They can choose for unethical strategies and you, your company could affect credibility and negative.
4. It may mean the end of it all for entrepreneurs whose website are located on the last few pages of the search engine.

Summary

Search engines are different on whether they are crawler-based search engines or human-powered directories. Although major search engines tend to become hybrid search engines providing both types of results, they still favor one type of results over another as their main results. A meta-search engine is a search engine that searches the search engines. Each type of search engines work well for certain types of search tasks but not for all. Considering only crawler-based search engines, they differ on their crawling features, the indices they build (mainly on sizes) and the search engine software they use to search against the search base. Search engine software is implemented differently to support different search commands or interpret the same commands differently. Different search engine software also present the search results in different manners (mostly user-customizable) and rank matches based on different ranking algorithms.

Knowing about the differences between search engines will help the users find the right place to go to when having a certain search task to accomplish. Learning different search engines by comparing them will also help to make full use of them to explore the Internet to a greater extent.

While SEO and SEM are used to make a website the right choice for more and more visitors, it helps in crafting the website popular and relevant.

Summarizing, the top trends of SEO in 2013 includes The rapid increase in the importance of mobile optimization, The death of old, black-hat SEO webspam tactics, Longer, richer content is ranking better than shorter content, Social media marketing is imperative to a successful SEO campaign, Integrating multimedia within text-based content improves and enhances content rankings, Google authorship is playing a major role in search engine rankings and click-through rates.

REFERENCES

1. 28 November 2000. "Introduction to Search Engines." *Yutrend.com*. Available online http://www.yutrend.com/topics/basics/index.php? clanak= seaeng & rubrika= beginners
2. Danny Sullivan. 14 October 2002. "How Search Engines Work." *Search Engine Watch. com* . Available online http://searchenginewatch.com/webmasters/article.php/2168031
3. Danny Sullivan. 29 April 2003. "Major Search Engines and Directories." *Search Engine Watch. com* . Available online http://searchenginewatch.com/links/article.php/2156221
4. "Meta-Search Engines." *UC Berkeley - Teaching Library Internet Workshops*. Available online http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch. html
5. "Introduction to Search Engines." *Kansas City Public Library*. Available online http://www.kclibrary.org/resources/search/intro.cfm
6. Danny Sullivan. 5 December 2002. "Search Engine Features For Webmasters." *SearchEngineWatch.com* . Available online http://searchenginewatch.com/webmasters/article.php/2167891
7. Danny Sullivan. 2 September 2003. "Search Engine Sizes." *SearchEngineWatch.com* . Available online http://searchenginewatch.com/reports/article.php/2156481
8. Laura Cohen. September 2003. "Boolean Searching on the Internet." *University Libraries of University at Albany*. Available

online http://library.albany.edu/internet/boolean.html

9. Danny Sullivan. 26 October 2001. "Power Searching For Anyone." *SearchEngineWatch.com* . Available online http://searchenginewatch.com/facts/article.php/2156031
10. Danny Sullivan. 26 October 2001. "Search Features Chart." *Search EngineWatch.com* Available online http://searchengine watch.com /facts/article.php/2155981
11. Greg R. Notess. 31 December 2002. "Search Engine Statistics: Relative Size Showdown." *SearchEngineShowdown.com* . Available online http://www.searchengineshowdown.com/stats/size.shtml
12. Danny Sullivan. 31 July 2003. "How Search Engines Rank Web Pages." *earchEngineWatch.com* . Available online http://searchenginewatch.com/webmasters/article.php/2167961