

Search here

8 July 2011

This paper was written for school purpose.

University of Utrecht // Faculty of Humanities
Degree/program // pre-MA New Media & Digital Culture
Course // BA-thesis

Student

Salko Joost Kattenberg // 3614875

Search here // A historical analysis of search engines development

Supervisor

Mirko Tobias Schäfer

Table of Content

Paper

Introduction	4
Growth	5
Human vs Computer	6
Changing core-businesses	7
It's all about the money	7
Uying business and market shares	8
Geographical positioning	9
From open to closed systems	9
Conclusion	10

Appendix A – A History of Search Engines

Introduction – Why this Appendix?	12
1990-1993 The invention of internet search engines	13
Archie	13
Veronica and Jughead	14
W3 Catalog	14
World Wide Web Wanderer	14
Jumpstation and WWW Worm	15
Aliweb	15
1994-1996 Internet, from lots to a legion	17
Webcrawler	17
MetaCrawler	18
Lycos	18
Inktomi	19
InfoSeek	19
Excite	20
AltaVista	20
Yahoo!	21
Daum	22
Go.com	22
AOL/America Online	22
Dogpile	23
1997-1999 The garden of search and the beginning of the internet bubble	24
InfoSpace	24
Hotbot	24
Ask.com	25
Yandex	25
Google	26
MSN Search / Windows Live / Live Search / Bing	27

1999-2002 Surviving the bubble, and the start of the market share wars	28
AlltheWeb	28
Naver	28
Teoma	29
Baidu	29
2002-2010 The smoke clears, everything is quite now	30

Appendix B – Numbers and percentages

Introduction	31
1992-1995	31
1995-2004	32
2005-2011	33

Appendix C – Visuals and Graphs

Introduction	34
Historical internet chart: takeovers and launches	35

References	36-41
-------------------	-------

Introduction

During the 1980's the internet was introduced into the world. The internet as it was then, consisted of a web of linked computers, which enabled people to access information within the boundaries of the internet's infrastructure. During the early years of the internet a lot of university computers around the world started to join the internet, focusing on the innovation of the internet and thereby exploring the possibilities of a global information and communication structure (Castells, 2001). Over the course of the last twenty years the internet has grown, developed and innovated beyond anyone's wildest imagination. Billions of internet pages and applications have become available to us through the use of internet, making the internet the largest accumulation of data in the whole history of the human race. Coping with the immense number of pages and the incomprehensible amount of data are search engines. They are our guides to this vast and enormous cyberspace.

Over the last 20 years internet search engines have become inseparable with our use of internet. Search engines help us to cope with the immense size of the available information on the internet, and supply us with categorized, indexed and logical information/data. With the rise of major internet search companies, we have begun raising questions about the way these search engines operate, and how they affect our view of the internet. This is mainly because internet is considered a medium that stimulates democracy or freedom of speech, and thereby freedom of information. A *cyberspace*, which should be free of governance and economic values, if we believe J.P. Barlow's *Declaration of Independent Cyberspace*. It are these discussions about privacy, transparent data harvesting, advertisement income and search result manipulation that fill today's general debate about internet search engines. In this paper I try to address internet search engines from a different perspective, and will not raise questions regarding the general opinion or ideological connotations of internet search engines.

In this paper I will try to write a full historical review on internet search engines. By looking at past innovations in the last 20 years of search engine development, I mean to provide a clear view of the course of development over the last 20 years. In my research I stumbled upon 132 different search engines that have existed or still exist upon this day. From the 132 search engines, I have chosen 30 characteristic search engines to analyze and research deeper. These 30 search engines will be used to construct the main thesis in this paper. Since I mean to provide a clear and historical view of internet search engines, I included a detailed description and analysis of every search engine that is used in this paper. All these descriptions can be found in appendix A, which can also be used as reference during the reading of this paper.

Providing a clear view of the course of development of search engines, is meant to provide better insight regarding the way search engines operate. It is my opinion that the general ideas do not always align with the actual way internet search engines operate or have been operating. To address some of the innovations and changes internet search engines have made during the last 20 years, I have divided this paper into different sections. Every section will provide new insight into the historical development and innovations of internet search engines.

Growth

The size of the web has always been a question for many scientist and internet companies. While we have a pretty clear number of the total internet users in the world, around 2 billion as of March 2011, this only accounts for 30% of the world's population (Internet World Stats, 2011). How great these numbers might seem, they are nothing compared to the size of the internet itself. The growth of the internet has had a major influence on the way search engines developed and innovated over the years. While the first search engines were more focused on how to retrieve information, this later shifted to coping with the extreme growth of the internet. Today internet search engines still don't manage to index the whole internet. To get a better understanding of the development of the internet, I will provide some milestones of the size and indexing capabilities of search engines in the last 20 years. A full report of numbers is given in Appendix B.

In 1990, the same year that Tim Berners-Lee together with Roger Cailliau introduced the world wide web on the internet (Berners-Lee, 1992), the first search and archiving program was introduced, Archie made by Alan Emtage (Emtage, 1992). Up to that point the internet was not searchable. You needed to know internet addresses to gain access to other computers. There were no URL's, no databases of web addresses and no hyperlink referencing. Some sites that had listings of different web addresses, found themselves trying to cope with the enormous growth of servers and available content. The size of the entire internet was unknown.

Archie was meant to index ftp-servers (see Archie) and cope with a relative small internet. In 1992 the internet consisted of a mere 535.000 hosts (Gray, 1996) where Archie had indexed about 1 million online documents (Emtage, 1992). From 1992 to 1996 the Word Wide Web Wanderer, made by Matthew Gray, was designed to determine the expanse and growth of the internet at the time (Gray, 1996). Gray noticed that the internet had a doubling period of 3 to 6 months. Resulting in the fact that the internet consisted of a 9,5 million internet hosts in 1995, and had around 100 million pages (Gray, 1996). While first Archie, Veronica and Jughead were the ones that tried to cope with the enormous growth, they quickly figured out their programming scripts were not capable of coping with the immense size of the web (See Veronica and Jughead). It was around the mid-nineties that many of these early internet search engines became slow in producing search results. Jumpstation and WWWorm were search engines that tried to work with webcrawlers or spiders to cope with the vast growing internet (See Jumpstation and WWWorm). However, this began to raise questions about the use of internet spiders or webcrawlers, which would make the web even slower and less accessible. Concepts like the *Robots Exclusion Standard* that was presented by Martijn Koster in 1994, founder of the Aliweb search engine, were raised around that time. (Koster, 1994).

In the mid-nineties the development of search engines really started to change. There was a lot to gain by providing quick results to internet users. Lycos is one of the first search engines that were especially designed to cope with the immense size of the internet (see Lycos). Lycos was determined to keep up with the fast growth of the internet. This is clearly visible in the growth of their search index. By the launch of Lycos in July 1994, the search index counted 54.000 indexed documents (Mauldin, 1997). This was less than Archie, Veronica and Jughead at the time. But Lycos servers were a lot faster than all its predecessors, and had indexed 1,5 million documents in January 1995 (Mauldin, 1997). Finally surpassing Veronica, Jughead and Archie by the end of 1995, Lycos had indexed 60 million documents in November 1996 (Mauldin, 1997). In about two years' time Lycos managed to index about 75% of the web in 1996, which now consisted of about 80 million indexed pages (Mauldin, 1997). This was astonishing, Yahoo!, Excite and AOL were nothing compared to those numbers around that same time.

That was the benchmark, until AltaVista showed up at the end of 1995 (See AltaVista). Their 64-bit servers were capable of searching 2,5 million pages a day, silencing the competition (Lewis, 1995). In the years to come, not many search engines could match the speed of AltaVista. It was not until 2000 that Google and Inktomi (Powering Hotbot and Yahoo!) servers were also capable of coming up with greater results. In 2002, the newly founded AlltheWeb search engine, was able to surpass Google and indexed 2,1 billion documents to Google's 2,07 billion indexed documents (Bowmen, 2002). AlltheWeb was incorporated into Yahoo! and lost most of its search index (See AlltheWeb). Around 2005 the internet was estimated to

count about 11,5 billion internet pages, were Google had only indexed about 8,1 billion pages, a mere 70% of the internet (Kuner, 2007). Today in 2011 the indexed web is estimated to contain about 45 to 50 billion webpages, the size of the actual web is unknown (Kuner, 2011). Saying that Google still manages to index 70% of the internet (\pm 45 billion), would lead to conclude that the whole internet counts about 64,3 billion pages. However not able to index the entire internet, Google provides today's search results within tenths of a second.

It is astonishing to see how fast the internet has grown, and with that how fast internet search engines have become. In the last 20 years the internet has expanded from around 1,5 million pages in 1990, to probably around 65 billion pages in 2011. However it is important to see that from during 1994 until 2000 a lot of the development and innovations of the internet were conducted around the server capabilities and webcrawler-scripts, this in order to index the vast expanding internet and refresh the indexed pages more frequently. It is clearly visible that the expansion of the web leaves its marks on the development of internet search engines. With only 30% of the entire world population having internet access at the moment, the growth of the internet will continue to be a factor in the innovations and development in future years.

Human vs Computer.

While Archie, Veronica and Jughead used scripts to index ftp and gopher servers, the first www webcrawler or spider that was used was made by Matthew Gray for his World Wide Web Wanderer (Gray, 1996). The crawler was designed to index the size of the internet. The concept of using spiders or webcrawlers has come a long way since then. The first spiders were very simple, using hyperlinks to crawl further along the internet, indexing the headers of websites, like the spiders of Jumpstation and WWWWorm (See Jumpstation and WWWWorm). Later Webcrawler and other search engines would develop new spider programming scripts, and raise the quality of spider to a whole new level with full text indexing (Pinkerton, 1994). This allowed spiders to search through the entire content of webpages. Today spiders are capable of indexing hundreds of parameters and index relevant keywords out of endless pages of scripts and content and measure contexts and human interaction on the site.

However, the use of spiders as a way of indexing has always been challenged. One of the first to oppose the way spiders work is Oscar Nierstrasz, developer of the W3 Catalog in 1993 (Nierstrasz, 1996). He found out that a lot of people were constructing high-quality list online. These lists were added and checked by men not by computers or spiders. Filling his own internet index with a dozen of those lists meant he could deliver very good quality search results (Nierstrasz, 1996). While other search engines had major trouble coming up with the right search results, Nierstrasz saw great potential in using men made list.

The trend of not using spiders to rely on filling your search index is always been present. Metasearch engines use the top search results of other search engines to fill up their own search index (See MetaCrawler and Dogpile). However, there is still one big presence left today of the vision Nierstrasz had. Yahoo! founded by David Filo and Jerry Yang in 1995 (Yahoo!, 2005), does not rely on spiders or webcrawler technology, instead their system relies on pure human intelligence. They construct their web index by humanly categorizing web pages and thereby indexing and ranking their index. While this proves to be a time consuming task, Yahoo! is respected for their great search results. Today Yahoo! still is one of the biggest players in the global search engine market and shows that this technique is a very good alternative to spiders and computer scripts (comScore, 2011).

While the conflict between computer and human generated search indexing seems to be faded away, the concept was very lively in the development of search engines. In the mid 90's search engines had real problems with coming up with the right search results. One way to cope with this problem is using more human indexed search indexes and thereby improving the search results. Later more advanced methods like Google's Page Rank would lead spider and webcrawler technologies to such a high level, that those systems would come up with the right search results and rendering human made indexing less relevant. However, Page Rank is also just an algorithm that can be manipulated by web developers and so is vulnerable for making mistakes. Yahoo! does not seem to have the same failures.

Changing core-businesses.

In the early nineties search engines had a clear goal. Provide search capabilities to internet users. As Micheal L. Mauldin, the founder of Lycos have said: "Although the Web contains over million pages of information, those millions of pages are useless if you cannot find the pages you need" (Mauldin, 1997) The main goal of a lot of search engines was to provide new or innovating ways to provide the best search results to the large online audience. This shifted by the end of the nineties to a wider range of goals. Lycos together with Hotbot, AltaVista, Excite and Yahoo! were the first to expand their search engine capabilities and provide a wider range of services to their online costumers.

Yahoo! and Lycos started mail capabilities, offering free or relative cheap online email addresses and online inboxes. Hotbot and AltaVista started to expand their services by allowing users virtual online hosting. Especially Hotbot's online hosting servers were vastly used around the turn of the century. Lycos and Yahoo! continued to expand their services and came up with online news sites and online community pages and forums. This was done to provide users with a so called portal, a place where people could go to online and be provided with everything they would want online. Others later followed, OAL and Excite would also invest in instant messaging services and expand their online services though chat capabilities.

This meant that the main goals of search engines were slowly shifting towards other services and possibilities outside the normal range of search engine technology. This formed a base for many of the later established search engines, who were now also more focused on becoming an online portal than only becoming a search engine. This was clearly visible in the versions of Windows Live/Live Search, Yandex and Naver that were set up to provide a range of services next to search engine capabilities. Other search engines, whose main goal had been providing search, also started to incorporate more services. You can see that the core-businesses of these search engines quickly expanded and changed. AOL for instance almost completely discarded their search engine capabilities as core business and merged with Time Warner in 2001 (See AOL). AOL would become one of the biggest news and community sites, and still is a huge player in the search engine market today because of its dedicated member base. Microsoft search engines even tried to go a step further. Exploiting the fact that Microsoft also functions as an operating system, Microsoft tried to merge the offline and online environment and link them together in Windows search capabilities. This would erase the bridge between offline and online search capabilities on the computer (See MSN Search). Sadly, the idea didn't work out for Microsoft and Bing currently doesn't have these features anymore.

After the turn of the century Google would take over. Google's Image search together with Gmail and Google Maps opened up a whole new range of new possibilities. What made these programs unique was that they presented new information to the user. Image search was only just available on the market when Google launched their image search in 2001 (See Google). Google Maps was something completely different, offering a range of new information that was presented to online users. Google has been a leader in adding and programming new services and applications since 2001, coming up with new ways to form and index information that has not yet been indexed or even digitalized. This could be seen as one of the reasons why Google is currently global market leader in most of the countries of the world. Providing more information than just mere web pages, but supply a range of new and good information to online users, forming today's Google Network, which is more widespread than any other search engine.

It's all about the money

If we look at the development of search engines is hard not to look at the way search engines make their money. Advertising has been, and still is the number one income for all search engines. Since the early years of search engine history advertisement has been a part of search engines.

Advertisement has changed a lot since it first appeared in the online environment. Advertisement back in the early nineties was nothing else then a banner or tittle animated-gif on the homepage of internet search provider pages. This changed a lot when search engines started to realize they could deliver advertisement far more accurate. Search engines possessed the ability to know what people are looking for, if you know that, you can adjust your search engine advertisement to that search query that was entered. This made way for personal or search word related advertisement. Infoseek was one of the

first internet search engines that produced a program called Ultraseek (See InfoSeek). The program could react to the behavior of users and adjust the advertisements accordingly. This changed the way users were addressed, but also boosting the advertisement income. Now advertisers could be guaranteed that search engine users were given more accurate advertising resulting in better advertisement deals and more income. Other search engines, Like Yahoo! started using pay per click capabilities (See Yahoo!). This meant that a price per click was guaranteed to advertisement banners. This also worked well for Yahoo!, and many companies followed.

Today, most search engine companies can be described as advertisement companies, a point that Konrad Becker and Felix Stalder make in their book *Deep Search*. The advertisement income of Google accounts for 98% of its revenues (Becker et al. 2009). But how Google's advertisement algorithms work is something of a black box. We know that paid search results are shown independent of the regular search results. But how do searches on Google relate to my advertisements on Gmail? How does my location-software on Android relates to the advertisement I get when searching on Google.com? A research done by a competitive search engine Dogpile 2007 showed that: "On average 20-29 percent of first page search results are sponsored" (Dogpile, 2007). This is a lot, considering that most of us don't look much further than the first page results.

Questions about advertisement income in relation to the core businesses search engine should have -free and open search-start to show here. For instance, how are advertisement income/deals affecting my view of the internet? The point I am trying to make is that advertisement was there since the beginning of the first search engines. Advertisement is, on the moment, the only way for search engines to make their money. Therefore search engines will be and always has been entwined with the use, innovations and development of online advertisement.

Buying business and market shares

Looking at the historical course of internet search engines you can see that buying and selling companies is almost as usual as hiring staff. More important is why internet search engines have been buying and selling companies as they have. In my research I have come across two main reasons why internet search engines exchange ownership so much: to buy technology/knowledge or to buy market share.

Technology is mainly bought to expend the already existing search engines capabilities. We have already read that many search engines wanted to become internet portals. For becoming an online portal you need expertise in Mail servers, hosting providers, instant messaging, image search, mobile applications etc. Mostly companies with these expertises were bought en implemented into the search engine capabilities. Over the yeas this means that hundreds of small companies have been bought and incorporated within search engines. It happened so much that I can say that I have come across no search engines that still exists today that has never bought another technology or knowledge from other companies.

Next to that are many takeovers, which happened in the last 20 years. Many smaller and sometimes smarter internet search engines have been bought by larger and more competitive search engines. Yahoo! for instance has incorporated, InfoSeek, Go.com, Inktomi (partial) and AltaVista (See Yahoo!). All these search engines had great market shares and even some of them were at some point larger then Yahoo! itself. But controlling the market is far more important than the size of its users. Internet search engines pay handsome for gaining the users and search queries of other search engines. With these takeovers search engines would gain millions of new users and expend their advertisement contracts. Therefor earning the money back eventually, because the search engine grew in market share and size and thus establishing more grip on the search engine market. A perfect example of this is the Daum Corporation buying Lycos in 2004 in order to expand their search market across borders and thus expand in market share (See Daum).

Other takeover also happened because of technology. The WebCrawler, MetaCrawler, Hotbot, Teoma and AlltheWeb search engines were not only bought to expend market share and increase user amounts. These were also bought for their

technological programming scripts or fast server capabilities. These scripts were then incorporated into new search engines that provided better and faster result to online users.

Geographical positioning

The internet originates from the United States, originally started as a military operation (Castells, 2001). In the years before 1990 the internet was expanding mostly across America though many Universities. Later abroad Universities would also connect to the internet and start to expand the internet there. Looking at the way the internet expanded in those early years we can directly say that there is no such thing as a global information and communication network. While our general perception would think that internet is not bound by geographical borders, the internet is. Also taken in mind that still only 30% of the world's inhabitants are connected to the internet in 2011 (Internet World Stats, 2011).

But if we are willing to say that the access to internet is only a matter of time, our use of different internet search engines is also still bound by geographical borders and even linguistic boundaries. In my research I have included several 'non-Western' internet search engines: Daum (South-Korea), Yandex (Russia), Naver (South-Korea), Baidu (China). It is clear that these search engines are responsible for a whole different approach to search engine technology. Baidu, for instance is especially developed with the Chinese language in mind. Daum an early Korean search engine, founded in 1995, was struggling to find enough online Korean information and Korean internet pages. So, Daum started to provide online chat rooms, community pages and online question/answering forums in order to generate information and web pages. This worked so well that Daum is still one of the biggest search engines today in South-Korea, and manages to have a completely different approach to the way search engines developed. Yandex a Russian centered search engine is still the biggest internet search engine in Russia and Baidu is still the biggest search engine in China. So globally there are differences in our use of internet search engines, and therefore also difference in the way we use the internet. With Russian and Chinese markets starting to grow far more rapidly in the last few years, these countries are starting to provide these search engines with huge potential markets.

Knowing that Google has the biggest global market share and is the biggest search engine in most of the world's countries, you could say that there is still global way we search on the internet. But yet again you are wrong. Google, like many other search engines today, uses positioning software to manipulate advertisements and search results. While everything may look the same the search results are not. Google knows from where you conduct your search queries and responds accordingly. Try to enter the same search query in to different places some kilometers apart from each other. I'll bet you find very different results. While Google might give the impression that you are searching the entire web, Google is actually confining his search results down to geographical or even invisible borders. Off course, these programs are designed to come up with better search results, while this may be true, there is no such thing as a global search network or a global internet index. We are only presented with parts of the whole.

From open to closed systems.

During my research I have read many papers and articles about search engine technology. It has come to my attention that there is difference in the way search engines present themselves. Today I can still read how the search engines like Archie, Veronica, Aliweb, Inktomi, Metacrawler worked. How webcrawlers were programmed and how search indexes were filled, and what kind of information was stored. Hotbot and AltaVista were also very open in the way their search engine were operating and allowed users to adjust key features in their search engine. However, today's information is a bit more obscure. I can find information regarding the first years of Google and Yahoo!, but how they operate now a days appears to be in 'advanced' or 'intelligent' algorithms. It seems that together with the digital age has come a kind of closure. Have technical scripts become too hard to explain or has there been no innovation at all?

It is interesting to read the paper that was written in front of the launch of the Teoma search engine in 2000, *DiscoWeb: Applying Link Analysis to Web Search*. In that paper University professors address the concept of Google's Pagerank systems. The system seems to be a mere "straightforward technique of counting the number of incoming edges to a page" (Gerasoulin, 1999). They situate their concept to take off were Google left it. After the page ranking systems of Google, no

great innovations were accounted for. Even Teoma which was bought by Ask.com, is not considered to be 'better' than Google (See Teoma). The last years, without new technology or innovations, seem unreal looking at past developments. Although search engines merged, new search engines launched, advertisement algorithms advanced and new services and applications were added, no great innovations in search engine technology are accounted for. No, great new search engine has appeared on the market.

A better answer is given to us by Wade Roush in his paper *Search Beyond Google*. He interviews Craig Silverstein Google director of technology in 2004 and asks him how Google will keep up with the competition. Silverstein answers: "Google pays hundreds of researchers and software developers, including more than 60 PhDs, to man the front lines in this technology war (...). "We hope the next breakthrough comes from Google—but who knows?" (Roush, 2004) Wade goes on and addresses many different scientists that say that the reign of Google could be over in a matter of months. But now, 7 years later, Google is still ahead of the competition and we are nowhere near of how Google's search engine technology has grown. Maybe there is something to this 'technology war' Silverstein speaks of. In a war you don't reveal your secrets to the enemy. Reading Roush its paper it is clear that other scientist are working on better solutions, they are just not getting through to the mainstream media, or they are bought so fast by existing search engines that we never hear about them.

Conclusion

In this paper I have tried to look at past innovations in the last 20 years of search engine development. While this paper was presented in a couple of sections, all presenting different approaches to internet search development, you can easily see how these sections have influence on each other. More advertisement income will maybe result in buying new technologies or companies. While investing in new companies might lead to more users leading again to more advertisement income. You can see how advertisement, technology, users, services, market shares and search indexing algorithms all are related to each other and affect the success and growth of a search engine company. In this way it is very hard to speak of one aspect without addressing one of the others.

However there are some interesting conclusions to draw from this historical analysis. If we compare the growth of the internet in relation to geographical positioning we stumble upon interesting new insights. Asia is on the moment one of the biggest internet users markets of the world, and accounts for about 44% of the world's internet users (Internet World Stats, 2011). However Asia is just beginning to develop internet access to their population, and has an internet penetration percentage of a mere 24% (Internet World Stats, 2011). This means that the potential markets for Daum, Naver and smaller Chinese, Korean and Indian search engines is huge. So huge in fact that they could compete with Google. In Russia and surrounding countries Yandex has the same opportunity, with Russia only having an internet penetration percentage of a 43% (Internet World Stats, 2011). So in the future we might see huge changes in the search engine market.

What is clearly visible in the above insight, are the linguistic boundaries of the Chinese, Russian and Indian language. One of the aspects to keep an eye on is how Africa with also a lot of native languages mixed with western languages will react when internet is widely introduces there. Africa accounts for one seventh of the world's population but has the lowest internet penetration percentage of only 11% (Internet World Stats, 2011).

Looking further, we can see a correlation between the newly emerging closed systems and the changing of the core-businesses of internet search engines. Today Google, Yahoo! and Gmail don't present themselves anymore as internet search engines. They talk about their new mobile applications and new services like Google Books and Scholar. For years now all the attention of the media and search engines has been on the innovation and evolution of new services and not on search engines technology. We are quickly bored with difficult talks about 'intelligent' algorithms and new concepts of ranking web pages. We are more interested in the latest 'Google Discovery'. This means that Google is also spending more time developing the new 'Google Discoveries'. I think it would be wise if we would start asking more questions about how todays search engines operate and confine our internet search capabilities.

Other smaller insight can be provided looking closer at the different historical reviews I wrote in Appendix A. One of the most interesting cases I researched during the writing of this paper has to be Daum. Daum really combined modern social media with search engine capabilities from the beginning of 1995. For me it was clear that mergers between search engine technology and social media would be explored in the future. And as it turned out, Google announced Google Buzz in February 2010 and during the writing of this paper launched Google+ in 2011 (Google, 2010/2011). Google+/Buzz are kinds of social media platform that integrates with all the other Google Network applications and services. So Google is already busy with introducing social media aspects into their search engine capabilities.

Overall I believe that internet search development is for a long time now waiting for change. It hard to believe that in 10 or even 20 years' time we could come up with the ultimate search indexing solutions. While the first ten years of search engine development was very hectic and wild, later search engine development seems to be slowing down, and far more focused on services and applications than innovating search technologies.

Appendix A – A History Of Search Engines

Introduction – Why this Appendix?

In this appendix you will find an analysis of internet search engines throughout history. In my research I have come across many different historical reviews and analysis of search engines from 1990 until 2010. Many of these historical papers, articles and theses have varying opinions, analysis and standpoints from which the writers or companies write. Next to that, I found many histories of internet search engines that were already available on the internet and in publications. All of these show different opinions and known facts, and sometimes lack the right references or are simply written without references at all. Most of these historical papers also address a few years of history in length and do not address the whole development.

Within mind that there are varying opinions out there, I wanted to include my historical findings and analysis. I hope that with this analysis, the reader will get a clearer view of my research regarding different historical facts and search engines. While some search engine analyses go in to more detail than others, I hope it will give you an idea of what my research and analysis have brought to light. In this analysis I have tried not to express my own opinion too much and rely on the findings of my research. I have tried to use and analyze as much first hand literature as possible. Sadly, this is not always possible because new developments in the search engine market were mostly not developed anymore by University students, but by big corporate companies.

This historical analysis is divided into different sections because I found that they represent different stages in the development, innovation, use and size of internet search engines from those times. Each section covers a certain amount of years. However, the analysis of each search engine is not constrained within the boundaries of the different sections, but is placed in that section only because of the launch date of the search engine. For example, Google was launched in 1997 and is placed in the section of 1997 until 2000. But the analysis of Google contains a historical review from 1997 until 2010.

1990-1993

The invention of internet search engines

In order to understand the introduction of the first internet search programs in 1990, we need to know what internet was back in the 1980's. Manuel Castells addresses the early development stages of the internet in his book *The Internet Galaxy*. Castells writes about ARPANET as the roots of the internet as we know it today. In 1971 ARPANET consisted of a few nodes, all nodes were basically US University computers or research centers related to US Universities. The fact that Universities are entwined in the early development of the internet is later clearly visible in the introduction of the first internet search engines. In the seventies ARPANET started to grow and incorporate many other networks, this was mainly possible by the introduction of the inter-network protocol (IP) and uniting it with the transmission control protocol (TCP), forming the TCP/IP protocol which is still used today by internet. This allowed for many new nodes to be formed as more computers and network now connected to each other. In 1984 the US National Science Foundation founded NSFNET, a separate computer network from APRANET. In 1990 NSFNET would shed ARPANET's militaristic skin off, being a project of the American Defense Department, resolving in the decommissioning of ARPANET. The more open and privatized NSFNET, together with the introduction of computers with network capabilities, boosted the number of users and nodes of the network. By 1995 NSFNET was shut down making way for today's totally privatized internet.

We can learn for Castells that during the 1980's internet was beginning to grow at a fast rate. More universities and computer networks began to join ARPANET and later NSFNET. As we look at the dedicated work of Robert Zakon we see an extraordinary growth of internet hosts from 1987 to 1990. Where in 1987 the internet consisted of a mere 10.000 hosts, by 1990 this number was ten folded to 100.000 hosts worldwide. So the amount of accessible data and information was skyrocketing. Mostly this information was available through the use of the file transfer protocol servers (FTP-server) that where hosted. FTP servers shared files and data and made them available to all the internet users. These servers were nothing like a website or a webpage, they had no graphical interface just a list of files available to the visitor. Wes Sonnenreich writes that new available content on FTP-servers was mostly contributed by a sort of mouth to mouth communication, "Somebody would post an e-mail to a message list or a discussion forum announcing the availability of a file" (Sonnenreich, 1999). Someone looking for applications or information had to look up different FTP-sites and hope to find the wanted information. Searching information or files was far more difficult in those years compared to how easy it has become today.

Archie - 1990

Few other areas in the field of computer science hold out such promise for significant performance gains in the coming years as the field of computer networking. (Alan Emtage & Peter Deutch, 1992)

Alan Emtage worked and studied at the University of Montreal since 1983 (Metz, 2007). While earning his bachelor's and master's degree in computer science he worked as system administrator for the university. In 1986 Montreal University had one of the two internet connections in Canada (Metz, 2007). Emtage was instructed to search en gather new applications from the internet. He first began, as was usual at that time, to search FTP-server after FTP-server by hand and logging them in his own database. This was a time consuming task, and as you can imagine not quite fun. It was then that Emtage came up with a script that could search the internet by itself and fill his database with the needed information. It collected the file lists of FTP-servers and added them to his database. In PC magazine Alan Emtage is quoted: "Basically, I was being lazy" (Metz, 2007). Later on Emtage -with the help of Peter Deutsch- would open up his database to all the internet users, thereby creating the first internet search engine ever. Archie, as he named the search engine, became a hit. By the end of 1990 Archie was responsible for the half of Canada's total internet traffic (Metz, 2007).

Veronica – 1992 and Jughead - 1993

With Archie steadily growing with more users and larger search indexes, new technologies became available. The gopher-protocol was invented in 1991 by Mark McCahill, Farhad Anklesaria, Paul Lindner, Daniel Torry and Bob Alberti at the University of Minnesota (Frana, 2001). Where the file transfer protocol (FTP) was focused on distributing and retrieving files on the internet, the Gopher protocol was also invented to search the internet better. With the Gopher protocol it became possible to search the content of files and make search indexes relate to the content of files. This was not possible with the file transfer protocol which only provided search on the basis of file names and server listings. The introduction of the Gopher Protocol led to the creation of two new search engines, Veronica: developed by Steven Foster and Fred Barrie at the University of Nevada in 1992 (Jean Polly & Steve Cisler, 1995) and Jughead: developed by Rhett Jones at the University of Utah in 1993. Both these search engines used the Gopher protocol to make search queries possible. By using this new protocol the inventors hoped to improve Archie's capabilities and introduce better and smarter search engines. Veronica claimed to be "the grandmother of all search engines" (Salient Marketing) , this is mainly because it reached new high scores in search engine history. Veronica could search through 5,500 Gopher servers and search in 10 million Gopher indexed files (Jean Polly & Steve Cisler, 1995). But due to the immense size of the index and the constant updating of the index by Veronica, the system overloaded often. This resulted in errors and undeliverable search results. Jughead wanted to cope with this problem by searching only one Gopher server at the time, making it possible to show faster search results than Veronica. The downside to that was that searching the entire database of Gopher servers took ages. Veronica never really exceeded in surpassing their predecessor Archie, Gopher servers were running alongside of the already existing FTP servers, this resulted in the fact that Archie as well as Veronica and Jughead coexisted in the early years of the internet.

W3 Catalog - 1993

"Although the navigational possibilities of the Web were self-evident, it was not clear how one could (or should) provide query facilities for Web resources" (Oscar Nierstrasz, 1996)

In September of 1993 W3 Catalog was released by its developer Oscar Nierstrasz at the University of Geneva (Nierstrasz, 1996). The internet search engine had a new approach to search indexing. The idea behind Nierstrasz W3 Catalog was not the index all of the internet like Archie, Veronica and Jughead but to rely on the work of individuals that made listings of available content, as he himself wrote: "I noticed that many industrious souls had gone about creating high-quality lists of WWW resources" (Nierstrasz, 1996). Nierstrasz made scripts that allowed him to mirror these high-quality lists and reformat them so they were useful for his own W3 Catalog. So instead of mapping the internet by himself he relied on others to do the indexing for him. In this way he could guarantee that his search results would come up with the right kind of data, because these were checked and added by people and not programming scripts. This process of relying on others for indexing databases will be visible throughout the whole development of search engines.

World Wide Web Wanderer - 1993

"Want to know how big the Internet is? How fast is it growing? Will it ever stop?" (Matthew Gray, 1996)

In 1993 the now fully operation world wide web of Berners-Lee was doing very well. Adapting to the introduction of many new Gopher and FTP servers worldwide, the internet was starting to grow exponentially. In June that same year, Matthew Grey¹ developed the World Wide Web Wanderer working at the Massachusetts Institute of Technology (Gray, 1996). This program was design for only one purpose, mapping the extent and growth of the World Wide Web (WWW). The program was later named the first web-crawler to search and index the content of the WWW. A web-crawler being a program that

can search and index the web independently, mostly working systematically and intelligent to update and search the entire WWW. Later on the Wanderer started to not only count the number of sites but also index their Uniform Resource Locator's (URL's) (Gray, 1996). This list of page URL's was later called Wandex. Wandex was shortly used as a search engine but mainly by other search engines to expand and direct their own indexing scripts. The work of Grey is a real milestone in the development of internet search engines. From this moment on web-crawlers or what we call them today: spiders, web-robots or bots are inseparable with the way modern search engines work.

¹Now working at Google as software engineer

Jumpstation – 1993 and WWW Worm - 1993

“A fundamental problem with the World Wide Web (WWW) is the enormous number of resources available and the difficulty of locating and tracking everything. Manually generated archives are extremely laborious to develop, and are intrinsically non-scalable” (Oliver McBryan, 1994)

Next to the new technologies and search engines that were released in 1993, there was another great development that started in that same year. This was the introduction of Mosaic, one of the first graphical web browsers. Mosaic was far more appealing to the general public and easier to use. This resulted in again more users on the World Wide Web, all demanding quicker and better search capabilities. Robert Zakon addresses the introduction of Mosaic as: “Mosaic takes the internet by storm” (Zakon, 2010). Although in 1993 the growth of internet traffic of the WWW is 342% and therefore far less than Gopher's traffic growth of 997%, one year later the WWW far exceeds total amount internet traffic (Zakon, 2010).

To cope with the new pressure of hundreds of new sites and host, and not to speak of thousands of new search queries, two new search engines were launched by the end of 1993. Jumpstation developed by Jonathon Fletcher, student at the University of Stirling (Stirling University, 2009) and the WWW Worm developed by Oliver McBryan at the University of Colorado (Price, 2005). Both search engines used the idea of World Wide Web Wanderer to start indexing the WWW with the use of a web-crawler. This meant using spiders, before known as web-crawlers or robots, which crawled the internet site by site and picking up information as they went along. These were very simple spiders using the URL's presented on the home pages to continue their search through the WWW. “JumpStation's web bot gathered information about the title and header from Web pages” (Sonnenreich, 1997)

Although these spiders collected many data from the net, they were easy manipulated, and had no idea what the difference was between the top search result and the lowest. WWW Worm wanted to change that and tried to implement page ranking systems to cope these problems. With page ranking the WWW Worm hoped to apply hierarchy in their search engine results.

Aliweb – 1993/1994

“Robots have been operating in the World-Wide Web for over a year. In that time they have performed useful tasks, but also on occasion wreaked havoc on the networks.” (Martijn Koster, 1995)

Along with the WWW came the Hypertext Markup Language (HTML). This was a major difference compared with the Gopher and FTP protocol Archie, Veronica and Jughead were using. Internet was now on its way to gain ground on the two already existing protocols. With WWW users growing Martijn Koster develops a new search engine for the HTML/WWW standard. Aliweb, what stand for 'Archie Like Indexing of the Web', can be seen as yet another milestone in the development of internet search engines (Sherman, December 2002). Martijn Koster was a strong believer that web-crawlers or robots

should not be used to index the WWW. This is made very clear in the Robots Exclusion Standard (Koster,1994) which Koster writes a year later, and presents this to Organisation Européenne pour la Recherche Nucléaire (CERN), to not allow robots access to all the available content online. Koster saw a huge problem coming up with the way we use robots and crawlers online, "certain robots swamped servers with rapid-fire requests, or retrieved the same files repeatedly"(Koster, 1994). This had to stop before the internet would be overflowing with robot traffic instead of human traffic.

Aliweb did not make use robots or web-crawlers to index the internet. On the website of Aliweb you could subscribe the location of your index files (Sherman, December 2002). In that way Aliweb did not search the whole web but only the provided addresses of the index files. The upside of this method of indexing, was that users could add page descriptions and keywords to those index files that Aliweb would use to make up his own search database. Sadly not enough users submitted their index files and Aliweb was never a big success (Sherman, December 2002). But the idea of page descriptions and keywords later found his way to the standard HTML and is still used today as one of the ways of indexing websites.

1994-1996

Internet, from lots to a legion

Internet was now developing and growing at an alarming rate. While previous internet search engines were focused on the introduction of new concepts, development took a turn in these years. Companies and students at Universities became aware that the present internet search engines would not be able to cope with the growing amount of users and search queries. This caused for a shift in development of new search engines. These engines were far more focused on providing fast and large internet queries. This led to much development in the area of server side scripting and advance and smart web crawlers or spiders.

With new spiders and more servers now searching and crawling the internet, there also came awareness for what impact spiders or crawlers could have on the internet traffic. With that in mind more metasearch engines were brought in to life. These search engines used no spiders and crawlers to search the whole internet, but rely on other internet search providers to search the internet. Using results of other engine these search engines make up their own index database for search queries

WebCrawler - 1994

"In large distributed hypertext system like the World-Wide Web, users find resources by following hypertext links. As the size of the system increases the users must traverse increasingly more links to find what they are looking for, until precise navigation becomes impractical." (Pinkerton, 1994)

Looking at the results of earlier mentioned Wandex, developed by Matthew Grey (See World Wide Web Worm). We can see that Internet was now growing faster and faster. Where in 1993 there were 1,3 million hosts on internet, in 1994 this number was already grown to 2,2 million and would continue to grow to 9,5 million hosts in January 1996 (Grey, 1996). This meant that a lot more users and pages were needed to be indexed and searched.

WebCrawler was released in April 1994 and was developed by Brian Pinkerton at the University of Washington (Pinkerton,1994). It was released with the idea of creating a system that really helped people search on the WWW, like they would search in the library, along with that process came difficult problems: " But the World-Wide Web is decentralized, dynamic, and diverse; navigation is difficult, and finding information can be a challenge" (Pinkerton,1994). For finding the right information Pinkerton developed a way of reading the entire HTML pages that were available. This became known as 'full text search', a way of retrieving the entire information a page has to offer, not only the header or links like previous search engines did. This meant that search queries could now be made more specific and more elaborated. Where earlier search engines would come up with wrong search result when entering too many search words, WebCrawler was far more capable of coping with larger queries. This resulted in massive use of WebCrawler, so much that the site would not always be online due to the enormous amount of stress to the hosting server (Pinkerton,1994). Greg Notess writes about these performance issues in Online magazine 1995 as a problem concerning not only WebCrawler at that time. "WebCrawler can sometimes be as difficult to reach as Lycos. Once again, it is a victim of its own Success" (Notess, 1995).

MetaCrawler - 1994

"A common belief is that the majority of Web search services are roughly the same. They all index the same content, and all use reasonable Information Retrieval (IR) techniques to map a query to potentially relevant URLs" (Erik Selberg & Oren Etzioni, 1996)

In June 1994 Erik Selberg (student) and Oren Etzioni (professor) at the University of Washington developed MetaCrawler. In the beginning they started to examine different internet search engine at the time en combining the search results. What they found is that there are major differences in the search results that search engines provide. Differences consisted of: the age of the found results (refreshing rate of the server indexes), the relevance of the age of the page is considered or not, and huge differences in the ranking algorithms. They conclude that: "One service might rank pages that are highly relevant to a given query in the top positions, whereas another may not" (Selberg & Etzioni, 1996). Combining these search results would result in interesting conclusion about where, when and how pages where indexed. MetaCrawler started to combine search results of major search engines and add possible searching features to widen or narrow the search results. The idea was that with knowing how different search engine operate, you could provide better 'trailer fit' search opportunities to users. MetaCraweler concept of using other search engine indexes to compose new search indexing is now called 'metasearch'. In this way MetaCrawler opened up a whole new notion to internet search and believed in the idea that the information out on the internet was in fact indexed (or being indexed), but not well presented to the public. "The user need know only what he or she is looking for the MetaCrawler Softbot takes care of how and where" (Selberg & Etzioni, 1996).

Quickly MetaCrawler was growing out of the University capabilities. It was in 1995 that NetBot was founded by the University of Washington to startup MetaCrawler together with 2 other Univeristy programas. With NetBot growing in different direction MetaCrawler initiators decided otherwise and was sold to starting up internet company Go2Net. Go2Net was purchased by InfoSpace Inc. in 2000.

Lycos - 1994

"One of the enabling technologies of the World Wide Web, along with browsers, domain name servers, and hypertext markup language, is the search engine. Although the Web contains over 100 million pages of information, those millions of pages are useless if you cannot find the pages you need." (Micheal L. Mauldin, 1997)

Lycos was developed by Micheal L. Mauldin in 1994 as part of a research program at Carnegie Mellon University (Mauldin, 2011). The spiders and crawlers of Lycos were made to rank and retrieve large amounts of text and data. Next to Lycos's ranking algorithms, Lycos incorporated word proximity results: results of what you might have meant with a search query. By the launch of Lycos in 1994, Lycos spiders had already indexed 54.000 documents (Mauldin, 1997). This number kept growing and growing, to a point that in 1994 Lycos ranked first in Netscape's list of: coming up with the most result on the query 'surf'. But still Lycos kept growing in size, by November 1996 "Lycos had indexed over 60 million documents--more than any other Web search engine" (Mauldin, 1997). These kinds of numbers were unreal for those times and showed the world how fast and huge the internet was growing.

During this enormous growth of Lycos and its popularity, Lycos began to set forth a trend that would last until today. Lycos began purchasing other internet search engines and incorporate their search indexes and market shares, while leaving the sites operational. Next to that Lycos began transforming their search engine to a wider online platform/portal. As stated on their website "Lycos's award-winning products and services include tools for blogging, web publishing and hosting, online games, e-mail, and search. (...) The Lycos Network of sites and services include Lycos.com, Tripod, Angelfire, HotBot, Gamesville, WhoWhere, and Lycos Mail" (Lycos Inc. 2011) Were other sites at the time, like Excite, never really exceeded in

becoming a large online portal, Lycos capital and timing allowed for an immense growth of users and market share by the end of the century.

In 2000 Lycos was bought for an astonishing 5.4 billion dollars to Telefónica, during the great internet bubble around 2000. Four years later in 2004 Lycos was again sold, then for a mere 95.4 million, to Daum Communications Corporation (Daum, 2011). Today Lycos is still a player in the search engine industry now focusing on mobile, social media and location based applications and servers, and was Sold to Ybrant Digital in 2010 (Business Wire, 2010).

Inktomi - 1994

"We had an opportunity to not only affect the search engine space but the architecture of the Internet" (Eric Brewer in Bowen 2000)

Inktomi was developed by Eric Brewer, professor at the University of California, Berkeley in 1994. The idea behind Inktomi was to create and form a whole new kind of search engine that could cope with the upcoming internet 'boom' Brewer foresaw (Brewer, 2010). Brewer manages to form computer networks to cope with the immense pressure of thousands of search queries, distributing the workload among many servers. (Brewer, 2001) This allowed him to build the unique search engine HotBot in 1996 which was powered by Inktomi servers and software. The success of HotBot was quickly visible and other search engines and search providers started using Inktomi. In 1997 Microsoft started using Inktomi clients for their search engine and in 1998 AOL also started using Inktomi software and servers (Brewer, 2010). Later Yahoo! would also follow Microsoft and AOL.

Inktomi was resolved into all kinds of smaller companies after the internet bubble by the end of the century and was almost gone by the end of 2002 (Brewer, 2010).

InfoSeek - 1994

"Our goal is to enable someone to do a comprehensive and accurate world-wide distributed search in less than 1 second." (Steve Kirsch, 1996)

"Infoseek Corporation is dedicated to making information easy to find. The company's services benefit all Internet users, whether they are casual World Wide Web surfers or business professionals whose jobs depend on up-to-date market information" (InfoSeek, 1996)

Steve Kirsch invented new methods of retrieving information from multiple large databases. This meant that servers were now better and faster accessible. Next to that Kirsch foresaw great problems if every search engine in the world would start indexing the internet (Kirsch, 1996). Kirsch thought that this would ultimately lead to the clogging up of the internet in a few years' time. Because of these irritations, along with some other critiques, Kirsch founded InfoSeek in 1994 (Kirsch, 1996). Infoseek would grow out to be one of the best search engines in 1996, having the largest categorized web pages at that time (InfoSeek, 1996). In the ideology of Kirsch that the web was clogging up with useless information, InfoSeek would come up with one of the first scripts for behavioral targeting in 1997 (Kirsch, October 1999). The program was called UltraMatch, this meant that advertisement would be more adaptive to the users behavior (Kirsch, July 1999).

In 1998 InfoSeek was almost completely bought by Disney (CNN Money, 1999). This caused InfoSeek to team up with Disney's search engine Go.com. Part of the software was later sold to Inktomi in 2000.

Excite - 1995

"When we (me and my partners) started Excite we were buying compilers, development environments, web servers, etc. The software infrastructure costs were real. Now, the open source community has done such a good job of making rock-solid infrastructure that this cost is all but gone." (Joe Kraus, 2005)

In 1994 Architext was developed by 6 Stanford Graduates: Graham Spencer, Joe Kraus, Mark Van Haren, Ryan McIntyre, Ben Lutch and Martin Reinfried (Red Herring, 1995). Starting in a garage in Silicon Valley, the boys wanted to create a search tool for coping with large amounts of data (Sherman, October 2002). The large amounts of data on the internet provided the ideal pool of data in order to test their program. Competing against other companies to produce new software the team quickly found themselves working to make deadlines and show off working demos. After that the boys got their first large financing deal of \$250.000, from Vinod Khosla working at Kleiner Perkins Caufield & Byers (Red Herring, 1995). This led to ever greater successes, and investments would lead up to 1,5 million from Khosla and another 1,5 million from Geoff Yang of Institutional Venture Partners. "I saw six really smart guys, who were extremely entrepreneurial, and pioneering a technology applicable in a hot new field" (Geoff Yang in Red Herring, 1995). In 1995 the software program called Architext was newly released under the name Excite. (Graham Spencer, 1995)

Excite would skyrocket in the upcoming years. The success quickly started and in 1995 with Excite buying two other internet search engines, Magellan, developed by David Hayden in 1993 (ResumeBucket.com) and WebCrawler (See Webcrawler). This was one of the first major takeovers in search engine history. In the end of December 1995 Excite was worth over 70 Million dollar (Kraus, 2005) and continued to grow. "The company went public in April 1996, at a valuation of \$177 million. Over the course of the next several years, its market value rose tenfold" (Sherman, October 2002). During 1996 and 1997 Excite would start a whole new type of search engine, combining the search engine with mail services, photo searching, personal ads and many other applications and services. Excite grown out to be one of the first internet portals, not only providing search but everything someone would need online. An internet portal.

The company was fused with the @Home network in 1999, due to major losses in first quarter of 1999 (Rohrlich, 2010), eventually forming Excite@Home. Many years later in 2005 Excite was acquired by Ask Jeeves known as Ask.com today (See Ask.com).

AltaVista - 1995

"Finding a 'cyber needle' in an ever-growing 'cyber haystack' has long been a dream of Web users. This technology is a major step in that direction. It features the most complete and precise Web index. In fact, it's the only one with the potential to keep up with the phenomenal growth of the Web" (Samuel H. Fuller in Digital Equipment Corporation, 1995)

AltaVista was developed by the Digital Equipment Corporation in 1994 and released in December 1995. The idea of AltaVista was to cope with the immense growth and expansion of the web (Digital Equipment Corporation, 1995). Other search engines had difficulty in expanding their web indexes or cope with the immense number of search queries by users. With the development of 64 bit servers instead of 32 bit servers Digital hoped to extend their hardware capabilities providing users with a new 'Super Spider' (Digital Equipment Corporation, 1995). This new spider could crawl past 2,5 million pages a day and was 100 times faster than already existing internet spiders. Providing these kinds of numbers and combine this with a full text search like WebCrawler was groundbreaking. Brian Pinkerton, developer of the WebCrawler search engine, commented in The New York Times "Webcrawler was asked to conduct 30,000 search requests a day. Today, the volume has increased to 2.5 million requests a day, he said" (Pinkerton in Lewis, 1995). On the first day, after the launch AltaVista, de site had coped with 300.000 hits (Lewis, 1995). This number would increase to 80 million hits per day two years later.

That AltaVista's services were phenomenal was not only visible in the number of users and market share, but in 1996 AltaVista became exclusive provider of search results for Yahoo!. During that same time they started multi-lingual search, searching in Japanese, Korean and Chinese. Because most of the programming scripts were already there, AltaVista began hosting BabelFish, a translating service that was popular well after the success of AltaVista.

AltaVista was purchased in 2003 by Overture, what in its place was purchased by the end of the same year by Yahoo!. AltaVista's website still exists today, but as a mere copy of the Yahoo! Search engine.

Yahoo! - 1995

"It was irreverent, it was reflective of the Wild West nature of the Internet, and a lot of people found it easy to remember, which we thought was probably good." (Jerry Yang in Michael Krantz, 1998)

David Filo and Jerry Yang started out as two Stanford University students (Yahoo!, 2005). There they were trying to keep track of the large amounts of data that was available on the web and especially on their University infrastructure. So they started making lists of accessible data (Yahoo, 2005). Quickly they started to realize there was so much information for one list and so they started to categorize the list. Eventually the categories were also too large, so they divided them as well into smaller categories. It was in 1994 that this 'guide' was published on the web as "Jerry and David's Guide to the World Wide Web", later the name was transformed into Yahoo!, what stands for "Yet Another Hierarchical Official Oracle" (Yahoo!, 2005)

In 1995 Filo and Yang would search for investment money to launch Yahoo! with the right support of hardware and people. They found that money in the same year in a 2 million dollar investment from Sequoia Capital (Yahoo!, 1997). With that investment Yahoo! quickly started to develop and grow, hiring staff and expanding their search engine speed and possibilities. Where AltaVista, Lycos and other major search engines were mostly relying on spiders and crawlers to categorize and rank pages, Yahoo! was hiring staff to do that exact same work. "No technology could beat human filtering," (David Filo in Stross, 1998). This was mainly possible because of the public offering of shares in 1996, earning the company 33.8 million dollars (Yahoo!, 1997).

Over the next year Yahoo! would do the exact same thing MSN and Lycos and Excite were doing at the time, buying other smaller companies. Yahoo! acquired many companies from 1997 until 1998 and launched Yahoo! Mail, Yahoo! Games, Yahoo! Groups and Yahoo! Messenger (Stross, 1998). All these extra services were mostly other internet companies, which were bought and transformed to Yahoo! related services and applications. This was done in order to form a Yahoo! Network and establish Yahoo! as an internet portal.

The speed of Yahoo!'s internet services was becoming a real problem for the company (Stross, 1998). With rivaling companies boosting their search speeds using new algorithms and better server hardware systems, Yahoo! finds himself coping to keep up the demand of his users. In 2000 Yahoo! signed an agreement with Google, allowing them to power all the searches made on yahoo.com (Google, 2000). With the quick and smart algorithms Google servers had at the time, Yahoo! could provide faster and better search results to their customers. In 2002 this service would be done by Inktomi.

Over the last ten years Yahoo! has endured the presence of Google, but is nothing compared to their previous glory, looking at market shares. Although shrinking in size, Yahoo! is still one of the most innovating and biggest search engines in the U.S. today (comScore, 2011)

Daum - 1995

Daum was founded in 1995 as one of the first search engines in Asia, South Korea, with Jeawoong Lee as one of the co-founders (Daum, 2011). With the internet market opening up in Asia, the company quickly grows to a sizeable and important Asian search engine. With the launch of Hanmail in 1997, the Daum corporations free mailclient being one of the first in Asia, users kept rolling in (Cho Jin-seo, 2007). In the first nineteen months Hanmail had over one million registered members¹ (Daum, 2011). 1999 Daum's network grew even more with the social network platform Daum's Café, which attracted new users to the online forums and news groups. By the end of the century Daum was the second biggest portal in the whole of Asia just behind Yahoo! (Daum, 2011).

With Daum now expanding en growing in the whole of Asia there was room for even more room for expansion. In 2003 Daum released Media Daum, a platform where people could connect and chat with each other (Daum, 2011). But above all Media Daum was a news site, and become the most visited news site at that time.

In 2004 Daum was struggling to keep up with Naver and Nate, to rivals of Daum in the Asian search engine market. To counter this, Jeawoong Lee still CEO of Daum at that time, bought Lycos to expend across borders and innovate Daum's abilities (Cho Jin-seo, 2007). "The acquisition of Lycos is a springboard for Daum to venture into the U.S. market and become a global player" (Lee, 2004). Daum has been trying new innovative ideas ever since, but lost a lot of market share in the whole of Asia (Schneider, 2010).

¹Today Hanmail has approximately 38 million registered users.

Go.com - 1995

"Radio stations shouldn't have to pay to have a Web presence" (Jeff Gold in Atwood, 1997)

Jeff Gold founded Go.com in 1995. The site was meant as an entertainment portal for the web in the early years. Go.com was one of the first with chat room capabilities and was working together with almost 3000 radio station (Atwood, 1997). Over the course of the next 3 years almost 1 million people became members of the network. Due to the success Gold was having with Go.com as entertainment provider online, Disney bought the company in 1998 with the idea of extending Disney entertainment for the whole family (Du Bois, 2001). Gold's Go.com was mixed up with InfoSeek, also bought that same year, to create the new online portal for Disney (Du Bois, 2001). Disney's Go.com never really exceeded further but is still used today as a network site of the Disney online entertainment network. Some say it was Disney's own demise that let the site down, and people don't look for a corporate site to start their search from. "Why try to create a whole new Internet identity when they already have good brand names on the shelf? There hasn't been a huge demand [for Go.com] from consumers." (David Marks in Du Bois, 2001)

AOL/America Online - 1995

"AOL is a good family brand. The reason you use it is because it is a Web site you can trust" (Jason Helfstein in Holahan, 2006)

In 1989 AOL started off as an instant messenger service with a background for game development (AOL, 2011). It was not until 1991 that the company was called America Online after an employee contest. In 1993 the company started to connect households to the internet, and by 1995 connected 1 million households to the internet (AOL, 2011). With so many people

now connection to the internet, AOL stated to focus to their online presence and launched AOL.com (AOL, 2011). In one year time the website would have five million subscribed members (AOL, 2011).

By 1998 AOL was so huge that is bought CompuServe: Internet provider and ICQ: instant messaging service. In 1999 AOL bought Moviefone: Movie network site and Netscape: Internet browser as well (AOL, 2011). With this purchases AOL was provider of the internet network, internet provider, browser provider and search and news provider all in one. In 2001 the power of AOL grew even larger with the merger with Time Warner: one of the biggest entertainment en media companies that ever existed. That same year OAL had 35 million subscribers online (AOL, 2011).

Over the last 9 years AOL has developed themselves as a unique company for online news, articles, software, advertisement and messaging, together with the many hardware and smart offline services and software. In 2007 USA Today set AOL as the fourth biggest change in the history of the internet (USA Today, 2007). AOL is still one of the biggest and most influential companies regarding internet usage and access in 2010.

Dogpile - 1996

Dogpile returns all the best results from leading search engines including Google, Yahoo! and Bing, so you find what you're looking for faster. (Dogpile, 2011)

The Dogpile search engine was developed in 1996 as a metasearch engine¹. Combining the top search results from other search engines provides Dogpile with its own index. This was done with the belief that the top search engines have hardly any overlap in their search result. With that conclusion they made started to combine the search results of the top search engines. In a publication written in 2007, Dogpile shows that major search engines such as Google, Yahoo!, Ask and Live have hardly any combining top search result (Dogpile, 2007).

Dogpile was faster and better adjustable then MetaCrawler at that time and therefor far more successful. In 2001 InfoSpace bought Dogpile in order to convert their own search engines to the Dogpile metasearch engine. In 2002 InfoSpace launched the new search engine for all their allready owned search engines (InfoSpace, 2002). Thereby claiming the new search engine would be twice as fast en sufficient then the previous engine. Today "Dogpile aggregates the most relevant searches from Google, Yahoo! and Bing and delivers them to you in a convenient search package" (Dogpile, 2011).

¹see MetaCrawler

1997 - 1999

The garden of search and the beginning of the internet bubble.

In the last 2 years new internet search engines were beginning to provide not only search engine capabilities anymore. Lycos and Excite were among the first to begin adding extra features and services to their search engines. Mostly these were news, mail and image search capabilities. While these search engines were developed initially as mere search engines, the search engines were now changing in becoming internet portals. Portals that could provide any wanted information and application the user wanted. Around 1997 to 1999 a new breed of search engine entered the stage. Search engine that were not only designed as mere search tools, but also as mail, messenger, news and blogging companies.

This was also possible due to the fact that the speed of server scripts and intelligent spiders was beginning to fade. This was because the rate in with new chips and processors were invented was exponentially growing. So now more work could go to the development of adding new features and possibilities to search engines.

By the end of the century the internet market caused for an extreme growth of market values, later referred to as 'the internet bubble'. Internet based companies saw their stock price rise to hundreds of dollars, due to the fact that the economy and investors were spending millions to starting and existing internet companies. It was a golden age for internet based companies.

InfoSpace - 1996

"Since 1996, our mission has been to make finding what users need fast and easy" (InfoSpace, 2011)

InfoSpace was founded in 1996 by Naveen Jain, an Indian entrepreneur from India tired of working for Microsoft (Jain, 2010). Starting InfoSpace was a great step for Naveen and in an interview in the Red Herring he said about founding InfoSpace "like a teenager going to his father and saying, 'Dad, I can do it on my own'" (Jain in Red Herring, 1997).

Indeed, InfoSpace was doing really good, by 1998 InfoSpace went public and raised 75 million dollars. (Wired, 1998) In 2000 InfoSpace starts its buying streak. The first was Go2Net a search provider and small portal also providing search for MetaCrawler and Dogpile (Hu, 2000). Later other wireless providing companies and game developers would follow. With MetaCrawler and Dogpile on board, along with a couple smaller search engine and niche websites, InfoSpace survived the internet bubble by the end of the century and still exists upon today. All the internet search engines are now transformed to metasearch engines following the idea of MetaCrawler in mind. (InfoSpace, 2011)

Hotbot - 1996

"So, incremental scalability could provide a large economic advantage over the longer term. Not only the hardware, but the software design must take all this into account" (Eric Brewer, 1998)

Hotbot was launched in 1996 by the Inktomi Corporation in collaboration with Wire Magazine, and finds its roots in the Berkeley NOW project (Eric Brewer, 1998). Hotbot was one of search engines that effectively made adjustment to the way server scripts worked. Because the new script were much faster and much more efficient. This caused Hotbot the search

much more faster than any search engine at the time, and could search 10 million pages a day (Brewer, 1998). This was a lot faster than Hotbot rival, AltaVista at that time.

With all that search capability HotBot was able to update internet pages a whole lot faster than the other search engines. This would lead to a database that would have over 54 million pages indexed, which was a record at that time (Hock, 1997). Hotbot can be seen as a metasearch engine, but not directly. It took the results of other search engines at the time. But instead of relying on the index pages of others, Hotbot scanned them himself again and placed them into his own database. It was in this time that it was discovered that the major search engines at that time had not many overlap in terms of indexed pages (Brewer, 1998).

Some unique features were added in Hotbot's search engine. You could choose in which search engine you wanted to conduct your search queries. This allowed users to set off unwanted search engine in their search, and narrowing down unwanted information. Next to that Hotbot later introduced a way of searching in previously done searches (Brewer, 1998). This was a function well used by Hotbot users, and was considered one of the innovations HotBot has made.

In 1998 Lycos bough Hotbot from Inktomi and Wire Magazine and remained in their possession until today (Wired, 1998).

Ask.com / Ask Jeeves - 1996

"Every day, millions of people come to us with questions - personal and professional, easy and difficult, silly and serious. At Ask.com, our mission is to answer these questions with the best information from the web and from real people - all in one place." (us.Ask.com, 2011)

Ask Jeeves was founded in 1996 in Berkeley, California by David Warthen and Garrett Gruener (sp.Ask.com). Ask Jeeves was not designed to perform as a usual search engine. Warthen and Gruener wanted to change the way people where always bound to use keyword based search methods. They wanted to make search more easier and closer to the way people want to get information. So they came up with the idea that people always start asking questions if one would seek some sort of information. With this idea in mind, Ask Jeeves was created: a search engine that would be able to answer anny questions users might ask the program (Sherman, 2003). The 'Jeeves' in Ask Jeeves, stands for the iconic figure of a butler, that provides these services of answering questions to the users. Jeeves was erased from the company's image in 2005, also changing the name of the company to Ask. From 1996 to 2010 Ask never really reached the top of the search engine market, but Ask never had problems remaining a player throughout these years. Ask is a very clear and good example that usual search engines methods are maybe not the best way of providing search capabilities to users.

In 2001 Ask took over Teoma, a small but very good search engine at that time that relied on their own search engine scripts (Sherman, 2003). Since 2001 Ask.com is being powered by Teoma's search engine scripts.

Yandex - 1997

Today the word Yandex has become synonymous with internet search in Russian-speaking countries. (Yandex, 2011)

Arkadia was an search company that dates back all the way through the mid-eighties, back than the company started off as an indexing company providing search en index programs (Yandex, 2011). In 1993 Arkady Volozh and Ilya Segalovitch, founders of the company, invented the word Yandex and renamed the company (Yandex, 2011). It was not until 1997 Yandex launched their website yandex.ru with their own search engine and own indexing system. The system accounted

for language morphology and used algorithms to define relevance of internet pages. Three years later Yandex would be the leading search engine in Russia, and still is today. (Yandex, 2011)

During the last 10 years Yandex has innovated themselves using new talents and new startup programs to innovate their own search capabilities. This includes a School of Data Analysis, a two year master program. Next to that are to investment programs called Yandex.Start and Yandex.Factory which supports new upcoming talent in the data analyzing and indexing sector. (Yandex, 2011)

Google - 1997

“Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems” (Sergey Brin and Lawrence Page, 1998)

In 1996, two Stanford University students, Sergey Brin and Lawrence Page, begin their work on a project called BackRub, a search engine that would be used by Stanford University students and others (Google,2011). In 1997 the search engines grew out of proportions and the university could no longer provide the search engine with the right bandwidth. By the end of 1997 Brin and Page decide that BackRub should be renamed, and so rename their search engine to Google (Google, 2011). Google was inspired by the web performance and size of the internet at that time.

The two students talk about internet search engines in in their paper regarding the ideas and ideology of Google, *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (Brin and Page, 1998). “Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all esoteric topics. Automated search engines that rely on keyword matching usually return too many low quality matches.” (Brin and Page, 1998) You can conclude that Brin and Page were not satisfied with the way internet search engines operated at that time. The idea of a search engine that could answer all the questions and come up with all the information easy and quickly was far from realized. “Our main goal is to improve the quality of web search engines. In 1994, some people believed that a complete search index would make it possible to find anything easily. (...)However, the Web of 1997 is quite different.” (Brin and Page, 1998). To come up with a solution to these problems they introduce a page ranking system called PageRank, a system that takes the amount of links that follow to a website in account. If more links lead to one website, this website should contain good and probably valid information. Later Lawrence Page would get U.S. patent right over this PageRank system (Page, 1999).

With the page ranking system retrieving good quality webpages, providing valuable and good information to its users, Google immediately starts to expand in users and capital. While in 1999 the two founders find that Google is taking up too much of their time, they try to sell Google to Excite, but they refused their offer (Google, 2011). In the next years Google tries to expand their capabilities with taking in more employees and introducing more search features next to extra and new services. In 2000 Google already had the largest web index that ever existed counting 1 billion indexed webpages (Google, 2011). December that same year Google would introduce the Google Toolbar, making search queries possible without having to go to google.com.

In the upcoming years Google would introduce many unique or good scalable and adjustable programs. Google image, Gmail, Google maps, Google Search Appliance, Picasa, Blogger, Google Books, Google Scholar and a whole lot more applications and services. Google is currently the market leader worldwide in internet search providers (Alexa.com, 2011). This dominant position, together with the worldwide use of their services and applications, forms the entire Google Network. This Google Network can almost be addressed as a monopoly in the global search engine market.

MSN Search 1998-2005 / Windows live 2005-2007 / Live Search 2007-2009 / Bing 2009-2011

MSN Search was launched by Microsoft in 1999. Over the course of 1999 to 2005 MSN Search was related to Microsoft's email client Hotmail together with Microsoft's MSN instant messaging service. Microsoft had not relied on their own search engine but used the search engine Inktomi, Looksmart, Altavista and Yahoo! to come up with their search results.

Windows Live Search was introduced in 2005 and was the first self-made search engine of Microsoft (Perez, July 2005). Windows Live Search replaced MSN search and took over all his functionality and integration in all Microsoft products. Now searches were not coming from Yahoo! anymore but were processed by Microsoft's own search engine (Perez, February 2005). The engine had the capability to search Microsoft's Encarta encyclopedia together with MSN Music. In 2006 Windows Live Search added their own scripts of searching pictures as well. In 2006 Windows Live Search imbedded a new feature trying to cope with the growth and expanding from Google. The idea was that Windows Live Search would be a cross-media search engine, able to cope with not only online content but also could search the whole computer (Lai, 2006). In this way combining the PC and online searches under one engine, making search faster and more efficient.

Live Search was launched in 2007, two years after launching Windows Live Search. Live Search was designed to have more income and more appeal compared to Windows Live Search. With combining the existing search engine with Microsoft's adCenter (Foley, 2007). This new merger meant that the capability of combining online and offline search was discarded, due to the fact that you don't want commercials and ads in an offline environment.

Coping with identity issues, because of Live Search is still too much attached to all the other Windows Live products and services, Microsoft renames and transforms Live Search in 2009 to Bing (Singel, 2009). Bing has a lot more complex features such as improved local search capabilities and indexing review sites and therefore showing ratings of services, drinks, foods, wait time and so on. (Singel, 2009) With the introduction of Bing also came one of the biggest deals of the past years in the search engine market. Yahoo! decided they needed a new partner in the battle against Google and finds this in the form of Bing. Bing's search engine will power Yahoo! searches, in return Yahoo gets advertising capabilities on Microsoft sites. In recent years Bing is not a real competitor to Google's success, but it does show higher growing rates than Google at the moment (McGee, 2011).

1999 - 2002

Surviving the bubble, and the start of the market share wars.

The internet bubble had to burst at one time. That happened in 2000 when the whole internet market collapsed. Major investment firms selling millions of stock belonging to major internet companies. That together with the loss in faith because of major lawsuits and large internet companies stalling in their growth and innovation caused the stock prices to reach newly discovered deeps. Some internet search engines buckled under the immense losses and were bought for almost nothing compared to a few months before.

With only a handful of search engine companies left, and most of the rest bought by that handful, a new way of market domination began. Search engines started to reach out globally investing millions of dollar to South America, Asia and East Europe. Yahoo! and Google were fighting all over the world to expand their search engine market and buy or outsmart native search engines in those countries.

AlltheWeb - 1999

"Fast's mission is to deliver leading real-time search and information retrieval technology to our clients" (John Lervik in Bowman, 2002)

AlltheWeb was developed by Fast Search and Transfer (FAST) in 1999. The company developed new server scripts that provided the company with much faster servers for searching large indexes (Gran, 1999). Were old serves where only capable of doing one search at the time, the FAST servers were able to coop with more than 1 search at the time. This gave the company a real advantage in comparison to AltaVista and Yahoo! (Gran, 1999). In 2002 AlltheWeb surpassed Google in the amount of indexed pages, indexing 2.1 billion pages versus Google's 2.07 billion (Bowman, 2002). But this was not the biggest advantage AlltheWeb had. Due to the immense speed of the FAST servers, AlltheWeb's refresh rate of their index was much greater than Google's (Microdoc, 2003). This was seen as one of the major advantages AlltheWeb had compared to Google. In 2003 AlltheWeb was bought by Overture, which was itself taken over by Yahoo! in 2004. From that day on Yahoo! used AlltheWeb scripts in their own search engine. In 2011, Yahoo! closed AlltheWeb.com, and let it redirect to www.yahoo.com (Rao, 2011).

Naver - 1999

"the emperor of the domestic Internet portal business is Naver" (IT Times, 2007)

Naver was developed by the Next Human Network (NHN) Corporation in 1999 (Bonfils, 2011). Naver derives from the idea that an internet portal should be something where people can ask question to the community (internet), and get answers in return from the community (internet). This changes the whole idea of a search engine upside down. Were regular internet search engines do not provide content or provide the means to make user generated content Naver does. The idea of the internet search provider is laid out in the simple basics of helping one another. Choe Sang-Hun describes this in an article in the New York Times. "Tapping a South Korean inclination to help one another on the Web has made Naver.com the undisputed leader of Internet search in the country" (Sang-Hun, 2007) What Naver does, is giving the users the ability to ask questions, and on the other side answer these questions. What emerges is a system of people helping each other out. The man on the street is quoted in the New York Times saying, "No one pays me for this. But helping other people on the Internet is addictive" (Sang-Hun, 2007).

Choe Sang-Hun writes about the difficult market that is South Korea, the place where Naver originates from and still is by far the number one search engine. For Naver there was no indexed material available in 1999, so they came up with this idea to fuel their database with content to allow for search queries. Today Naver combines his own question database with many other databases, online magazine and internet pages forming the ideal balance of hand to hand user generated content with a state of the art advertisement algorithm.

With Asian markets now being fuel by these new kinds of Asian internet engines, meaning Naver together with Baidu, the new Asian search market was open to raise to new highs. Within three years Naver had surpassed Yahoo!, which was market leader in South Korea around the end of the century (It Times, 2007). With opening the internet cafes of Daum for Naver, both companies again rose to new highs in users and search queries (Sanh-Hun, 2007) In 2007 Baidu (China) and Naver(South Korea) became global players in the internet search engine business and ranked at the number 3: Baidu and number 5: Naver (Lipsman, 2007). This shows that the Asian search engine market is expanding and is growing in size and capital. Today in 2011, "One-third of the Korean population visits Naver.com every day, according to Koreanclick. More than 130 million daily queries are conducted daily on the search engine" (Bonfils, 2011). Naver is now trying to reach more Koreans across borders and is expanding their search engine across the sea to America. (Bonfils, 2011)

Teoma - 2000

"While the search engines and related tools continue to make improvements in their information retrieval algorithms, for the most part they continue to ignore an essential part of the web – the links." (Apostolos Gerasoulis, 1999)

In 2000 Professor Apostolos Gerasoulis at the Rutgers University in New Jersey, together with some fellow colleagues, develops the Teoma search engine. In their paper *DiscoWeb: Applying Link Analysis to Web Search* they address the irritation and the lack of innovation in the search engine market. Teoma was supposed to dig deeper into the internet's structure and collect better information using 'authority pages' (Roush, 2004). These pages could be found by analyzing the link structures of certain sections of the web. Establishing which sites were linked the most, is then not enough. You need to find dedicated communities and find the real mostly used pages or 'authority pages'. These pages can then be indexed for links and content providing better and more accurate search queries. (Roush, 2004). This Discoweb program had some problems, being very slow for instance, but showed real promise. "We believe that there is much promise in the use of link analysis for improving search engine retrievals as well as other tasks"(Gerasoulis, 1999).

Later the concept of Discoweb would get shape under the name Teoma (Roush, 2004). The new internet search algorithm was bought by Ask Jeeves in 2001 and went to provide the search queries for Ask.com. (Gerasoulis, 2010) Teoma is still used upon this day by Ask.com.

Baidu - 2000

"And as more and more Chinese come online, Baidu continues to innovate to meet their increasingly diverse tastes. With our goal of best serving the needs of our users and customers with intelligent and relevant solutions, we look forward to a robust future" (Baidu, 2011)

Baidu is the largest internet search provider in China at the moment, having 73% market share (Resonance China, 2011). Baidu was developed in 2000 by Robin Li (Baidu, 2011). Li had worked for InfoSeek in the past and had developed RankDex in 1996 apart from that (Rankdex, 2011). RankDex was an internet search engine that has many similar techniques to

Google's PageRank system. Rankdex is even cited in Google's PageRank patent written by Lawrence Page (Page, 2001). Li launched Rankdex in 1996, meaning that Google wasn't the first using page ranking capabilities in a search engine.

After Li decided to develop Baidu's search engine, Li reacquired the licensing for the RankDex search engine and used it for Baidu's search engine. Baidu was developed with the idea of having more knowledge of the Chinese language. Where Google and Yahoo!, still remained search engines that were written for the western language and script, Baidu was one of the first internet search engine to be developed with the Chinese language in mind (Baidu, 2011).

Over the last 10 years Baidu has introduced many of Google's innovations among other web innovations such as social media applications and an online encyclopedia, all focused on the Chinese language instead of the western languages. With these applications and services dedicated to China, Baidu has secured their position in the potentially largest internet user population of the world, now already counting 477 million users (Baidu, 2011). They are now looking to expand their search engine across the whole of Asia and maybe across seas, reaching out the worldwide Chinese speaking communities and cultures.

2002 - 2010

The smoke clears, everything is quite now.

While continuing to expand search engines throughout the whole world market shares become divided between only a few search engines. Google, Yahoo!, AltaVista, Ask, AOL and Microsoft are leading the search markets in most of the world. Were in the beginning of the 21st century there was a lot of fight for the dominant positions in countries and applications, all have become quite. Around 2010, Google, Yahoo! and Bing suddenly go quite in new innovating search technology. More focuses on other applications and the rise of social/personal media cause the major search engines to focus more on those aspects.

New technologies and innovation new grounds as mobile search and mobile web applications are immediately bought by the major search engines and incorporated into their own branding. This causes for a quite internet search market. But also brings along a lot of understanding and questions regarding the dominant internet search engines.

In recent years Asia is starting to fight for their own dominant search engine. Baidu is now rated in the top ten search engines globally and start to break through in the global search engine market.

Appendix B – Numbers and percentages

Introduction

With the research that I conducted it is possible to create a clear estimate of how the internet/world wide web has grown and how large today's internet search indexes and engines have become. The below numbers and percentages can be seen as a guideline for the real estimate. However, many of the numbers presented here are from the makers or developers of a lot of these search engines and will give us a clear view how internet search engines together with the internet has grown. The numbers below are gathered from all my references used in this paper.

1992

	Internet	535,000 hosts over 30 countries
	Archie	Search index of over 1 million documents
November	Veronica	4000 queries a day

1993

July	Internet	1,3 million internet hosts
------	----------	----------------------------

1994

	WWWorm	Web index of 110.000 pages
	WWWorm	1500 queries per day
	Webcrawler	50,000 documents indexed
	Webcrawler	6000 queries a day
April	Yahoo!	1 million hits a day
June	Internet	2,2 million internet hosts
July	Lycos	54,000 indexed documents
August	Lycos	349,000 indexed documents

1995

January	Lycos	1,5 million indexed pages
February	Veronica	searches 6500 gopher servers a month indexed 15 million items
June	Internet	4,9 million internet hosts
September	Excite	1,3 million indexed pages
November	Veronica	50,000 queries a day
December	Webcrawler	30,000 search requests a day
December	Altavista	Spiders crawled 2,5 million pages a day
December	Internet	Consisted of 30 million pages

1996

	AOL	5,000,000 members
June	Internet	9,5 million internet hosts
November	Lycos	60 million indexed documents

1997

	Google	servers were capable of index 600.000 pages per minute
February	Internet	±100 million pages
April	Go.com	Implements 500 radio stations worldwide
November	AltaVista	20 million queries a day
November	WebCrawler	Web index of 2 million pages
November	WebCrawler	indexed 100 million web documents
December	HotBot	54 million indexed sites

1998

	Hotbot	10,000,000 search words in index
April	Google	24 million pages indexed

1999

	AlltheWeb	1,2 million page views a day
--	-----------	------------------------------

2000

	Daum	100 million page views a day
--	------	------------------------------

2001

July	Google	Image search launched and contained 250 million images
August	AOL	Servers could cope with 10 billion queries a day
August	Inktomi	Servers could cope with 80 Million queries a day

2002

	Google	available in 72 languages
May	AlltheWeb	2,1 Billion searched website in the index
May	Google	2,07 Billion searched website in the index
October	Dogpile	4.3 million unique visitors a day

2004

January	Google	6 Billion indexed items, 4,28 billion pages, 880 million images
November	Google	8 million indexed web pages

2005

Internet	±45 Billion web pages public available
Internet	±5 Billion web pages private available
Internet	±200 Billion web pages database-driven
Yahoo!	345 unique visitors a month
Google	8,1 Billion searched website in the index
Yahoo!	6,6 Billion searched website in the index
Ask	5,3 Billion searched website in the index
Live	5,1 Billion searched website in the index
Internet	11,5 billion known indexed pages to search online

2007

August	Google	37,094 million searches conducted a month
August	Yahoo!	8,549 million searches conducted a month
August	Baidu	3,253 million searches conducted a month
August	Live Search	2,166 million searches conducted a month
August	Naver	2,044 million searches conducted a month
August	AOL	1,212 million searches conducted a month
August	Ask.com	743 million searches conducted a month
August	Lycos	441 million searches conducted a month
August	Internet	±550 million searches conducted every day

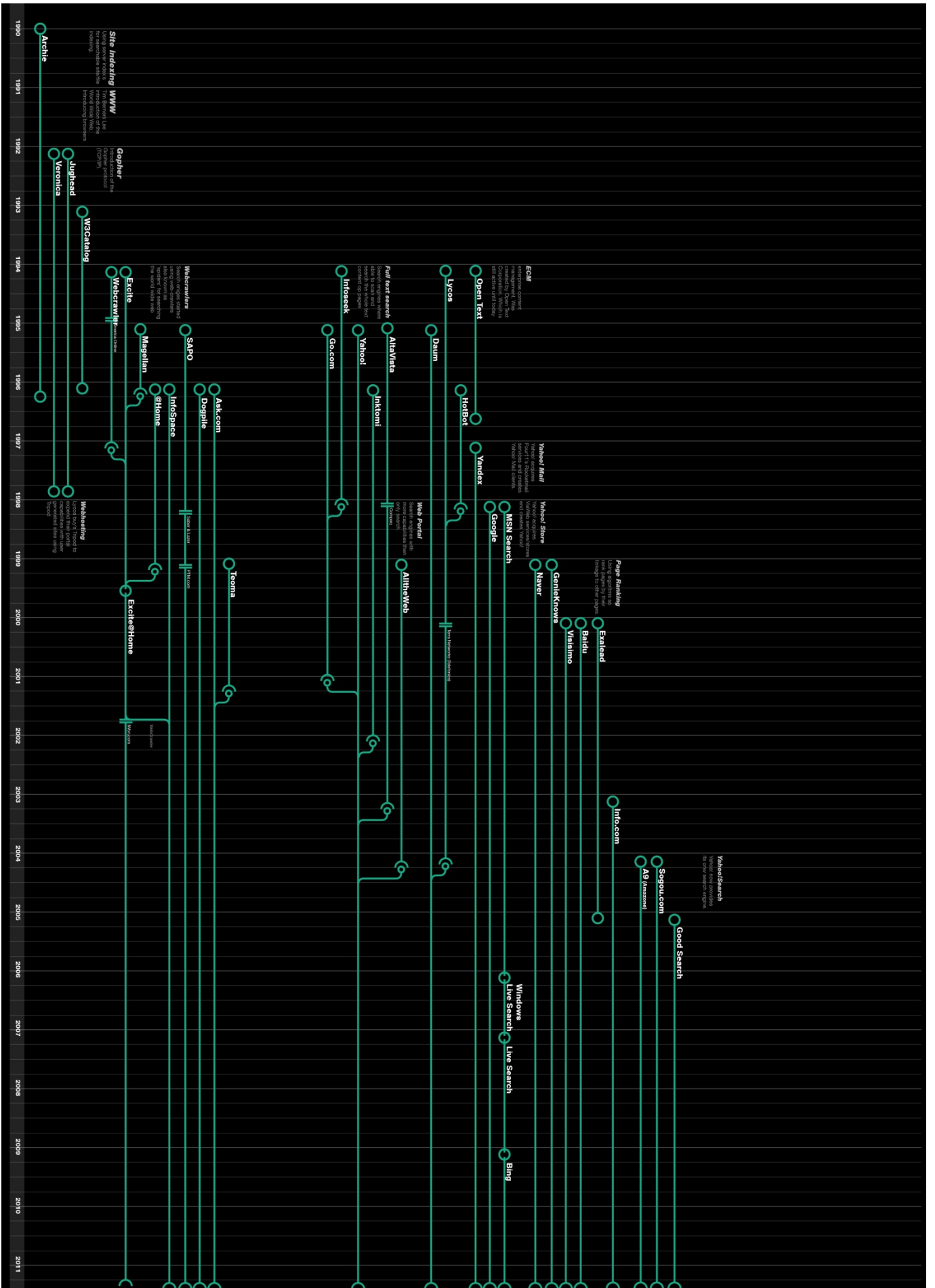
2011

	Lycos	12-15 million monthly users
	Ask.com	90 million monthly users
	Baidu	Tens of Billions of queries
	Daum	38 million subscribers/830 million page views a day
April	Naver	130 Million search queries per day
May	Google	19,26 Billion Core searches monthly
May	Yahoo!	12,20 Billion Core searches monthly
May	Microsoft	3,78 Billion Core searches monthly
May	Ask.com	502 million Core searches monthly
May	AOL	254 million Core searches monthly

Appendix C – Visuals and Graphs

Introduction

Sometimes it becomes hard to cope and understand certain changes and developments that I discuss in this paper. For the purpose of creating a better understanding of the innovations and development of internet search engines I made some graphical charts and visuals to help during my research.



References

- Alexa. *Google.com*. Alexa.com, 2011. 25 June 2011. <http://www.alex.com/siteinfo/google.com#>
- AOL. *About AOL: Overview*. AOL Inc., 2011. 22 June 2011. <http://corp.aol.com/about-aol/overview>
- Ask.com. About Ask. Us.ask.com, 2011. 23 June 2011. <http://us.ask.com/about>
- Ask.com. Fact Sheet. Sp.ask.com, 2011. 23 June 2011. http://sp.ask.com/en/docs/about/fact_sheet.shtml
- Atwood, Brett. *Go.com provides radio outlets Internet presence with web site*. Billboard, 19 April 1997, Vol. 109, Issue 16.
- Baidu, Inc.. *The Baidu Story*. Ir.baidu.com, 2011. <http://ir.baidu.com/phoenix.zhtml?c=188488&p=irol-homeprofile>
- Barlow, John Perry. *A Declaration of the Independence of Cyberspace*. 1996. 30 May 2011. <http://homes.eff.org/~barlow/Declaration-Final.html>
- Becker, Konrad and Stalder, Feliz et al. *Deep Search: The Politics of Search beyond Google*. Studien Verlag, Innsbruck, 2009.
- Berners-Lee, Tim et al. *World-Wide Web: Information Universe*. Electronic : Research, Applications and Policy, April 1992.
- Bonfils, Michael. *Search Marketing Guide to Naver, Korea's Most Popular Search Engine*. Searchenginewatch.com, 11 May 2011. <http://searchenginewatch.com/article/2070244/Search-Marketing-Guide-to-Naver-Koreas-Most-Popular-Search-Engine>
- Bowmen, Lisa M. *The little engine that could beat Google*. Cnet.com, 17 June 2002. 26 June 2011. <http://news.cnet.com/2100-1023-936757.html>
- Bowen, Ted. *Eric Brewer*. InfoWorld, 10 September 2000, Vol. 22 Issue 41.
- Brewer, Eric A. Algorithms in the real world. Carleton Miyamoto and Guy Blellock, 1998. <http://www.cs.cmu.edu/~guyb/real-world/class-notes/all/25.ps>
- Brewer, Eric A. *The Rise and Fall of the Inktomi Corporation*. Kape at Teknolohiya 2010: Wireless and Rural Connectivity, Technologies and Start-Ups. 8 January 2010 at UP-AyalaLand TechnoHub. <http://www.enterprise.upd.edu.ph/?p=292>
- Brewer, Eric A. *Lessons from Giant-Scale Services*. IEEE Internet Computing, July/August 2011.
- Brin, Sergey and Page, Lawrence. *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. In: Seventh International World-Wide Web Conference (WWW 1998), 14-18 April, 1998, Brisbane, Australia.
- Business Wire. *Ybrant Digital Buys Lycos for \$36 Million*. Business Wire, 16 August 201. 20 June 2011. <http://info.lycos.com/overview.php>
- Castells, Manuel. '4. *Lessons from the history of the Internet*', in: *The Internet Galaxy*. Oxford: 2001, 9-35
- CNN Money. *Disney absorbs InfoSeek*. 12 July 1999. 21 June 2011. <http://money.cnn.com/1999/07/12/deals/disney/>

- comScore.com. comScore Releases February 2011 Search Engine Rankings. 10 June 2011. 21 June 2011.
http://www.mometrics.net/dut/Press_Events/Press_Releases/2011/6/comScore_Releases_May_2011_U.S._Search_Engine_Rankings
- Daum. *Corporate History*. 21 June 2011. <http://info.daum.net/DaumEng/info/corporateHistory.do>
- Digital Equipment Corporation. Original press release for Altavista. Digital Equipment Corporation, 15 December 1995. 20 June 2011. <http://www.samizdat.com/altavist.html>
- Dogpile. About Dogpile. Dogpile, 2011. 22 June 2011. http://www.dogpile.com/dogpile/ws/about/_iceUrlFlag=11?_IceUrl=true
- Dogpile. *Different Engines, Different Results: Web Searchers Not Always Finding What They're Looking for Online*. Dogpile.com, In Collaboration with Researchers from Queensland University of Technology and the Pennsylvania State University, April 2007.
- Du Bois, Grant. *Disney's Go.com site goes no more*. eWeek, February 5 2001. Vol. 18, Issue 5.
- Emtage, A., Deutsch, P., *Archie - An Electronic Directory Service for the Internet*. Winter Usenix Conference Proceedings, 1992. Pages 93-110. <http://www1.chapman.edu/gopher-data/archives/Internet%20Information/whatis.archie>
- Foley, Mary J. *Microsoft servers Live Search from the rest of the Windows Live family*. ZDnet.com, 21 March 2007. 25 June 2011. <http://www.zdnet.com/blog/microsoft/microsoft-severs-live-search-from-the-rest-of-the-windows-live-family/339>
- Gerasoulin , Apostolos. *DiscoWeb: Applying Link Analysis to Web Search*. Department of Computer Science, Rutgers University, 1999, New Jersey.
- Gerasoulin , Apostolos. *Apostolos Gerasoulis, Professor.*: Rutgers.edu, 15 August 2010. 27 June 2011. <http://www.cs.rutgers.edu/~gerasoul/>
- Google. *Yahoo! Selects Google as its Default Search Engine Provider*. Google, 26 June 2000. http://www.google.com/googlefriends/alert2_2000.html
- Google.com. *Google History*. Google, 2011. 25 June 2011. <http://www.google.com/intl/en/about/corporate/company/history.html>
- Google. *Introducing Google buzz*. Google, 9 February 2010. 6 July 2011. <http://googleblog.blogspot.com/2010/02/introducing-google-buzz.html>
- Google. *Introducing the Google + project*. Google, 28 June 2011. <http://googleblog.blogspot.com/2011/06/introducing-google-project-real-life.html>
- Gran, Even. *Faster net search using Norwegian technology*. Ntnu.no, 1 January 1999. 26 June 2011. <http://www.ntnu.no/gemini/1999-01/36.html>
- Gray, Matthew K. *Internet Statistics*. 1996. MIT, Wandex, 11 June 2011 <http://www.mit.edu/~mkgray/net/index.html>
- Hock, Randolph E. Sizing up Hotbot. Online, November/December 1997, Vol. 21, Issue 6.

- Holahan, Catherine. *Will less be more for AOL?*. Businessweek.com, 31 July 2006. 22 June 2011.
http://www.businessweek.com/technology/content/jul2006/tc20060731_168094.htm
- Howe, Walt. *A Brief History of the Internet*. 24 March 2010. 9 June 2011.
<http://www.walthowe.com/navnet/history.html>
- Hu, Jim. *InfoSpace to buy Go2Net to expand content delivery*. CNET News, 26 July 200. 22 June 2011.
<http://news.cnet.com/2100-1023-243697.html>
- InfoSeek. *People are saying*. InfoSeek, June 1996. 21 June 2011.
<http://web.archive.org/web/19970702014347/http://info.infoseek.com/doc/people.html>
- InfoSeek. *Company History*. InfoSeek, June 1996. 21 June 2011.
<http://web.archive.org/web/19970702015939/http://info.infoseek.com/doc/Reference/History.html>
- InfoSpace. Inc. *New Dogpile Meta-Search Engine is Now Twice as Fast*. InfoSpace.com, 2 October 2002. 22 June 2011
<http://investor.infospaceinc.com/phoenix.zhtml?c=119056&p=irol-newsArticle&iD=339871&highlight=>
- InfoSpace. *Our Story*. InfoSpace Inc., 2011. 22 June 2011. <http://www.infospaceinc.com/ourstory/default.aspx>
- Internet World Stats. Internet Usage Statistics. Internet World Stats.com, 31, March 2011. June 5 2011.
<http://www.internetworldstats.com/stats.htm>
- It Times. *Naver - Myth of Internet Age Success*. Koreaittimes.com, 15 August 2007. 27 June 2011.
<http://www.koreaittimes.com/story/3083/naver-myth-internet-age-success>
- Jain, Naveen. *Life Story*. www.naveenjain.com, 2010. 22 June 2011. <http://www.naveenjain.com/naveen-jain-biography/>
- Jin-seo, Cho. *Daum at Another Crossroads*. The Korea Times, June 28 2007.
- Kirsch, Steve. *Infoseek's approach to distributed search*. Report of the Distributed Indexing/Searching Workshop, Cambridge, Massachusetts. 28 May, 1996.
- Kirsch, Steve. *Secure, convenient and efficient system and method of performing trans-internet purchase transactions*. U.S. patent, 5 October 1999.
- Kirsch, Steve. *Real-time document collection search engine with phrase indexing*. U.S. patent, 6 July 1999.
- Koster, Martijn. *Robots Exclusion Standard*. Robotstxt.org, 30 June 1994. 11 June 2011
<http://www.robotstxt.org/orig.html>
- Koster, Martijn. *Robots in the Web: threat or treat?*. Robotstxt.org, April 1995. 11 June 2011
<http://www.robotstxt.org/orig.html>
- Krantz, Michael, et all. *Click till you drop*. Time, Juli 20 1998 Vol. 152 Issue 3.
- Kraus, Joe. *Bnoopy: Personal Blog Joe Kraus*. 2004-2005. 20 June 2011
<http://bnoopy.typepad.com/bnoopy/>

- Kuner, Maurice de. *Geschatte grootte van het geïndexeerde World Wide Web*. University of Tilburg. 27 March 2007.
- Kuner, Maurice de. The size of the World Wide Web. worldwidewebsite.com, 7 July 2011. 7 July 2011.
<http://www.worldwidewebsite.com/index.php?lang=EN>
- Lai, Eric. Microsoft Plays Catch-up In Search Tools Market. ComputerWorld, 22 May 2006.
- Lewis, Peter H. *Digital Equipment Offers Web Browsers Its 'Super Spider'*. The New York Times, 18 December 1995.
- Lipsman, Andrew. *61 Billion Searches Conducted Worldwide in August*. comScore.com, 10 October 2007. 26 June 2011.
http://www.comscore.com/Press_Events/Press_Releases/2007/10/Worldwide_Searches_Reach_61_Billion
- Lycos, Inc. *Company Overview*. Lycos, Inc., 2011. 20 June 2011.
<http://info.lycos.com/overview.php>
- Mauldin, Micheal L. *Curriculum Vitae*. 20 June 2011
<http://robot-club.com/lti/vita.html>
- Mauldin, Micheal L. *Lycos: Design choices in an Internet search service*. Lycos, Inc. January/February 1997. 20 June 2011
<http://robot-club.com/lti/pub/ieeee97.html>
- Metz, Cade. *Innovators Alan Emtage*. PC Magazine, April 24 2007
- MCBryan, Olivier A. *GENVL and WWW: Tools for Taming the Web*. Dept. of Computer Science University of Colorado, to appear in Proceedings of the First International World Wide Web Conference, ed. O. Nierstrasz, CERN, Geneva, May 1994. 20 June 2011.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.11&rep=rep1&type=pdf>
- McGee, Matt. *Bing Search Volume Up 29% in 2010, Google Up 13%, comScore Says*. Searchengineland.com, 8 February 2011. 26 June 2011. <http://searchengineland.com/bing-search-volume-up-29-in-2010-google-up-13-comscore-says-64075>
- Frana, Philip L. Oral History of Mark P. McCahill. National Science Foundation, 13 September 2001.
- Microdoc. *AllTheWeb Now Better Search Experience Than Google*. 25 June 2003. 26 June 2011.
<http://web.archive.org/web/20030704024758/http://www.microdocs-news.info/newsGoogle/2003/06/25.html>
- Nierstrasz, Oscar. *W3 Catalog History*. 8 November 1996. 9 June 2011
<http://scg.unibe.ch/archive/software/w3catalog/W3CatalogHistory.html>
- Notess, Greg R. *Searching the World-Wide Web: Lycos, WebCrawler and more*. Online, Jul/Aug95, Vol. 19 Issue 4.
- Page, Lawrence. *Method for node ranking in linked database*. U.S. Patent, 4 September 2001.
- Perez, Juan C. *Microsoft Spotlights Its Search Engine*. PCWorld.com, 1 February 2005. 25 June 2011.
http://www.pcworld.com/article/119512/microsoft_spotlights_its_search_engine.html

- Perez, Juan C. *Microsoft Revamps MSN Search*. PCWorld.com, 1 July 2005. 25 June 2011.
http://www.pcworld.com/article/116766/microsoft_revamps_msn_search.html
- Pinkerton, Brian. *Finding What People Want: Experiences with the WebCrawler*. www.thinkpink.com, 1994. 12 June 2011
<http://www.thinkpink.com/bp/WebCrawler/WWW94.html>
- Polly, Jean Armour & Cisler, Steve. *Travels with VERONICA*. Library Journal, 1 January 1995, Vol. 120 Issue 1, p32.
- Price, Gary. *Google's Usenet Timeline and Early Search Engine Announcements*. Searchenginewatch.com, 10 January 2005. 11 June 2011. <http://searchenginewatch.com/article/2063066/Googles-Usenet-Timeline-and-Early-Search-Engine-Announcements>
- Selberg, Erik & Etzioni, Oren. *The MetaCrawler Architecture for Resource Aggregation on the Web*. Department of Computer Science and Engineering University of Washington, 8 November 1996.
- Rankdex. About Rankdex. Rankdex.com, 2011. 27 June 2011. <http://www.rankdex.com/about.html>
- Rao, Leena. *The Sun Will Set For Yahoo's AlltheWeb on April 4*. TechCrunch.com, 18 March 2011. 26 June 2011.
<http://techcrunch.com/2011/03/18/the-sun-will-set-for-yahoos-alltheweb-on-april-4/>
- Red Herring. *Smarter than Bill*. RedHerring.com, 30 June 1997. 22 June 2011.
- Resonance China. Q4 2010: China Search Engine Revenue Market Share. Resonancechina.com, 31 January 2011. 27 June 2011.
<http://www.resonancechina.com/2011/01/31/q4-2010-china-search-engine-revenue-market-share/>
- Resume Bucket. *David Hayden Profile*. ResumeBucket.com. 20 June 2011.
<http://www.resumebucket.com/davidhayden>
- Rohrlich, Justin. *Stupid Business Decisions: Excite Rejects Google's Asking Price*. Minyanville.com, 23 April 2010. 20 June 2011
<http://www.minyanville.com/special-features/articles/excite-google-microsoft-yahoo-apple-bankruptcy/4/23/2010/id/27013>
- Roush, Wade. *Search Beyond Google*. Technology Review, 2004.
- Salient Marketing. *1993 - Veronica, the grandmother - or Archie's girlfriend*. 9 June 2011.
<http://www.salientmarketing.com/seo-resources/search-engine-history/grandmother-search-engine.html>
- Sang-Hun, Choe. *South Koreans Connect Through Search Engine*. The New York Times, 5 July 2007. 27 June 2011.
http://www.nytimes.com/2007/07/05/technology/05online.html?_r=1&oref=slogin
- Schneider, Stefan. *Asian Search Engine Market Shares*. Online Marketing in China, February 13 2010. 21 June 2011
<http://www.my-life-in-china.com/online-marketing/asian-search-engine-market-shares-20092010/>
- Sherman, Chris. *Happy Birthday, Excite!*. Search Engine Watch.com, 1 October 2002. 20 June 2011
<http://searchenginewatch.com/article/2064708/Happy-Birthday-Excite>
- Sherman, Chris. *Happy Birthday, Aliweb!*. Search Engine Watch.com, 3 December 2002. 11 June 2011
<http://searchenginewatch.com/article/2065671/Happy-Birthday-Aliweb>

- Sherman, Chris. *Happy Birthday, Ask Jeeves*. SearchEngineWatch.com, 7 April 2003. 25 June 2011.
<http://searchenginewatch.com/article/2063934/Happy-Birthday-Ask-Jeeves>
- Singel, Ryan. *Hands On With Microsoft's New Search Engine: Bing, But no Boom*. Wired.com, 28 May 2009. 25 June 2011.
<http://www.wired.com/epicenter/2009/05/microsofts-bing-hides-its-best-features/>
- Sonnenreich, Wes. *A History of Search Engines*. 1997. 9 June 2011.
<http://www.wiley.com/legacy/compbooks/sonnenreich/history.html>
- Spencer, Graham. *Excite: internet navigation service-discussion*. Usenet post, 27 September 1995. 20 June 2011
<http://groups.google.com/group/comp.infosystems.www.announce/msg/474272ef391d9730>
- Stross, Randall E.. *How Yahoo! Won the Search Wars*. Fortune, 3 February 1998, Vol. 137, Issue 4.
- The Herring Reporter. *Inside Architext*. Red Herring, 1995, March issue.
- University of Stirling. *The World's First Web Search Engine*. Spring 2009. 11 June 2011.
<http://www.stir.ac.uk/about/jumpstation>
- USA Today. *How the Internet took over*. USA Today, 30 April 2007. <http://www.usatoday.com/tech/top25-internet.htm>
- Wired. *Lycos Acquires Wired Digital*. Wired.com, 6 October 1998. 23 June 2011.
<http://www.wired.com/techbiz/media/news/1998/10/15437>
- Wired. *A fine IPO for InfoSpace*. Wires.com, 15 December 1998.
- Yahoo! Inc.. *History of Yahoo! - How It All Started...* Yahoo! Inc., 2005. 21 June 2011.
<http://docs.yahoo.com/info/misc/history.html>
- Yahoo! Inc.. *Yahoo! Inc. Completes Initial Public Offering Of 2,600,000 Shares Of Common Stock At \$13.00 Per Share*. Yahoo! Inc., 1997. 21 June 2011. <http://yhoo.client.shareholder.com/releasedetail.cfm?releaseid=173429>
- Yandex Compagny. *History of Yandex*. Campagny.yandex.com, 2011. 25 June 2011.
http://company.yandex.com/general_info/history.xml
- Zakon, Robert H. *Hobbes' Internet Timeline 1993-2010*. 15 December 2010. 9 June 2011.
<http://www.zakon.org/robert/internet/timeline/>