History Of Search Engines

Tom Seymour, Minot State University, USA
Dean Frantsvog, Minot State University, USA
Satheesh Kumar, Minot State University, USA

ABSTRACT

As the number of sites on the Web increased in the mid-to-late 90s, search engines started appearing to help people find information quickly. Search engines developed business models to finance their services, such as pay per click programs offered by Open Text in 1996 and then Goto.com in 1998. Goto.com later changed its name to Overture in 2001, and was purchased by Yahoo! in 2003, and now offers paid search opportunities for advertisers through Yahoo! Search Marketing. Google also began to offer advertisements on search results pages in 2000 through the Google Ad Words program. By 2007, pay-per-click programs proved to be primary money-makers for search engines. In a market dominated by Google, in 2009 Yahoo! and Microsoft announced the intention to forge an alliance. The Yahoo! & Microsoft Search Alliance eventually received approval from regulators in the US and Europe in February 2010. Search engine optimization consultants expanded their offerings to help businesses learn about and use the advertising opportunities offered by search engines, and new agencies focusing primarily upon marketing and advertising through search engines emerged. The term "Search Engine Marketing" was proposed by Danny Sullivan in 2001 to cover the spectrum of activities involved in performing SEO, managing paid listings at the search engines, submitting sites to directories, and developing online marketing strategies for businesses, organizations, and individuals. Some of the latest theoretical advances include Search Engine Marketing Management (SEMM). SEMM relates to activities including SEO but focuses on return on investment (ROI) management instead of relevant traffic building (as is the case of mainstream SEO). SEMM also integrates organic SEO, trying to achieve top ranking without using paid means of achieving top in search engines, and PayPerClick SEO. For example some of the attention is placed on the web page layout design and how content and information is displayed to the website visitor.

Keywords: search engines; Internet; meta-search engine; Cern

INTRODUCTION

*W*eb Search Engine is a software program that searches the Internet (bunch of websites) based on the words that you designate as search terms (query words). Search engines look through their own databases of information in order to find what it is that you are looking for. Web Search Engines are a good example for massively sized Information Retrieval Systems.

HISTORY

During the early development of the web, there was a list of web servers edited by Tim Berners-Lee and hosted on the CERN web server. As more web servers went online the central list could not keep up. On the NCSA site new servers were announced under the title "What's New!".

The very first tool used for searching on the Internet was Archie.[1]The name stands for "archive" without the "v". It was created in 1990 by Alan Emtage, Bill Heelan and J. Peter Deutsch, computer science students at McGill University in Montreal. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites, creating a searchable database of file names; however, Archie did not index the contents of these sites since the amount of data was so limited it could be readily searched manually.

The rise of Gopher (created in 1991 by Mark McCahill at the University of Minnesota) led to two new search programs, Veronica and Jughead. Like Archie, they searched the file names and titles stored in Gopher index systems. Veronica (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives) provided a keyword search of most Gopher menu titles in the entire Gopher listings. Jughead (Jonzy's Universal Gopher Hierarchy Excavation and Display) was a tool for obtaining menu information from specific Gopher servers. While the name of the search engine "Archie" was not a reference to the Archie comic book series, "Veronica" and "Jughead" are characters in the series, thus referencing their predecessor.

In the summer of 1993, no search engine existed yet for the web, though numerous specialized catalogues were maintained by hand. Oscar Nierstrasz at the University of Geneva wrote a series of Perl scripts that would periodically mirror these pages and rewrite them into a standard format which formed the basis for W3Catalog, the web's first primitive search engine, released on September 2, 1993.

In June 1993, Matthew Gray, then at MIT, produced what was probably the first web robot, the Perl-based World Wide Web Wanderer, and used it to generate an index called 'Wandex'. The purpose of the Wanderer was to measure the size of the World Wide Web, which it did until late 1995. The web's second search engine Aliweb appeared in November 1993. Aliweb did not use a web robot, but instead depended on being notified by website administrators of the existence at each site of an index file in a particular format.

Jump Station (released in December 1993) used a web robot to find web pages and to build its index, and used a web form as the interface to its query program. It was thus the first WWW resource-discovery tool to combine the three essential features of a web search engine (crawling, indexing, and searching) as described below. Because of the limited resources available on the platform on which it ran, its indexing and hence searching were limited to the titles and headings found in the web pages the crawler encountered.

One of the first "full text" crawler-based search engines was WebCrawler, which came out in 1994. Unlike its predecessors, it let users search for any word in any webpage, which has become the standard for all major search engines since. It was also the first one to be widely known by the public. Also in 1994, Lycos (which started at Carnegie Mellon University) was launched and became a major commercial endeavor.

Soon after, many search engines appeared and vied for popularity. These included Magellan (search engine), Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Yahoo! was among the most popular ways for people to find web pages of interest, but its search function operated on its web directory, rather than full-text copies of web pages. Information seekers could also browse the directory instead of doing a keyword-based search.

In 1996, Netscape was looking to give a single search engine an exclusive deal to be the featured search engine on Netscape's web browser. There was so much interest that instead a deal was struck with Netscape by five of the major search engines, where for $5Million per year each search engine would be in a rotation on the Netscape search engine page. The five engines were Yahoo!, Magellan, Lycos, Infoseek, and Excite.

Search engines were also known as some of the brightest stars in the Internet investing frenzy that occurred in the late 1990s. Several companies entered the market spectacularly, receiving record gains during their initial public offerings. Some have taken down their public search engine, and are marketing enterprise-only editions, such as Northern Light. Many search engine companies were caught up in the dot-com bubble, a speculation-driven market boom that peaked in 1999 and ended in 2001.

Around 2000, Google's search engine rose to prominence. The company achieved better results for many searches with an innovation called PageRank. This iterative algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that good or desirable pages are linked to more than others. Google also maintained a minimalist interface to its search engine. In contrast, many of its competitors embedded a search engine in a web portal.

By 2000, Yahoo was providing search services based on Inktomi's search engine. Yahoo! acquired Inktomi in 2002 and Overture (which owned AlltheWeb and AltaVista) in 2003. Yahoo! switched to Google's search engine until 2004, when it launched its own search engine based on the combined technologies of its acquisitions.
Microsoft first launched MSN Search in the fall of 1998 using search results from Inktomi. In early 1999 the site began to display listings from Looksmart blended with results from Inktomi except for a short time in 1999 when results from AltaVista were used instead. In 2004, Microsoft began a transition to its own search technology, powered by its own web crawler (called msnbot). Microsoft's rebranded search engine, Bing, was launched on June 1, 2009. On July 29, 2009, Yahoo! and Microsoft finalized a deal in which Yahoo! Search would be powered by Microsoft Bing technology.

TYPES OF SEARCH ENGINES

Archie - (1990)

History of Search Engine can be said as started in A.D. 1990. The very first tool used for searching on the Internet was Archie. It was created in 1990 by Alan Emtage, a student at McGill University in Montreal. The Archie Database was made up of the file directories from hundreds of systems. When you searched this Archie Database on the basis of a file's name, Archie could tell you which directory paths on which systems hold a copy of the file you want. Archie did not index the contents of these sites. This Archie Software, periodically reached out to all known openly available ftp sites, list their files, and build a searchable index. The commands to search Archie were UNIX commands, and it took some knowledge of UNIX to use it to its full capability.

Gopher - (1991)

Later in A.D. 1991 Gopher came into the scene. Gopher was a menu system that simplified locating and using Internet resources. Gopher was designed for distributing, searching, and retrieving documents over the Internet. Gopher offered some features not natively supported by the Web and imposes a much stronger hierarchy on information stored on it. Gopher Software made it possible for the system administrator at any Internet site to prepare a customized menu of files, features and Internet resources. When you used the Gopher, all you had to do is select the item you want from the menu. Gopher was a protocol system, which in advance of the World Wide Web, allowed server based text files to be hierarchically organized and easily viewed by end users who accessed the server using Gopher Applications on remote computers. Initially Gopher Browsers could only display text-based files before developments such as Hyper Gopher, which were able to handle simple graphic formats.

Veronica and Jughead - (1991)

Archie, Gopher, Veronica and Jughead were three standard "finding" tools on the Internet. The rise of Gopher led to two new search programs, Veronica and Jughead. Like Archie, they searched the file names and titles stored in Gopher index systems. Veronica was a Resource-Discovery system providing access to information resources held on most (99% +) of the world's Gopher Servers. The Veronica Database was a collection of menus from most Gopher sites. When you did a Veronica Search, you were searching the menu items. Veronica used to build an on-the-spot menu consisting of just those items that matched your request. When the search was finished, Veronica would present you with a customized Gopher menu. Veronica would not only present you with a list of Gopher menu items, it would also act like a Gopher. Jughead on the other hand was distinct from Veronica. Jughead searched a single server at a time. Jughead indexed the servers quickly so it used to builds its database in memory. When Jughead used all of the available memory, it used to become unacceptably slow, limiting the size the servers it can index. Veronica does not have this problem.

W3Catalog & Wanderer - (1993)

Initially, the only widely available browsers were purely textual. Mosaic was the first browser to display images in line with text instead of displaying images in a separate window. While often described as the first graphical web browser. W3Catalog was one of the first search engines that attempted to provide a general searchable catalog for WWW resources.

Unlike later search engines, like Aliweb, which attempt to index the web by crawling over the accessible content of web sites, W3Catalog exploited the fact that many high-qualities, manually maintained lists of web resources were already available. W3 Catalog simply mirrored these pages, reformatted the contents into individual entries, and provided a Perl-based front-end to enable dynamic querying.

In 1993, Matthew Gray, then at MIT, produced what was probably the first web robot, the Perl-based World Wide Web Wanderer, and used it to generate an index called "Wandex". The World Wide Web Wanderer, also referred to as just the Wanderer, was a Perl-based web crawler that was first deployed in June 1993 to measure the size of the World Wide Web. Wanderer was developed at the Massachusetts Institute of Technology by Matthew Gray, who now works for Google. It was used to generate an index called the Wandex later in 1993. While the Wanderer was probably the first web robot, and, with its index, clearly had the potential to become a general-purpose WWW search engine.

Aliweb - (1993)

Second search engine, Aliweb appeared in November 1993. Aliweb allowed users to submit the locations of index files on their sites which enabled the search engine to include WebPages and add user-written page descriptions and keywords. Aliweb, the search engine, distinguished from its contemporaries such as AltaVista by the fact that it does not automatically index sites. If a Webmaster wanted a site to be indexed by Aliweb then he or she would have to write a special file and register it with Aliweb Server. Because of the difficulty of doing this, ALIWEB has a much smaller database than search engines such as Lycos and has suffered in popularity. Aliweb provided a tool allowing users to just keep track of the services they provide, in such a way that automatic programs could simply pick up their descriptions, and combine them into a searchable database.

Jump Station - (1993)

Jump Station (released in December 1993) used a web robot to find web pages and to build its index, and used a web form as the interface to its query program. Jump Station was thus the first WWW resource-discovery tool to combine the three essential features of a web search engine (crawling, indexing, and searching). Because of the limited resources available on the platform on which Jump Station ran, its indexing and hence searching were limited to the titles and headings found in the web pages the crawler encountered.
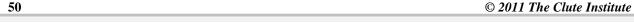
Jump Station used document titles and headings to index the web pages found using a simple linear search, and did not provide any ranking of results. Jump Station had the same basic shape as Google search. Brian Pinkerton, a CSE student at the University of Washington, starts WebCrawler in his spare time. At first, WebCrawler was a desktop application, not a Web service as it is today.

WebCrawler - (1994)

Brian Pinkerton, a CSE student at the University of Washington, starts WebCrawler in his spare time. At first, WebCrawler was a desktop application, not a Web service as it is today. WebCrawler went live on the Web with a database containing pages from just over 4000 different Web sites. WebCrawler was the first Web search engine to provide full text search. It went live on April 20, 1994 and was created by Brian Pinkerton at the University of Washington. It was bought by America Online on June 1, 1995 and sold to Excite on April 1, 1997. Pinkerton built a web interface to his WebCrawler program, which was released on April 20, 1994, with a database containing documents from over 6,000 web servers. The WebCrawler was unique in that it was the first web robot that was capable of indexing every word on a web page, while other bots were storing a URL, a title and at most 100 words.

MetaCrawler - (1995)

The concept of Meta-Search Engine came into existence in which a single interface provided search result that was generated by multiple search engines rather than a single Search Engine Algorithm. Daniel Dreilinger at

Colorado State University developed Search Savvy which let users searched up to 20 different search engines at one and a number of directories.

MetaCrawler improved on accuracy of Search Savvy with the addition of its own search syntax and behind the scenes, matching its syntax to that of the search engines it was probing. MetaCrawler searched through six search engines, yet while providing better results, still could not match those achieved by searching each engine individually.

AltaVista - (1995)

AltaVista was once one of the most popular search engines but its popularity waned with the rise of Google. The two key participants who created the engine were Louis Monier, who wrote the crawler, and Michael Burrows, who wrote the indexer. AltaVista was backed by the most powerful computing server available. AltaVista was the fastest search engine and could handle millions of hits a day without any degradation.

One key change that came with AltaVista was the inclusion of a natural language search. Users could type in a phrase or a question and get an intelligent response. For instance, "Where is London?" without getting a million-plus pages referring to "where" and "is."

Excite - (1995)

Yahoo! was among the most popular ways for people to find web pages of interest, but its search function operated on its web directory, rather than full-text copies of web pages. Information seekers could also browse the directory instead of doing a keyword-based search. In 1995, Open Text provided the search technology used by Yahoo! as part of its Web index.

In 1996, Netscape was looking to give a single search engine an exclusive deal to be their featured search engine. Resultantly five major Search Engines were Yahoo!, Magellan, Lycos, Infoseek, and Excite joined the deal. Excite is an Internet portal, and as one of the major "dotcom" "portals" of the 1990s (along with Yahoo!, Lycos and Netscape), it was once one of the most recognized brands on the Internet.

Excite first appeared at the end of 1995 and was one of a spate of launches by the new 'crawler' based search engines, sending out spiders to record websites and build a searchable index – others from this time were AltaVista, Lycos, WebCrawler and Infoseek. SAPO was created on September 4, 1995 at the University of Aveiro by seven members of the Computer Science Center of the University.

Dogpile, Inktomi, & HotBot - (1996)

Dogpile began operation in November 1996. The site was developed by Aaron Flin. Dogpile was a metasearch site. It searched multiple engines, filtered for duplicates and then presented the results to the user. Inktomi software was incorporated in the widely-used HotBot search engine, which displaced AltaVista as the leading web-crawler-based search engine, and which was in turn displaced by Google. The company Inktomi Corporation was initially founded based on the real-world success of the search engine they developed at the university. After the bursting of the dot-com bubble, Inktomi was acquired by Yahoo!

HotBot is one of the early Internet search engines and was launched in May 1996. It updated its search database more often than its competitors. HotBot was one of the first search engines to offer the ability to search within search results. HotBot also offered free webpage hosting, but only for a short time, and it was taken down without any notice to its users. HotBot proved itself to be one of the most powerful search engines of its day, with a spider capable of indexing 10 million pages a day. This meant HotBot not only had the most up to date list of available new sites and pages, but was capable of re-indexing all previously indexed pages to ensure they were all up to date as well.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

fastcase
Smarter legal research.