History and Web Search Engines Works

[1]Saurabh Kumar, [2] Maneesh Kumar, [3] Vipul Agarwal
M .Tech Scholar, Department of Computer Science and Engineering,
Dr. A.P.J Abdul Kalam Technical University, Lucknow (U.P),

Abstract: -- A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages(SERPs). The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

I.   INTRODUCTION

Search engine is a web software program or web based script available over the Internet that searches documents and files for keywords and returns the list of results containing those keywords. Today, there are numbers of different search engines available on the Internet, each with their own techniques and specialties. Search Engine Optimization is a technique to improve visibility of a website in search engine.

II.HISTORY

Timeline (full list)		
Year	Engine	Current status
1993	W3Catalog	Inactive
	Aliweb	Inactive
	JumpStation	Inactive
	WWW Worm	Inactive
1994	WebCrawler	Active, Aggregator
	Go.com	Inactive, redirects to Disney
	Lycos	Active
	Infoseek	Inactive
1995	AltaVista	Inactive, redirected to Yahoo!
	Daum	Active
	Magellan	Inactive
	Excite	Active
	SAPO	Active
	Yahoo!	Active, Launched as a directory
1996	Dogpile	Active, Aggregator
	Inktomi	Inactive, acquired by Yahoo!
	HotBot	Active (lycos.com)
	Ask Jeeves	Active (rebranded ask.com)
1997	Northern Light	Inactive
	Yandex	Active

Year	Engine	Status		Engine	Status
1998	Google	Active		GoodSearch	Active
	Ixquick	Active also as Startpage		SearchMe	Inactive
	MSN Search	Active as Bing	2006	Soso	Active
	empas	Inactive (merged with NATE)		Quaero	Inactive
1999	AlltheWeb	Inactive (URL redirected to Yahoo!)		Search.com	Active
	GenieKnows	Active, rebranded Yellowee.com		ChaCha	Active
	Naver	Active		Ask.com	Active
	Teoma	Inactive, redirects to Ask.com		Live Search	Active as Bing, Launched as rebranded MSN Search
	Vivisimo	Inactive	2007	wikiseek	Inactive
2000	Baidu	Active		Sproose	Inactive
	Exalead	Active		Wikia Search	Inactive
	Gigablast	Active		Blackle.com	Active, Google Search
2003	Info.com	Active	2008	Powerset	Inactive (redirects to Bing)
	Scroogle	Inactive		Picollator	Inactive
2004	Yahoo! Search	Active, Launched own web search (see Yahoo! Directory, 1995)		Viewzi	Inactive
				Boogami	Inactive
	A9.com	Inactive		LeapFish	Inactive
	Sogou	Active		Forestle	Inactive (redirects to Ecosia)
2005	AOL Search	Active		DuckDuckGo	Active

2009	Bing	Active, Launched as rebranded Live Search
	Yebol	Inactive
	Mugurdy	Inactive due to a lack of funding
	Scout (Goby)	Active
	NATE	Active
2010	Blekko	Inactive, sold to IBM
	Cuil	Inactive
	Yandex	Active, Launched global (English) search
2011	YaCy	Active, P2P web search engine
2012	Volunia	Inactive
2013	Halalgoogling	Active, Islamic / Halal filter Search
2013	Egerin	Active, Kurdish / Sorani Search engine

III.WORKING

A search engine maintains the following processes in near real time:

1. Web crawling
2. Indexing
3. Searching[14]

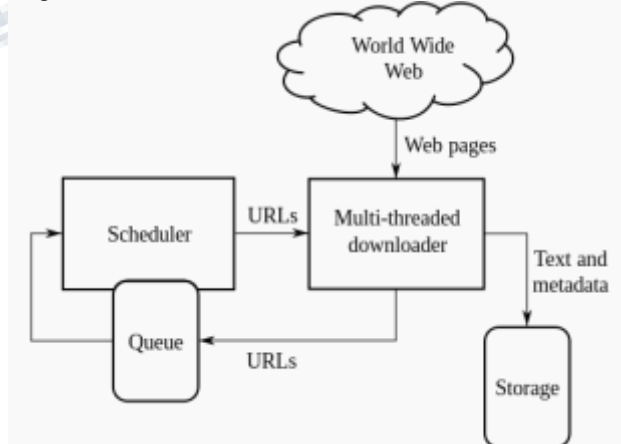Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename robots.txt, addressed to it, before sending certain information back to be indexed depending on many factors, such as the titles, page content, JavaScript, Cascading Style Sheets (CSS), headings, as evidenced by the standard HTML markup of the informational content, or its metadata in HTML meta tags.

Indexing means associating words and other definable tokens found on web pages to their domain names and HTML-based fields. The associations are made in a public database, made available for web search queries. A query from a user can be a single word. The index helps find information relating to the query as quickly as possible.[14]

Some of the techniques for indexing, and cacheing are trade secrets, whereas web crawling is a straightforward process of visiting all sites on a systematic basis.

Between visits by the spider, the cached version of page (some or all the content needed to render it) stored in the search engine working memory is quickly sent to an inquirer. If a visit is overdue, the search engine can just act as a web proxy instead. In this case the page may differ from the search terms indexed.[14] The cached page holds the appearance of the version whose words were indexed, so a cached version of a page can be useful to the web site when the actual page has been lost, but this problem is also considered a mild form of linkrot.

High-level architecture of a standard Web crawler

Typically when a user enters a query into a search engine it is a few keywords.[15] The index already has the names of the sites containing the keywords, and these are instantly obtained from the index. The real

processing load is in generating the web pages that are the search results list: Every page in the entire list must be weighted according to information in the indexes.[14] Then the top search result item requires the lookup, reconstruction, and markup of the snippets showing the context of the keywords matched. These are only part of the processing each search results web page requires, and further pages (next to the top) require more of this post processing.

Beyond simple keyword lookups, search engines offer their own GUI- or command-driven operators and search parameters to refine the search results. These provide the necessary controls for the user engaged in the feedback loop users create byfiltering and weighting while refining the search results, given the initial pages of the first search results. For example, from 2007 the Google.com search engine has allowed one to filter by date by clicking "Show search tools" in the leftmost column of the initial search results page, and then selecting the desired date range.[16] It's also possible to weight by date because each page has a modification time. Most search engines support the use of the boolean operators AND, OR and NOT to help end users refine the search query. Boolean operators are for literal searches that allow the user to refine and extend the terms of the search. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search, which allows users to define the distance between keywords.[14] There is alsoconcept-based searching where the research involves using statistical analysis on pages containing the words or phrases you search for. As well, natural language queries allow the user to type a question in the same form one would ask it to a human.[17] A site like this would be ask.com.[18]

The usefulness of a search engine depends on the relevance of the result set it gives back. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another.[14] The methods also change over time as Internet usage changes and new techniques evolve. There are two main types of search engine that have

evolved: one is a system of predefined and hierarchically ordered keywords that humans have programmed extensively. The other is a system that generates an "inverted index" by analyzing texts it locates. This first form relies much more heavily on the computer itself to do the bulk of the work.

Most Web search engines are commercial ventures supported by advertising revenue and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. Search engines that do not accept money for their search results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads.

IV. MARKET SHARE

Google is the world's most popular search engine, with a market share of 67.49 percent as of September, 2015. Bing comes in at second place.[20]

The world's most popular search engines are:

Search engine	Market share in September 2015	
Google	69.24%	
Bing	12.26%	
Yahoo!	9.19%	
Baidu	6.48%	

Search engine	Market share in September 2015	
AOL	1.11%	
Ask	0.24%	
Lycos	0.00%	

10) Jump up^ Allen, Kevin (25 September 2015). "A look at Google search numbers".Kansas City Internet Marketing. Retrieved 26 October 2015. External link in|publisher= (help)

11) Jump up^ Gandal, Neil (2001). "The dynamics of competition in the internet search engine market". International Journal of Industrial

REFERENCES

1) Jump up^ "RFC 812 - NICNAME/WHOIS". ietf.org.

2) Jump up^ http://ftp.sunet.se/pub/Internet-documents/matrix/services/KIS-id.txt

3) Jump up^ "World-Wide Web Servers". W3.org. Retrieved 2012-05-14.

4) Jump up^ "What's New! February 1994". Home.mcom.com. Retrieved 2012-05-14.

5) Jump up^ "Internet History - Search Engines" (from Search Engine Watch), Universiteit Leiden, Netherlands, September 2001, web: LeidenU-Archie.

6) Jump up^ Oscar Nierstrasz (2 September 1993). "Searchable Catalog of WWW Resources (experimental)".

7) Jump up^ "Archive of NCSA what's new in December 1993 page". Web.archive.org. 2001-06-20. Archived from the original on 2001-06-20. Retrieved 2012-05-14.

8) Jump up^ "Yahoo! And Netscape Ink International Distribution Deal" (PDF)

9) Jump up^ "Browser Deals Push Netscape Stock Up 7.8%". Los Angeles Times. 1 April 1996