1

A Concise Search Engine History

The Early Days

The modern Internet owes its existence to the small Soviet satellite, called Sputnik, that was lunched into orbit in 1958. As a response to this Soviet technological shock, the Advanced Research Projects Agency, ARPA, was created under directive of President Eisenhower. Over the years the agency had changed its name a few times. Today it stands as the Defense Advanced Research Projects Agency, DARPA.

One of the key achievements by ARPA was the creation of the so called ARPA network, also known as ARPAnet. The very first message over ARPAnet was sent on October 29. 1969. A UCLA student, Charley Kline, with the guidance of UCLA Professor, Leonard Kleinrock, sent the first message from a UCLA terminal to another terminal at Stanford Research Institute. The message just said "lo". The intent was to send "login" but the system crashed. Eventually the full message made it through!

With the introduction of ARPAnet email application in 1972, and the creation of File Transfer Protocol, ARPAnet had its very first, albeit basic, email system.

ARPAnet network was not the only 'kid on the block'. There were other public networks that sprung up over the years including X.25, UUCP and FidoNet. In the beginning they were all disjointed. With the introduction of packet switching protocols, TCP/IP, different networks could be joined together seamlessly, collectively forming a network of networks or what we know today as Internet. In 1983, the Domain Name System, DNS, is designed with the first top level domains including ".com", ".gov", ".mil", ."org", ".net", and ".int".

And it would be pretty hard to imagine the internet existing as it does today without the power of search engine helping to bring it all together. The humble roots of today's powerful search engines can be traced back to the 1960's and a Harvard professor, Dr. Gerard Salton, who created Salton's Magical Automatic Retriever of Text or SMART. SMART is considered to be the world's first digital search engine.

In 1990, while working at CERN, Tim Berners-Lee and Robert Cailliau created the WorldWideWeb project. A year later, Hypertext Markup Language HTML was born. The era of mainstream Internet was about to begin. No one could have anticipated that this emerging technology would ever move past its intended scientific and military purposes. There was only one piece missing; the glue; the Search Engines.

From Archie to Cuil

In this section we cover some of the most important Search Engines, Web Directories and Web Spiders. Some of these no longer exist, but deserve a mention for their pioneering legacies.

Archie, 1990

Archie was created by Peter Deutsch and others at the McGill University in Montreal, Canada. It was the very first tool used to search the Internet.
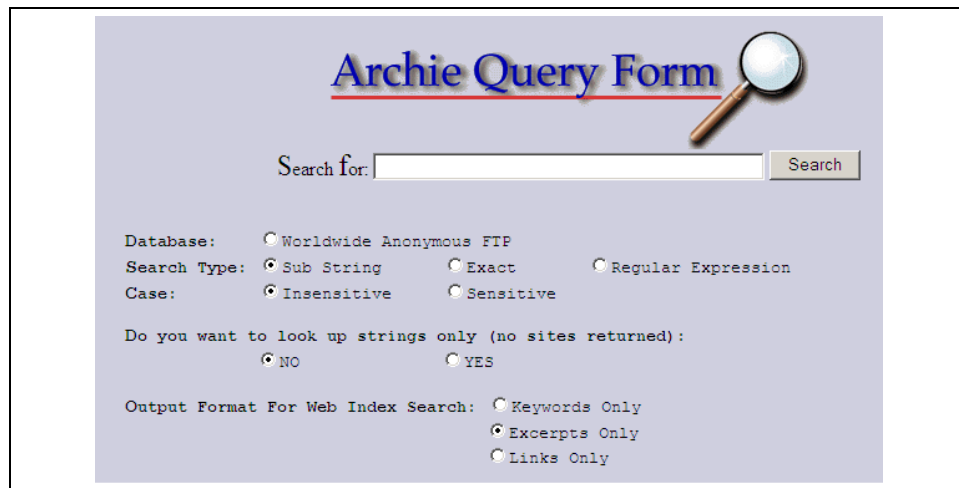

Figure 1-1. Archie Search interface

Serving as an 'archiver' for ftp sites, Archie would frequently go out to all known, openly available, ftp sites, list their files, and build a searchable index. This index would contain only the names of the files and would be stored locally as text files.

Initially and in order to fully search Archie, you had to be proficient in UNIX and use the *grep* command for your searches. Several front-ends were developed to simplify its usage. Archie quickly spread around Universities in the early nineties.

Veronica & Jughead, 1991

University of Minnesota developed a menu system called Gopher. Gopher sites indexed text files. Gopher systems were accessible via the Gopher, Telnet or FTP clients.

In any particular Gopher system, instead of UNIX commands, you would simply type or click on a number to select the menu option you wanted to explore. Shortly thereafter the University of Nevada at Reno developed the VERONICA searchable index of gopher menus. It stood for "Very Easy Rodent-Oriented Netwide Index to Computerized Archives". Veronica's indexed gopher menus around the world.

Jughead was very similar to Veronica. It was developed at the University of Utah Computer Centre. Its author, Rhett "Jonzy" Jones insisted that Jughead stood for "Jonzy's Universal Gopher Hierarch Excavation and Display" and had no relation to the Archie comic book character. Jughead introduced Boolean search capabilities but was limited to search individual servers.

There are very few gopher servers left around the world. At the time of this writing, the best one available was gopher://gopher.floodgap.com. If you are interesting in viewing gopher sites you can use Firefox web browser. Please note that Internet Explorer has discontinued gopher support from 2002 and that Firefox will do the same in their next version, Firefox 4.

World Wide Web Wanderer, 1993

World's first web spider, The World Wide Web Wanderer was a Perl script design by Matthew Grey. Here is what Matthew said about Wanderer in his report at http://www.mit.edu/people/mkgray/growth/ link:

> "In Spring of 1993, I wrote the Wanderer to systematically traverse the Web and collect sites. I was initially motivated primarily to discover new sites, as the Web was still a relatively small place. As the Web grew rapidly, the focus quickly changed to charting the growth of the Web." --

Wanderer did not capture any content related information. It only collected the crawled URLs. The index of these URLs was called Wandex.

Excite, 1993

Created by a group of six Stanford students, Excite was one the very first "web spider" based search engines. Excite was also one of the first search engines to use more advanced ranking mechanisms in order to increase the search results quality.

The homepage of Excite in many ways looked like a portal. Its popularity grew over the years and in 1996 Excite acquired WebCrawler. To see how Excite looked in its heydays visit the Internet Archive at:

- http://web.archive.org/web/*/www.excite.com

Over the years, Excite provided search services for Microsoft and Netscape before getting acquired by @Home in 1998. In March 2004, Ask Jeeves acquires the www.excite.com portal from the debris of @Home bankruptcy.

WebCrawler, 1994

WebCrawler was the first search engine with the ability to index entire HTML documents -- not just a few hundred words. It was created by Brian Pinkerton at the University of Washington. Its popularity exploded in relatively short period of time before getting acquired by several companies including AOL, Excite, and finally Infospace in 2001.

The WebCrawler design was similar to other portal-looking websites. To see how WebCrawler looked in its heydays visit the Internet Archive at:

- http://web.archive.org/web/*/www.webcrawler.com

In its current form it resembles other Meta Search Engines such as www.dogpile.com. Meta search engines collect search results from several search engines including Google, Microsoft Live and Yahoo in order to provide aggregate view of top search results across all the popular search engines. The basic idea was to allow web users to come to one place to see results from all of the key search engines.

Lycos, 1994

Lycos was born at Pittsburgh's Carnegie Mellon University. Fathered by Dr. Michael Mauldin and his team, Lycos was released to public with an index of over fifty thousands documents on the Web.

With Lycos, end-users were able to get better, more focused search results. It was one of the first to add metrics such as "ranking by relevance", "prefix matching" and 'word proximity".

By end of 1996, Lycos had over sixty million web pages indexed, a feat no other search engine could claim at the time.

You can traverse through Lycos history using the Internet Archive at the following link:

- http://web.archive.org/web/*/http://www.lycos.com

Lycos enjoyed a good growth spurt in late nineties before being sold to a Spanish company, Terra Networks. Currently Lycos is owned by a Korean company called Daum Communications Corporation.

Infoseek, 1994

Infoseek introduced us to the world of advanced search modifiers including Advanced Boolean Modifiers, Text Quotes, as well as parentheses.

At first, Infoseek's creator Steven Kirsch offered it as a pay service only to reintroduce it later as a free search engine. What brought Infoseek into the spot light was a deal brokered with Netscape to have Infoseek as their default search engine.

In 1999, Infoseek gets acquired by Disney. Ultimately some Infoseek technology gets sold to Inktomi while Disney's go.com starts using Yahoo search.

Visit Internet Archive to see Infoseek in its early days:

- http://web.archive.org/web/*/http://www.infoseek.com
- http://web.archive.org/web/*/http://www.go.com

AltaVista, 1994

AltaVista was created by Luis Monier and Michael Burrows. Its multi-threaded crawler called Scooter was the most advanced web spider at that time.

Initially, AltaVista was made public on http://altavista.digital.com link. This was due to the fact that the AltaVista creators worked for Digital Equipment Corporation, DEC. In 1998, DEC is sold to Compaq who purchases www.altavista.com domain, the new internet address of AltaVista search engine.

Over time, AltaVista provided search services for Yahoo!, its eventual owner. Shortly after AltaVista was bought by Overture, Yahoo! purchased Overture. To see AltaVista beginnings go to these links:

- http://web.archive.org/web/*/http://altavista.digital.com
- http://web.archive.org/web/*/http://www.altavista.com

Ask Jeeves, 1997

Ask Jeeves was designed by Gary Chevsky. The concept of the Ask Jeeves search engine was different from any other as it tried to use a "natural language" approach. Web surfers would make searches using natural English by asking specific English language questions. This concept was not fully embraced by the mainstream audience. Towards the end of 2005, Ask Jeeves dropped "Jeeves" and it became ask.com.

To see AskJeeves in its early days visit:

- http://web.archive.org/web/*/http://www.askjeeves.com

Google, 1998

And so we come to Google. Another made-up word (a mistake, actually) known by all – well, most. Google is the current leader; the most popular; the most used; of all the Search Engines with millions upon millions of hits daily – and some cool extras.

Google was created by Larry Page and Sergey Brin at Stanford University. In many ways it was similar to most of the other major search engines, but its interface did not look like much, consisting of a simple form and search button. For many, this simplicity, so different from the cluttered portal-like search engine sites, was a welcome relief.
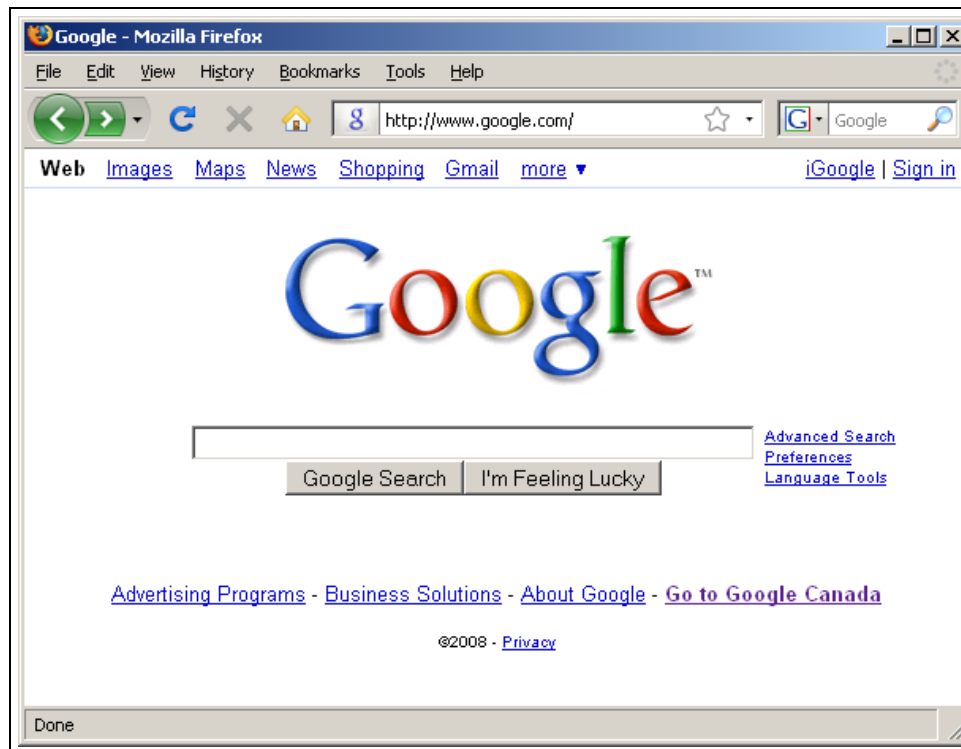

Figure 1-2. Google: The King of Search Engines

Over the years Google evolved into the multi-product company it is today, providing many other applications along with the search engine, including Google Earth, Google Moon, Froogle, AdWords and a very comprehensive web analytics tool for webmasters. In 2008 the company even launched its own standards-compliant browser, Chrome, into what some feel is an already saturated market.

Amazingly enough Google's home page has changed very little over time:

- http://web.archive.org/web/*/http://www.google.com

DMOZ, 1998

Here is a quote from http://www.dmoz.org/about.html link.

> "DMOZ was initially created by Rich Skrenta. DMOZ or "The Open Directory Project is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors.
>
> The web continues to grow at staggering rates. Automated search engines are increasingly unable to turn up useful results to search queries. The small paid editorial staffs at commercial directory sites can't keep up with submissions, and the quality and comprehensiveness of their directories has suffered. Link rot is setting in and they can't keep pace with the growth of the Internet.
>
> Instead of fighting the explosive growth of the Internet, the Open Directory provides the means for the Internet to organize itself. As the Internet grows, so do the number of net-citizens. These citizens can each organize a small portion of the web and present it back to the rest of the population, culling out the bad and useless and keeping only the best content."

As good as the DMOZ idea sounds, its popularity with mainstream Internet users is questionable. Part of this problem, may be, in the way the information is presented. There are literally endless "category and subcategory" levels to go through until you get to the information you need. If you have ever used DMOZ before you already know what I am talking about. In terms of SEO value of being listed on DMOZ it is also debatable since your listing would probably end up deeply buried (e.g. 7 or more levels deep) under a specific sub-sub-sub-sub-category.

Frankly, the only reason most people list their website URLs in DMOZ, is because Google specifically mentions DMOZ as one of the places where you need to list your sites.

DMOZ has also changed very little over time as evident in the Internet Archive link:
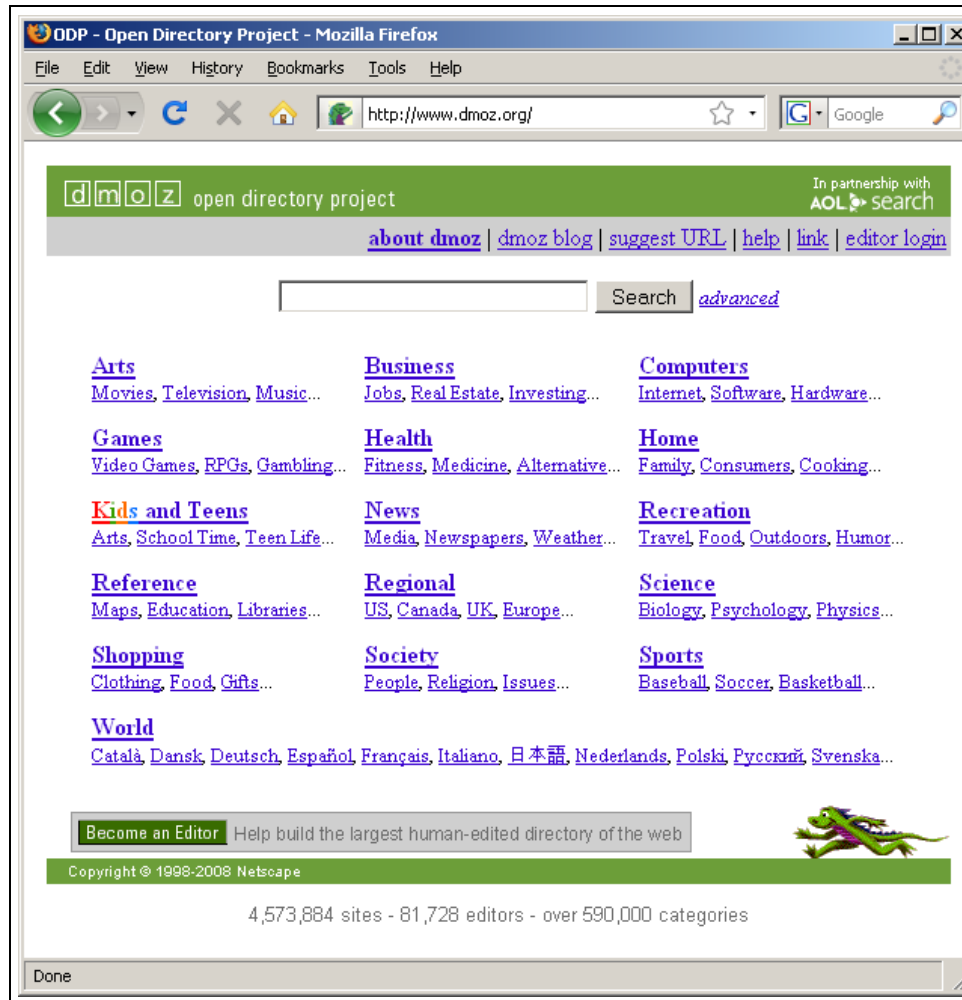
- http://web.archive.org/web/*/www.dmoz.org/

Figure 1-3. dmoz: open directory project

Yahoo!, 1994/2004

Founded by Jerry Yang and David Filo, Yahoo! emerged in the clumsy form of http://akebono.stanford.edu/yahoo link. It was made up of directory of websites. For the longest time, Yahoo! leveraged the search services of other companies. This all changed in 2004, when it dropped Google as its search provider in favor of its own acquired technology.

Yahoo has been Google's biggest rival for many years. What sets Yahoo apart from Google is its interface. When you go to www.yahoo.com website you actually see a web portal with all kinds of information.

Once you get to Yahoo's search result pages, things start to look similar to Google. Google does have an answer with its own portal igoogle.com where you can customize

news, weather and many other streams of information which can ultimately be streamlined / customized to specific keywords of interest in convenient square blocks.

In 2008, Microsoft makes a failed attempt to buy Yahoo! who responds by forming an advertising relationship with Google.

To see how Yahoo! has changed over the years use this link:

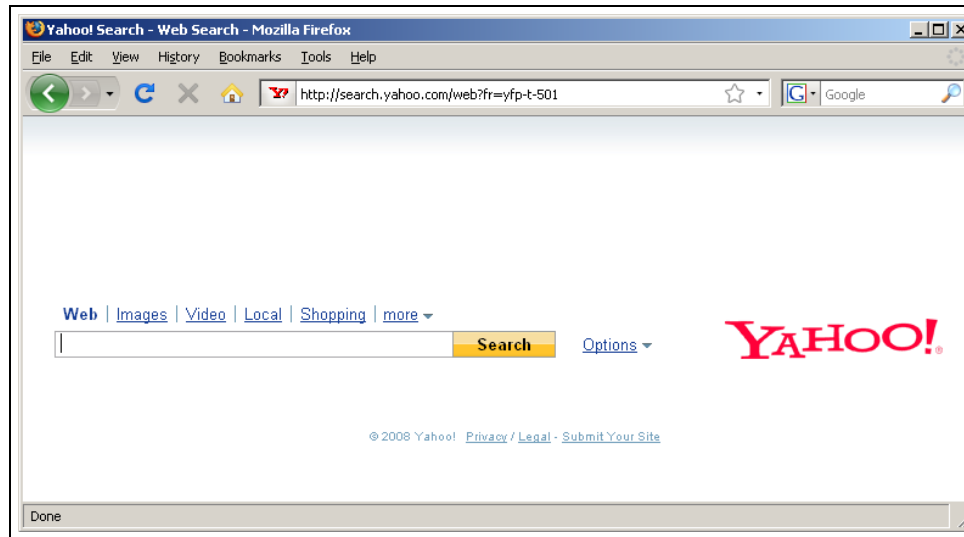- http://web.archive.org/web/*/http://yahoo.com

Figure 1-4. Yahoo!

Please note that the figure above is not showing the portal home page of Yahoo! but only its search interface.

In the later part of 2009, Yahoo! Signed an agreement to use Bing as its search results provider. It will take some time before Bing results are offered for Yahoo!'s users.

Live Search, 2007

Microsoft Live Search is the successor of the MSN search engine which failed to gain enough traction. Live Search is still struggling to cut into Google's market share.

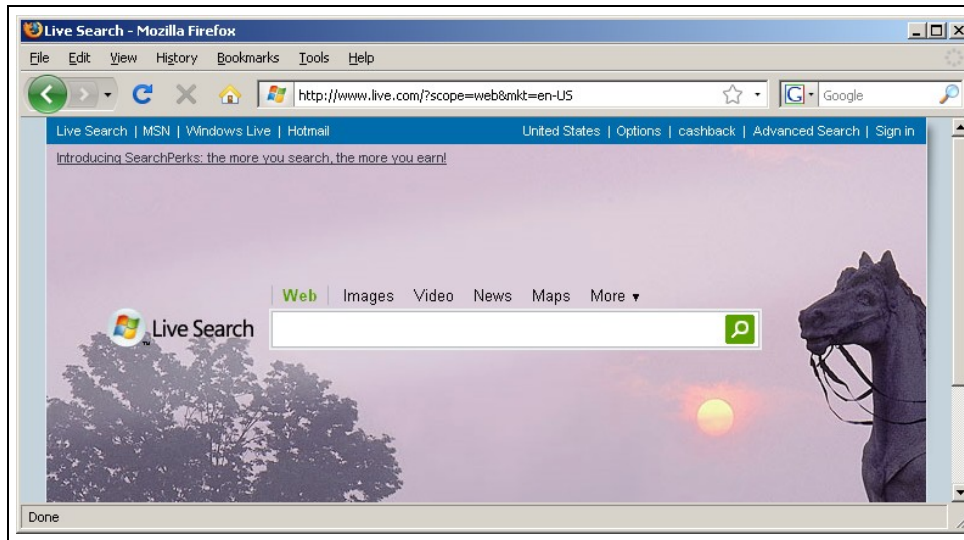- http://web.archive.org/web/*/http://www.live.com

Figure 1-5. Microsoft Live Search

It may be a long time before people switch to Live Search. The fact is that the word "Google" has become synonymous for many with the term "Search Engine". It has even become a slang-verb in pop culture with "I'll Google that later" a common term or phrase..

Having said that, never underestimate Microsoft. We all remember Netscape and the browser wars. Well, maybe not all of you reading this are old enough!

Cuil, 2008

Cuil, pronounced as "cool", was founded mostly by former Google employees. It is one of the latest search engines added to the mix. Its ultimate goal seems to be to overtake Google.

It is worthy to note some key differences of Cuil when compared to Google. Cuil claims to be very concerned about web user privacy. Here is what it says on Cuil's privacy page, http://www.cuil.com/info/privacy/:

> "Privacy is a hot topic these days, and we want you to feel totally comfortable using our service, so our privacy policy is very simple: when you search with Cuil, we do not collect any personally identifiable information, period. We have no idea who sends queries: not by name, not by IP address, and not by cookies (more on this later). Your search history is your business, not ours."

At the time of this writing, Cuil is claiming to have indexed more then 124 billion web pages, more then any other search engine.

Initially it would seem that Cuil was gaining a lot of traction among web users. Time will tell if Cuil can do what others have failed to do thus far -- in trying to overtake Google.

Bing, 2009

Bing, the Microsoft Live successor, was release with much fanfare. It was touted as a 'better way to search'.