University of Fribourg
Faculty of Economics & Social Sciences
Information Systems Research Group

Meta-Search Engine Analysis

Seminar Thesis

Supervisor: Prof. Dr. Andreas Meier
Advisor: Edy Portmann

Kim-An Phan
08-215-469
Grenchenstrasse 58
4500 Solothurn
kim-an.phan@unifr.ch

September 2010

Abstract

This paper gives an overview of web information retrieval and explains the function and use of search engines. The world's major search engines Google, Yahoo, and Bing were compared based on the factors database size, actuality, capability, and technology. An analysis explains the advantages and disadvantages as well as the differences between general search engines and meta-search engines.

Furthermore, an answer to the question if merging search results are more relevant than results from one general search engine can be found. The paper concludes with the discussion on what is required for a search engine to improve its performance.

Keywords: Information retrieval, web search engines, meta-search engines

Table of Contents

Figures

Tables

1. Introduction

1.1 Problem Definition

The importance of search engines has grown through the years. They give people the opportunity to find information in an easy and quick way and have become a part of people's daily life.
Throughout this paper, the following research questions will be further investigated:
Why are search engines successful? What are the main differences of the most successful search engines? What are meta-search engines and how do they work? Can meta-search engines optimize the search query? And finally, which requirements need to be considered by an ideal search engine?

1.2 Objectives

One of the objectives of this paper is to give an overview of information retrieval and show how different types of search tools work. A closer look at web search engines will be taken in order to make a comparison of the major search engines. Afterwards, meta-search engines will be introduced and described to understand their function. The main goal is to examine if the hypothesis that multiple search engines can outclass single search engines and optimize the query can be verified. Finally, the question if there is such thing like the best search engine will be up for discussion.
The seminar thesis puts great emphasis on functions of search tools. However, the business aspect will, where it's important, also be regarded.

1.3 Methodical Approach

This paper, which is constructed in two different parts, will be based on several literature researches.
The first part gives an insight of theory that describes information retrieval in general as well as the characterizations and functions of search engines. There will also be explained why search engines are successful.
In the second part, a qualitative content analysis of the three most popular search engines should be made in order to expose what their main strengths and weaknesses are. They should be compared to meta-search engines in order to find out which category is more useful.

2. Information Retrieval

There are many methods for finding information, but one of the leading ways is through search engines. At this time, practically everyone uses search engines, mostly for research, school, business, shopping, or entertainment. As the biggest driver of traffic on the Web, they have a grand influence which is continually growing [Clay & Esparza, 2009].
To know where search engines come from and how they work, it's important to have an overview and clear understanding of Information Retrieval.

Information retrieval (IR) is „the process of searching within a document collection for a particular information need" which is called a query [Langville & Meyer, 2006].
Baeza-Yates and Ribeiro-Neto [1999] indicate that information retrieval deals with the „representation, storage, organization of, and access to information item", in order to give the user the possibility to easily access the desired information.

A distinction between traditional information retrieval and web information retrieval can be made:
Traditional or classic information retrieval is search in smaller, controlled collections that are not linked [Langville & Meyer, 2006]. These document collections are stored in physical form. An example therefore would be looking for information in books of a public library.
Nevertheless, nowadays, most of the documents are computerized that can be retrieved with the help of basic computer-aided techniques, also referred to as information retrieval models or methods.

Web information retrieval is, other than the traditional IR, search within the globally largest collection of documents that are linked, such as the well known search services on the internet like Google or Yahoo [Langville & Meyer, 2006]. In the next chapters, web information retrieval, or more specifically, web search services will be indicated.

3. Web Search Services

Web search is often preferred over other information sources. The findings of an internet survey through Pew Internet, for instance, show that the web is for 92% of users a good place for getting information every day [Manning, Raghavan, & Schütze, 2009].

There are a couple of factors why web search is successful today. One reason might be its convenience. Nowadays, web search tools allow information to be easily accessible, anywhere and anytime, and they are available to anyone who has internet. Most of the people, however, don't realize how much they take search engines for granted. Imagine there was no web search service, what would it mean for the way people work online? Everyone would probably agree that the products and services from search engines make the use of the web much easier, more time-saving, if not even more efficient.

Since most users discover websites through search services, they also have a high priority for webmasters and web designers. In order to reach the desired audience, webmasters strive to create good, effective, and well-known websites. But with the help of search engines, more people will be able to find their site or discover that it actually exists. Webmasters can take many advantages of the web, especially for business purposes. A lot of effort will be put into search engine optimization (SEO) or maximizing search engine visibility, online marketing strategies [Clay & Esparza, 2009]. In other words, on condition that a website is build search engine friendly, the traffic of a website potentially increases. A study, conducted by a research organization and constituted by Thurow [2003], found that, after finding a website through a search engine, consumers are five times more likely to purchase a product or service rather than through a banner advertisement. Thurow [2003] points out, that it can be cost-effective to maximize a site's search engine visibility and „a properly performed search engine marketing campaign can provide a tremendous, long-term return on investment".

Because of the search engines significance, it may be advantageous or even necessary to know how search services work and what their background is.
Basically, there are two different methods for search tools: Directories and search engines, as per statement in the two upcoming subchapters.

3.1 Search Engines

First off, it's crucial to know that when a person performs a web search, he's actually not searching the web but the search engine's index of the web. Due to speed, costs, and capabilities, it is plain not possible to search through all the web pages every time a user clicks 'search' on an engine [Sherman & Price, 2001].

A general search query procedure can be summarized in four steps:

(1) A web user submits a query by typing a term, words or phrases in the search box.
(2) Regarding the query, search engine looks through all the pages that it keeps in its database.
(3) Search engine sorts out the relevant web pages
(4) Results are listed on the Search Engine Results Page (SERP) in an order, beginning with the most relevant results.

The whole search process usually only lasts a fraction of a second, but what's behind a search engine's function is more complex as it seems.

Web search engines consist of three basic parts: Web crawler, indexer, and query processor. The components and tasks of web search engines, which are illustrated in Figure 3.1-b, will be described on the next site.

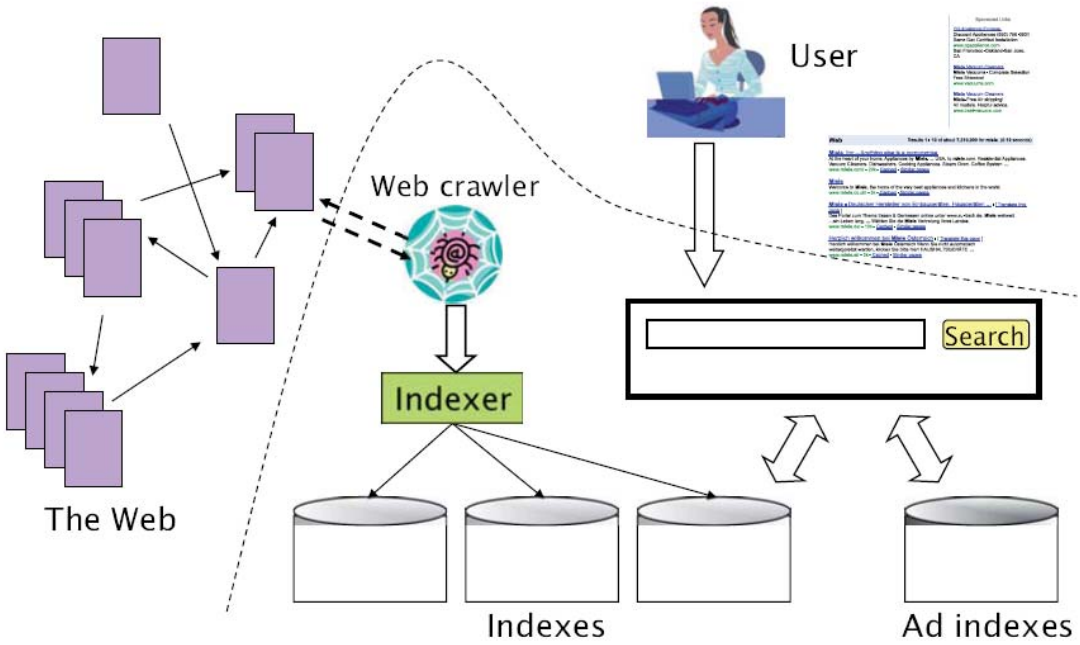

Figure 3.1-a Components of a Web Search Engine
Source [Manning, Raghavan, & Schütze, 2009, p. 434]

Crawling or spidering is an automated process to gather the data with web spiders. They can be visualized as little spiders and are also known as crawlers, robots, software agents, web agents, wanderers, walkers, or knowbots [Clay & Esparza, 2009]. Named after those special software robots, this type of search service is called “spider-based” or “crawler-based” search engine.
Spiders continuously crawl web pages by fetching them and build lists of words and phrases found to keep them as a full-text index in a database of the search engine. They find pages either through the URL, which web authors add to a list to notify of their web page’s existence, or through hypertext links embedded in most web pages [Sherman & Price, 2001]. In the latter case, spiders start by crawling a few web pages and follow the links on those pages. After fetching the pages they point to, they follow the links that are on the last pages. The same process will be continued until they have indexed a certain part of the web that includes pages they store across many machines, what leads to the next task.

Indexing is the second part of search engines. It is the process of “taking the raw data and categorizing it, removing duplicate information, and generally organizing it all into an accessible structure” [Clay & Esparza, 2009].

The stored full-text indexes of the crawled web pages are organized in a database, typically in an inverted index data structure [Sherman & Price, 2001]. It is ideal for keyword based queries, so that documents that include the typed keywords can be quickly retrieved.

Figure 3.1-a shows such an inverted index data structure which is sorted in an alphabetical order. In this example, there are four phrases with words to which numbers should be assigned. The first number is the identifier for each phrase (Doc #), hence, in this case numbers from one through four. The second number represents the position of the word within the phrase it occurs. Common words like “and”, “is”, “the” or “you” are discarded by some search engines. It would make no sense to contain those so called stop words, because they would only reduce search performance, since they are very ordinary.

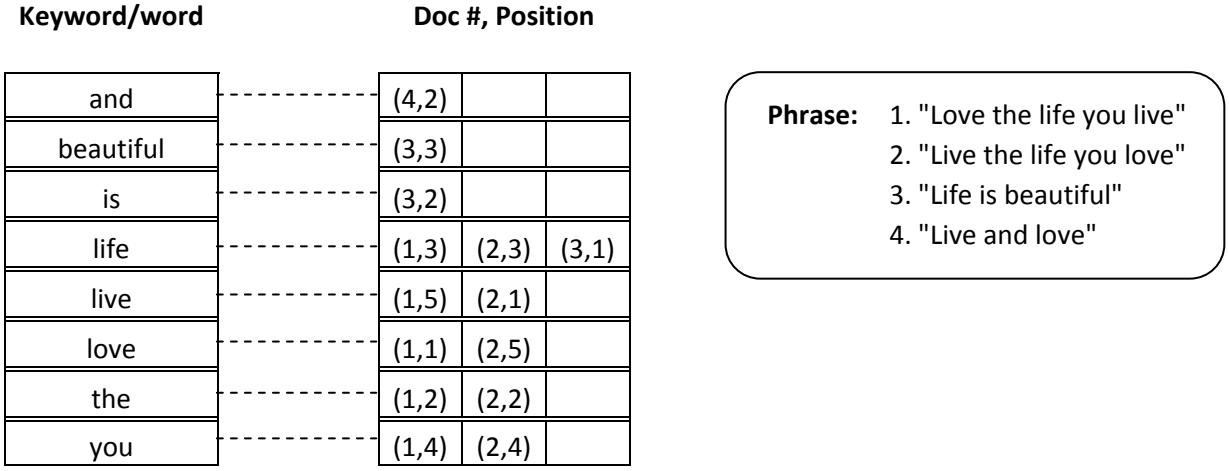

Figure 3.1-b Inverted Index Data Stucture
Source cf. [Sherman & Price, 2001, p. 20]

Technical and economical factors make it difficult to index the whole web. Not only technical limitations, but also cost restrictions don’t allow search engines to crawl the whole existing web [Lewandowski, 2005a].
The third and last part is called query processor which consists of the search form, the matching of the search query with relevant documents in the database, and the results-output formatter which is the search results page [Sherman & Price, 2001].

The search form on the user interface is basically the search box where the query can be typed in. Basic and advanced search forms are usually provided by several engines.

To find relevant documents that have been indexed for a particular query, search engines use special techniques. “The major differentiator of one search engine from another lies in the way relevance is calculated” [Sherman & Price, 2001]. Each search engine applies an algorithm that weighs various criteria and generates a result to decide which listings to display in the results form and in what order [Clay & Esparza, 2009]. Ranking algorithms are primarily math equations and very important to achieve search engine optimization.

Search engines rank results by using query-dependent factors, also on-the-page criteria, and query-independent factors, also called off-the page criteria.
Query-dependent factors are ranking methods to measure how good a page matches a specific given query. They include the measures in traditional Information Retrieval, such as word documents frequency, or language of the document and the query or the geographical distance [Lewandowski, 2005a].
Query-independent factors, on the contrary, attempt to determine the quality of a document, regardless of the given query. They are usually based on link analysis. PageRank for example belongs to the most famous query-independent factors. It is a measure of link popularity and will be explained more in detail later in section 4.2.4, referring to the search engine Google.

Many web designers try to manipulate the rankings in order to boost their rank position in the results form of search engines. Consequently, to prevent this manipulation, unwanted methods like spamming for improving the ranking, algorithms of search engines are highly confidential and change almost daily [Thurow, 2003].

Search results in a results page may be classified by two types: Organic search results (“natural” search results) that include web pages found through spiders, and paid results (sponsored listings) like advertisements based on keywords for which webmasters pay, so their web page will be placed above or to the right of the primary results [Clay & Esparza, 2009]. Paid results are always clearly identified with a designation.
Concerning advertisement on search results page, along with the page optimizing to ease the spider’s access to the web page, website owners have two ways to reach the users. Pay-for-inclusion model guarantees that the submitted web pages will not be dropped from a search index and new information added will be reflected very quickly. However, it does not assure that they will appear in top positions.

Pay-for-placement, on the contrary, ensures web pages a position on the paid results mentioned above, in favor of a bid between owners for particular keywords. Search engines that support this offer can be called as “pay-per-click” search engines, which means that an advertiser will be charged based on the number of times users click on a link to his web site through the search engine’s results [Thurow, 2003].
Fact is that people tend to access organic results more often than on the ads, but it is not possible to pay for a good position on that list. Nobody can buy his way to the top of organic results. The only way to earn a place in the top search results is with the aid of effective search engine optimization [Clay & Esparza, 2009].

3.2 Web Directories

Web directories, also catalogs, yellow pages, or subject directories, provide a context-based framework for structured browsing. Sherman and Price [2001] compare them to a table of contents in a book, because they use a hierarchical structure, just like a table of contents, to provide a high level overview of major topics, while search engines are more like an index of a book. Another comparison made by Sherman and Price [2001] is that search engines are akin to telephone white pages with a name and address list, whereas directories, yellow pages, respectively are organized by category and provide descriptive information.
Unlike search engines, which use autonomous software agents, directories manually place web sites and pages into specific categories with the help of human editors, why they’re also called “human-based” search engine. The way how the data is arranged is the biggest difference between an index and a directory, as Bruce Clay points out [2009].

Human editors evaluate and select by searching or browsing the web from site to site to decide whether the site or page is valuable enough to be added to their directory. A listing of classified topics will be created with links of web pages that are categorized in a hierarchical structure to simplify the query from a user’s perspective [Thurow, 2003].

Directories consist only of links arranged by subject and annotations. Since the links are hand-selected, directories are small and limited. Thus, their results should be supplemented with search engine partners’ results, so called “fall-through” or “fall-over” results that they display differently from the general directory listings (organic results) [Sherman & Price, 2001]. The other way around, some search engines pull information from directories.

Similar to search engines, directories support paid submission programs, as well as rank their web sites. Top directory listings are based on the directory category and the web site's title and description. Editors evaluate web sites by means of the websites unique content with good quality and how it's presented. Provided that all the predetermined conditions are met, the site will be added to the directory.

3.3 Meta-Search Engines

Meta-search engines, also known as multiple search engines, metasearchers, or metacrawlers, are special search tools that present the results by accessing multiple search engines and web directories. This way, they allow users to quickly receive combined results that are merged in one place at once. Thus, web users neither need to type the query several times nor have to access every single search engine by themselves. This job will be done for the users by meta-search engines, which might additionally suggest engines that the user had not considered before.

By performing a search query, meta-search engines transmit the typed terms simultaneously to multiple individual search engines. Multi-search engines don't do the crawling or maintain their own database like single search engines, but usually filter the results they found instead. Based on a specific algorithm, they eliminate duplicates and rank the results from their sources into a list. The list of collection will be displayed on the SERP, very similar to the search engines' results page, that relies on the indices of other search engines [Sherman & Price, 2001; Clay & Esparza, 2009].
There are also some meta-search engines that don't use an algorithm, but presents the resulted information of the sources. Meta-search engines differ from each other in the selection and quantity of search engines and in the presentation of results [Mohamed, 2004].

4. Analysis of Search Engines

4.1 Major Search Engines

When people look for information, they generally have at least one favorite search engine that they use regularly to satisfy their search needs. According to About.com, most web searchers expect three key features, namely relevancy results, uncluttered, easy to read interface and helpful options to tighten or broaden a search [Gil, 2010]. Hence, major search engines are generally popular thanks to those factors. They provide results that are both well-maintained and upgraded.

As far as web designers are concerned, major search engines as exceedingly relevant, because they want their site to be known and therefore to be listed in a place where they can generate a lot of traffic. Thus, those engines are most appropriate for SEO strategies.

Google is most likely recognized worldwide as the largest search engine. A global search survey conducted by ComScore, a leader in measuring digital world, proves that this statement is acknowledged: In 2009, Google dominated 66.8% of worldwide search with 87,809 searches, followed by Yahoo! with 9,444 searches, the Chinese search engine Baidu with 8,534 searches, and Bing that ranked fourth with 4,094 searches [comScore, 2010].

Table 4.1 shows the total of searches worldwide from the years 2008 and 2009 by people at the age of fifteen or more. The numbers are based on expanded search definition, which is the reason why not only search engines, but also other top properties with activity in search are contained. As it can be observed, Microsoft sites became with 70% change more popular during 2009, compared to the preceded engines. The most progress was made by the Russian search engine Yandex with 91%. Yet, it is not known worldwide at the moment.

Top 10 Search Properties by Searches Conducted December 2009 vs. December 2008 Total Worldwide, Age 15+ - Home & Work Locations Source: comScore qSearch			
	Searches (MM)		
	Dec-2008	Dec-2009	Percent Change
Worldwide	*89,708*	*131,354*	*46%*
Google Sites	55,638	87,809	58%
Yahoo! Sites	8,389	9,444	13%
Baidu.com Inc.	7,963	8,534	7%
Microsoft Sites	2,403	4,094	70%
eBay	1,327	2,102	58%
NHN Corporation	1,892	2,069	9%
Yandex	992	1,892	91%
Facebook.com	1,023	1,572	54%
Ask Network	1,053	1,507	43%
Alibaba.com Corporation	1,118	1,102	-1%

Table 4.1 Top Properties by Searches Conducted

Source [comScore, 2010]

Studies of Hitslink by Net Application show in its market share rankings of search engines that the positions of the last two years still remain in August 2010. Google ranks first again with 84.73% market share, and outperforms Yahoo! (6.35%), Baidu (3.31%), and Bing (3.30%), while the other engines only capture a total of 1.32%. Baidu recently outpaced Bing, namely from July to August 2010 with 1% [Net Applications, 2010].

Figure 5.1-a illustrates the total market share of search engines in August 2010.

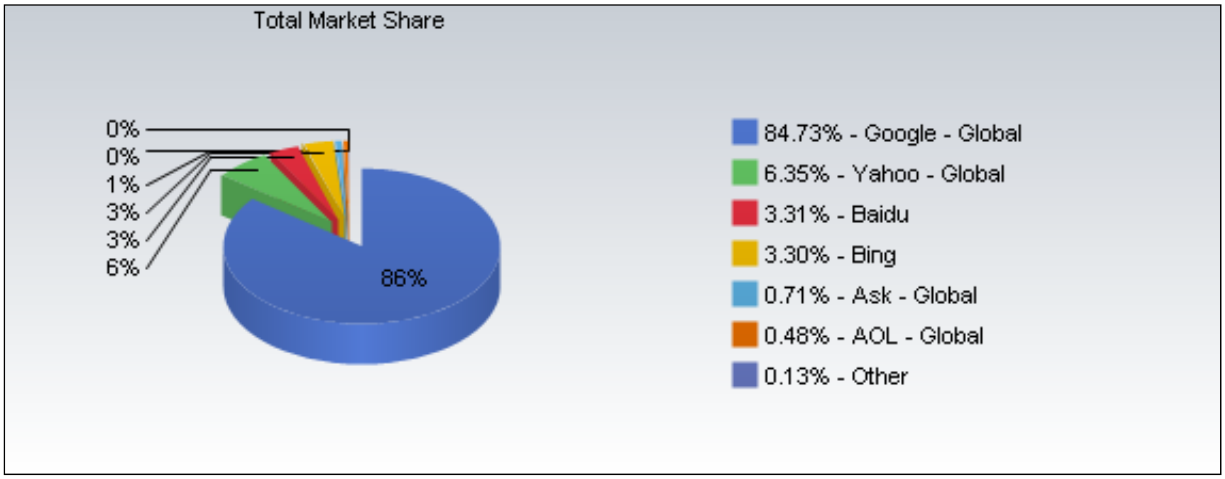

Figure 4.1-a Search Engine Market Share August, 2010

Source [Net Applications, 2010]

Webdevelopersnotes.com is currently making an online survey with a total of 13,304 participants in September 2010, voting on what search engine they think is the best in the world. The majority of the voters estimated Google as the best search engine, while only one fourth of the total voters chose either Yahoo, Bing, AOL, or Ask.

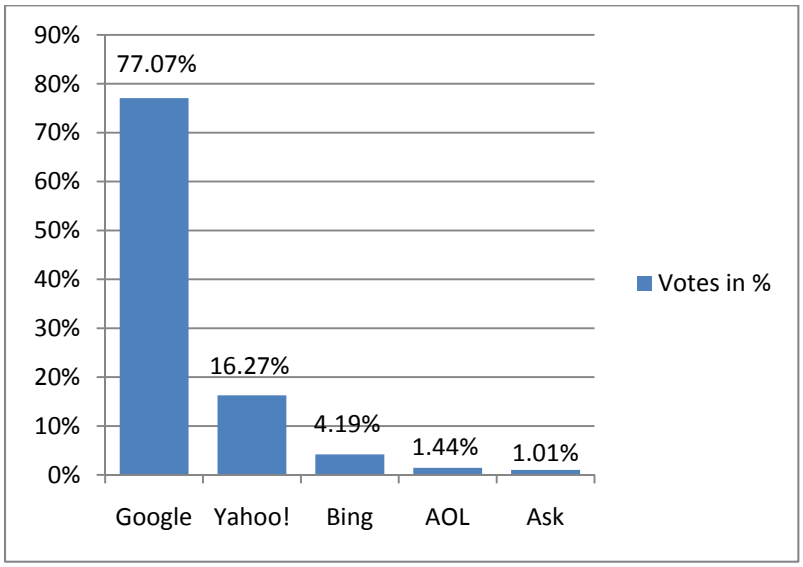

Figure 4.1-b Survey for the best Search Engine

Source cf. [Web Developers Notes, 2010]

Although the search engine Baidu, which is in fact the most used search engine in China, has taken the third rank, it is only available in the Chinese version and not widespread globally at the moment. Consequently, in this paper, it should not be examined more in detail.

In the following section, the three top search engines should be shortly introduced that are known worldwide.

4.1.1 Google

In 1998, Google was set up by Larry Page and Sergey Brin from Stanford University. Its name is based on the word “googol” for the number 10^{100} and symbolizes the huge information volume available on the web and at the same time its mission “to organize the world’s information and make it universally accessible and useful” [Google Inc., 2010].

Google, a spider-based search engine, is considered to be the most popular search engine which was affirmed in the previous chapter.

Google hosts and develops a number of web search features, as well as additional services and tools, known as Google products (e.g. Google Maps, Google Earth).

4.1.2 Yahoo!

Yahoo, founded in 1994 by Jerry Yang and David Filo, is an acronym for "Yet Another Hierarchical Officious Oracle". Its vision is "to be the center of people's online lives by delivering personally relevant, meaningful Internet experiences" [Yahoo, 2010b].

Yahoo is the web's oldest directory, though it started with crawler-based search results powered by Google. Since 2004, an own search technology is being used, with Yahoo's index and ranking mechanism [SearchEngineWatch, 2004]. Yahoo has now both, an own crawler-based index (Yahoo! Search) and a human-edited directory (Yahoo! Directory) with sites in subject categories, evaluated by editors from Yahoo.

4.1.3 Bing

Microsoft's search engine, formerly MSN Search, Windows Live Search, and Live Search, is known by the name Bing since June 2009.
Bing as well indexes by crawling the web. Microsoft acquired PowerSet, a search company which allows improvements in general searches with the addition of related searches [Microsoft, 2010; The Register, 2009]. With a user-friendly organization of search results, Bing also puts effort into a great visual presentation with vibrant pictures.
As can be seen in Table 4.1 of the previous chapter, Bing's implementation and its new technology made a contribution to Microsoft's 70% increase in search.

4.2 Comparison of Google, Yahoo, Bing

After the previous introduction to the three search engines, they will be examined in detail and compared on the basis of selected factors. In the attached Table 4.2, all the points of comparison will be summed up at a glance.

There are some factors thanks to which several search engines stand out. While many points can be focused on, in the following research, three main aspects of search engines should be analyzed: The database size, the actuality, the capabilities, and the technology of the results of search engines. These characterizations are thought to be core values for evaluating search engines.

4.2.1 Database Size

Throughout the years, the number of the web has such a tremendous growth that it cannot be counted. The exact number of web pages that is indexed in search engines is not known either. However, there are some estimations of the database size of each search engine.

Lewandowski [2005a] says the only way to find out the size of the web is to evaluate the size that is based on a representative sample. He indicates that Google has indexed around 8 billion documents, whereas estimation of Yahoo's index shows 5-7 billion, and formerly MSN search 4-5 billion documents. Those numbers, however, were from 2005.

A way to estimate which search engine has more web pages indexed, the same word can be typed in every search box and to see how many results each of them deliver. For a roughly actual size of each search engine database, it should additionally be a single word that is extremely common and most likely appears in every document. For example the word "the" can probably be found in every English page. As of September 18th, 2010, Google found approximately 12 billion pages, while Yahoo gives 9 billion, and Bing 0.9 billion results that contained the word "the". It's important to keep in mind that those numbers are only a certain percentage of the indexed documents, because the actual size of the entire index is much bigger. But they may show the ranking of the search engines' index size.

Nevertheless, the size of the database does not tell about the quality of a search engine. Due to duplicates and spam, search engines should not index the entire web [Lewandowski, 2005a]. There is the so called Invisible Web, Deep Web, or Hidden Web, which is below the surface web, with the static web pages normally being crawled. It is defined as information material on the web that general search engines don't add to their indices of web pages, either because their technology is limited in its capabilities or they consciously decided not to do so [Sherman & Price, 2001].

Google, Yahoo, and Bing work on reaching the Deep Web. Google involves the most important documents of the Hidden Web for patent data manually, but it has also developed a technology that allows an automatic approach of resources [Lewandowski, 2005a; Madhavan, Ko, Kot, Ganapathy, Rasmussen, & Halevy, 2008].
Yahoo in contrast, has its Content Acquisition Program in which it includes documents of the Deep Web through partnerships as content providers [Olsen, 2004].

There is no clear information about Bing, but it has also been exploring the Invisible Web, for example in Microsoft Research, a testbed for information extraction from Deep Web was proposed [Yamada, Craswell, Nakatoh, & Hirokawa, 2004].

4.2.2 Actuality

As described in section 3.1, databases consist of document copies that the crawlers found and indexed. But since the web content changes over time, spiders have to re-crawl those pages to get the updated version and keep the index quality. “Crawling is a never-ending process” [Langville & Meyer, 2006]. How well do the major search engines manage the up-to-dateness of their database?

In a research, Lewandowsky, Whalig and Meyer-Bautor [2005] tested the frequency in which the indices are updated by Google, Yahoo and MSN. During forty-two days, they observed four different groups with nine or ten websites, updated every day, to find out whether these search engines are able to index current contents on a daily basis. Googlebot, Google’s web crawling robot updates many sites daily and is the fastest concerning index quality. The amount of time for re-crawling usually depends on the link popularity and on the frequency how often the web page changes [GoogleGuide, 2007]. MSN on the other hand updates the index with MSNbot frequently, while Yahoo seems to update with its crawler Yahoo Slurp in a chaotic way (Lewandowski et al., 2005).

4.2.3 Capabilities

Search engines enable the web searcher to enter some simple keywords for a query, but also functions to narrow the search and receive more precise results, such as the basic and advanced search. However, even though they can contribute to a web searcher’s success in finding his information, they are either not used or even known by many people.

The next paragraph explains the most important operators that belong to the search engine’s query language.
One of the basic options are the Boolean operators AND, OR, NOT that are also used in the classic information retrieval. The operator AND, to join words, is usually an automatic default what means that it is already assumed between typed words (e.g. “question answer” = “question AND answer”). Both OR and NOT should be capitalized, and are commonly used for equivalent or related terms, and excluded words, respectively.
The minus (-) and plus (+) signs can be used to remove or include words and phrases.

Quotation marks (“”) allow the user to look for an exact phrase in a document. This function can be very practical if someone retrieves a known specific part of a text. In general, common words like “a” or “the” are ignored unless they are placed within quotation marks. For filling in the blanks (*), wildcards can be used. They replace any unknown whole word like a placeholder for terms.
Stemming is a technique that helps to search a word on the basis of its root, meaning that the search engine is able to find several results from a word that can have multiple endings. For example for the term “drive” the variations “driving”, “driver”, “drivers” etc. will be returned.
For Meta words search, there are some options that give results in consideration of the special keyword limitation. If the user wants to find search results in a particular website or class of site, he can use (site:). For instance pages about Informatics on the University’s website can be retrieved through the query *informatics:unifr.ch.* Meta words can be, amongst others. hostname (hostname:), link (link:), or URL (url:) and intitle (intitle:) for keywords as part of indexed titles.

Google supports the Boolean operators AND and OR, as well as the removing and including functions “-“ and “+”. Furthermore, it also features phrase search with quotation marks, wildcards, stemming and website specification, as well as Meta words search [Google, 2010c].

Yahoo also gives the possibility to use the Boolean operators and the function to require or exclude words. It allows the use of quotation marks, wildcards, stemming to expand search results, and Meta words as described above.
In addition, time can be saved by Yahoo! Shortcuts, symbols and keywords for specialized answers that appear directly on the results page, such as calculator or gas prices [Yahoo, 2010a].

As for Bing, like the other two search engines, Boolean operators can be applied, as well as Meta words, and placing words into quotation marks. It does not mention wildcards, but rather stemming. Similar to Yahoo Shortcuts, it has instant answers to get the information quickly [Microsoft, 2007].

Although the majority of users generally don’t leverage the advanced search options, they can be helpful to create very specific searches.

Google, Yahoo have an Advanced Search, which includes the previous discussed functions and additionally, settings for the number of results per page and the file type, such as PDF or Excel files, and safe search to filter adult content from search results are provided. There is also the possibility to specify the language of the web pages. While Google offers 46 languages, you can select within 32 languages at Yahoo as of September 2010.
Bing does offer Advanced Search options as well with a specification on site or domain, region and 41 languages. It does not contain as much options as the others, but there are settings to limit the search based on language as well as the language of the display site, and settings for safe search and results on each page.

All three search engines offer, additional to the text documents, other search sevices, namely from images, maps and news to audio files and videos.

4.2.4 Technology

Concerning technology of search engines, two aspects will be examined: The speed and the ranking for relevant results.

During experiments run by Google, the web search latency was increased to see users' reaction. The findings was, the longer the user has to wait, the fewer the number of his searches. Although the original latency returned, the loss of searches lasted for a time [Brutlag, 2009]. These experiments prove that the speed matter to users and thus, to search engines as well.

While Google declares that a query usually takes less than half a second and the time for retrieving a search query is always stated under its search box, unfortunately no information about the exact speed of Yahoo and Bing could be found.

In September this year, however, Google came up with a new search technology that, without much doubt, outperforms the other search engines. This technology concerns the factor speed. Previously, Google, Yahoo and Bing offer the list (from Yahoo called "Search Assist") under the search box with predicted results as suggestions for a quicker search process. While it still remains for the other two search engines, Google implemented early September "Google Instant", an addition to this feature. This enhancement gives users the opportunity to "search-before-you-type" [Google, 2010a]. What happens is that during typing the query, it does not only give the recommendation below the search box, but shows immediately the results on the SERP.

Without even typing the entire search term or pressing “Return”, it does not only predict and show more results, but it does it faster than before. Surprisingly, it saves the user two to five seconds of the whole search process [Google, 2010b]. Currently, Google Instant Search is only available in the English version on google.com.

In terms of ranking, search engines use methods to rank web pages by the relevance they expect from a user’s query, as explained in chapter 3.1. The goal of ranking is to maximize the usefulness and relevance of the results. Different weightings make the big differences between search engines. Ranking algorithms are not published and varies over time as the web changes and new techniques evolve. There are, however, some assumptions and known basic procedures of the algorithms. Especially webmasters have been accurately studied ranking factors in order to optimize their website and get listed in search engines, since they cannot pay for relevance ranking.

Google uses the famous and patented query-independent factor called PageRank that determines the link popularity. This algorithm, named after Larry Page, analyzes the whole link structure of the web and assesses which pages are most important. Google assumes that votes for a page’s importance can be assigned to links. The more important Google believes a page is, that means the more votes it has, the higher is its PageRank, and thus, its listing on the SERP. The link should ideally be from pages that are as relevant as possible. Not only the quantity of hyperlinks that point into a page, called inlinks, is a ranking factor, but also relevant content and the quality of the pages [Langville & Meyer, 2006]. With a query-dependent that is called Hypertext-Matching Analysis, Google evaluates the full content of web pages as well as the content of neighboring web pages to determine if they are relevant to the conducted query [Google Inc., 2010].

As for Yahoo! Search, the algorithm has some similarity to the Google algorithm. It’s been claimed that Pagerank affects the Yahoo ranking at some point. The ranking consists of the analysis of web page text, keywords in the title, description, source, and associated links. Most importantly, the title must contain major keywords. It can be advantageous to also include them in the description and category. Another part of its algorithm is click-popularity, what means that the number of document clicks from a results page will be counted [Yahoo! Help, 2010].

The Bing ranking is completely automated, its algorithm is complex and never human-mediated. Every time it updates the index, it changes the relevance rankings. Bing "analyzes the quality and quantity of indexable webpage content, the number, relevance, and authoritative quality of websites that link to web pages, and the relevance of the website's content to keywords" [Bing Webmaster Center, 2010]. In Bing community [DeJarnette, 2009], guidelines for SEO are given for improving site ranking. Bing mainly emphasizes on keywords, especially in the domain name and URL, and great, original content that are directed to the desired audience. Additionally, the architecture of the content should be well organized with regard to images to help MSNbot read and crawl the site. As already mentioned, to achieve optimal ranking, inbound links from authority sites, ideally of high quality, are very valuable. In general, those inlinks indeed pay a major role in ranking.

Search Engine Comparison Table			
	Google	**Yahoo!**	**Bing**
Database			
Index size	Rank 1 (> 12 bn pages) Includes Deep Web	Rank 2 (> 9 bn pages) Deep Web	Rank 3 (> 0.9 bn pages) Deep Web
Crawler name/ Actuality	Googlebot Fastest, many daily updates	Yahoo Slurp No clear frequency	MSNbot Frequent updates
Capability			
Search Operators Advanced Search	• Boolean operators • - to remove • + to include • "" quotation marks for exact words • Wildcards • Stemming • Meta word search • Advanced Search Form • 46 languages	• Boolean operators • - to remove • + to include • "" quotation marks for exact words • Yahoo shortcuts • Wildcards • Stemming • Advanced Search Form • 32 languages	• Boolean operators • - to remove • "" quotation marks for exact words • Instant answer • Stemming • Advanced Search Form • 42 languages
Technology			
Speed	Shown for every query Google Instant "search-before-you-type"	Not exactly known	Not exactly known
Ranking	PageRank, hypertext matching analysis	Keywords, click-popularity	Automated, emphasis mainly on keywords, inlinks

Table 4.2 Search Engine Comparison Table

4.2.5 Summary

All in all, even though Yahoo and Bing show good performance, Google deserves its first position of the major search engines ranking, because besides its popularity, it performs relative outstanding work, and keeps coming up with innovations.

In late July 2010, however, Microsoft Bing and Yahoo announced their merger. Yahoo has been struggling to make profits in the past years. As outlined in an article of PC World, not only Google but also Bing was growing at the expense of Yahoo [Newman, 2010]. This is also visible in the table 4.1: Referring to the percent change, Yahoo was far behind the other two search engines.
From now on though, Yahoo! search will be powered by Bing in both the US and Canada and they will together compete with Google. Within the next two years, Bing and Yahoo will also be officially merged in other countries to complete their transition. Because of the reason that this is not the case now, both were still separately examined in the previous chapters, based on their prior technology.

However, the consolidation has a big impact on the ranking factors and therefore the SEO. How does this Yahoo's assimilation by Bing affect their technology? Now that Bing took over Yahoo! Search, does it mean that only Bing ranking factors matter?
The Microsoft-Yahoo deal does involve the switch of Yahoo's technology to Bing. Consequently, whether the ranking position of a website has recently changed most likely arises from the merger. Bing determines the rankings of both. As soon as a web page is ranked well on Bing's list, it will be as well on Yahoo's.
Taking this into account, web designers should now concentrate on optimizing for Google and Bing. It should be kept in mind that while Google gives more weight to links, Bing focuses more on keywords. The answer to the question, whether "Binghoo" will keep up or even beat Google, can probably only be given over time after some observations [Link-Assistant.com, 2010].

4.3 Challenges

Search engines can be very useful in many aspects and they always strive to supply the users need need. However, there are challenges or issues with search engines that should be taken seriously.

The factors that have been investigated for the comparison in chapter 4.2 belong to the challenges of the search engines, since they need to be improved to meet the web searcher's need and also to be competitive. The criteria that differentiate one search engine from another can indeed be challenging.

First of all, search systems want to keep up not only with the growth of the web but they also want to crawl and index as much valuable information as possible, including the Invisible Web. Then, the spiders are expected to visit the ever-changing web pages to update the index. On the one hand, this can overcharge spiders to first crawl billions of pages and then visit them over and over again. There is not only a problem in the time lag between finding a page and re-crawling it to keep the actuality of the index, but also between a new page on the web and discovering it [Sherman & Price, 2001].
On the other hand, crawling in general causes a lot of costs. In [Sherman & Price, 2001], it's indicated that crawling is the most expensive part of maintaining search engines that are, for this reason, forced to set a limit for the database size as well as the re-crawl frequency. They also added that the potential pages, which have not been indexed due to the limits, are not part of the Deep Web, because these pages are visible, but just not chosen to be indexed. Search engines need to watch out for duplication and spam not to deliver irrelevant information to the end user.

Challenges in the capability of search engines are also a big issue. With the options of powerful features and additional search functions, they are trying to help the users to search more precisely and therefore, make it easier to them to get the desired information. The problem is, most people don't take advantage of the offered tools, but instead, they just type a few keywords for a query [Lewandowski, 2005a]. Thus, search engines have to find a way how to improve basic queries so that they can provide users successful research at the same time.

Another serious concern is indeed the technology that search systems keep trying to change and amplify. Both, great quality and speed of search results are demanded which, however, can exclude each other: Speed can constrain thorough searching of the expanding web and vice versa. Fortunately, search engines are making progress in technology improvements, also in ranking for better quality, as partially seen in the analysis of major search engines.

Problems also occur in the definition and measuring of relevancy. A document that is considered to be relevant to a person can be as well irrelevant to another. It should therefore be distinguished between the terms relevancy and pertinence. While relevancy allows the query and the engine to be objective measured, pertinence is corresponded to the usability of the results for the end user. In other words, only the user is able to determine the value of pertinence [Lewandowski, 2005b].

Last but not least, the Invisible web still turns out to be very challenging for search engines, because its size is comparing to the surface web supposedly 20 to 50 times larger, but often contains data of high quality. But technical barriers make it hard for general-purpose search engines to find them [Lewandowski, 2005a].

4.4 Merging Search Results for Best Performance?

Since all the three previous discussed search engines have their own advantages but also disadvantages, one would perhaps wonder which engine to use for best results in a quick way.
It's out of question that it takes too much time to type a query in every search engine's box to complete an efficient search. But if you only use one search tool, you might not get the information that is available in other search engines' index.
The solution proposed to solve this problem is meta-search engines that were described in chapter 3.3. They merge search results from multiple search engines and web directories, sometimes even the Hidden Web, and display the best combination of them in their SERP. Obviously, this way, the best information from all the selected search engines is gathered in one place simultaneously and it can reduce the amount of time spending on switching from one site to another.
It still needs to be clarified whether several search engines are more effective than only one favorite search engine to get to sufficient usable information. The question if a meta-search engine delivers more valuable search results will be examined next.

4.4.1 Overlap between search engines

To verify the assumptions of the previous section, results from the studies by Dogpile can be exemplified.

Dogpile is a meta-searcher which fetches results from leading search engines, including the three analyzed before, among which users are allowed to make a selection for their individual search. It belongs to one of the most popular meta-search engines and won the Best meta-search engine 2003. In 2006 and 2007, it was ranked highest in customer satisfaction by J.D. Powers and Associates, a global marketing information firm [Dogpile, 2010].

A very interesting research study by Dogpile was published in April 2007 that measured the overlap and ranking differences of the four leading web search engines back then, Google, Yahoo!, Windows Live and Ask Jeeves. Overlap occurs for a given query when a result from a search engine matches a result from another. The study from Dogpile, collaborated with researchers from Queensland University of Technology and the Pennsylvania State University, shows that there is a great difference in the top results of web search engines.

To get representative findings, over 19 thousand random user-entered queries had been tested with the help of a tool which automatically retrieved the search engines and stored the result data after capturing them from all the first pages. People rarely go beyond the first page of the results form [Lewandowski, 2005a]. The measure of only page one was the limitation of this study, because it's a barometer for the most relevant results from the search engines. A distinction between organic results and sponsored results was also necessary, because both have their own ranking position.
The findings showed that from the four tested engines, there were 88.3% total results unique, that means no overlap to one search engine. By any two search engines, the percentage of total results shared was 8.9%, by three engines 2.2% and by four only 0.6%. This means that the overlap across first results page of the evaluated search engines was only 0.6% for a given query [Dogpile, 2010].

In addition, if a person only searches Google, he won't get approximately 72.7% of the web's best results showing on the first page. If he, on the contrary, only uses Yahoo, he can miss 69.2%, and by utilizing only Live, 69.9% may result in a loss of the first page answers [Dogpile, 2007].

Compared to earlier researches from April and July 2005, done by Dogpile and collaboration, a trend could be observed: The content on search engines over time is unique and it's assumed that it will continue as each engine will keep modifying their crawling and ranking technologies [Dogpile, 2007].

To give a comparison between the overlap results of the evaluated search engines, Google Yahoo, Live and Ask, and the meta-search engine Dogpile, Table 4.4 can be studied.

	Percentage % of **G-Y-L-A** Total Results
Shared by all 4 engines	0.6%
Shared by any 3 engines	2.2%
Shared by any 2 engines	8.9%
Unique to 1 engine	88.3%

	Percentage % of **Dogpile.com** Total Results
Matched with all 4 engines	97.9%
Matched with any 3 engines	94.0%
Matched with any 2 engines	78.5%
Matched with any 1 engines	24.4%

Table 4.4 Overlap of Google-Yahoo-Live-Ask and Dogpile Total First Page Results
Source cf. [Dogpile, 2007]

To support Dogpile's research study, tests by Greg Notess, owner of Search Engine Showdown, can be regarded. He makes regular researches on overlap and also detected a very little overlap between major search engines [Sherman & Price, 2001].

As briefly mentioned, users don't look further than the first results page. This problem could be solved by clustering search results. Instead of the general known ranking list, hierarchical clustering of results could be created, so that users can simply select the category that is most appropriate to their needs. Some meta-search engines already support this method [Langville & Meyer, 2006].

In conclusion, especially due to each search engine's lack of overlap for assuring comprehensive results, retrieving more than one search engine might be very helpful and simplify the browsing.

4.4.2 A Web Searcher's Best Friend

In the next paragraphs it will be discussed whether meta-search engines are the most effective engines for retrieving information, and thus, their donation as the web searcher's best friend is justified.

As Mohamed [2004] reported, one of the serious challenges of meta-search engines is to make the best possible combination of search engines in order to provide the most relevant results.

The second challenge is to decide which method is most appropriate to aggregate the rank order of the retrieved sets. Each of them has to use effective and efficient merging techniques.
Meta-search engines' advantage that combine results of various sources for better performance can induce the delusive assumption that they give a much broader coverage of the web.
Though, Meta-search engines have limitations on the total number of results; they don't necessarily give all the pages that match the query [Langville & Meyer, 2006]. As outlined in [Sherman & Price, 2001], a portion of them are retrieved with less precision, at which expense they are increasing the potential relevance of results. While some difficulties can be managed by any of the single search systems, it most likely happens that some other problems, with which each of them is confronted, can't be solved by merging the results either. One of meta-search engines' disadvantages is also that they only give irrelevant results back from search engines or directories that don't support additional search functions as described.

On this account, meta-search engines are still in need of improvement.
In Mohamed's dissertation [Mohamed, 2004], he proposed a framework that can be used in the building process of meta-search developers. It should enhance data fusion technique of meta-search engines, which include the selection of databases, results combination and results merging. The goal was to make a research on how the optimal rank order for search engines can be defined, how to select the best combination from a set of search engines, and how to choose the best rank aggregation method for retrieved and combined results.
He came to the conclusion that larger search engines don't always retrieve more relevant information than engines with smaller databases. Therefore, meta-search developers should not depend on the size of the database in order to select search engines. Instead, their overall performance should be evaluated before ranking the database.
Another observation showed that it can be asserted, that there is more overlap of relevant, well-linked and popular pages between major search engines than irrelevant documents.
Finally, within three merging function being tested, the function which considers overlap documents, called Global Similarity Function, tends to perform better than the other two, interleave and rank similarity function [Mohamed, 2004].

While the mentioned points could be done to improve meta-search, more could be provided by an optimal search engine. After all the problems that a web searcher's best friend would have to face, the question may come up if it is even possible to build a best search engine.

The problem of building such an engine lies in finding the accurate information that the user desires.
Co-founder of Google Larry Page pointed this out by defining the perfect search engine: The ideal search engine must exactly understand the purpose of each research to provide exactly the information requested [Google Inc., 2010]. It's a big challenge, but at the same time, every search engine's main goal to understand and provide exactly what the user means and wants. Thus, it is appropriate that web is referred to as a huge haystack of information in which web users are looking for a needle [Baeza-Yates & Ribeiro-Neto, 1999; Sherman & Price, 2001].

User behavior can be problematic for search engine developers, since no user or user group is alike; they are one-of-a-kind and have difference experiences in searching the web, thus, they can't be treated the same way.
Their habits can also be inconvenient in the respect that, for example, many people neither use basic and advanced search options, as implied in 4.3, nor a set of words for a query, because they don't know how to utilize them or they are not aware of the benefits.

At least it is certain, that there is no perfect search engine as Larry Page described, because searcher's intention is subjective and difficult to quantify. As seen before, it is not possible for search engines to exactly understand every user's intent and define pertinence correctly at all times.

Because of the demonstrated problems that the web additionally poses, teaching the user how to properly profit from search services may be easier and more effective, as reported in [Baeza-Yates & Ribeiro-Neto, 1999]. This way, web searchers would be able to know how to search more efficiently, and also make use of the powerful search operations or tool and therefore, retrieve more successfully.

5. Conclusion

One of the objectives of this paper was to understand the search engines usability and significance. As could be seen throughout this paper, search engines are popular and successful, because they are convenient, not only in the common web user's perspective for finding information, but also from the web designer's point of view for their level of awareness and business purposes.

Throughout this paper, the major search engines Google, Yahoo, and Bing was being compared in matters of their characterization, such as index size, actuality of the database, the search engines capabilities and technology.
Google seems to deserve to be ranked first under the major search engines, not only due to its comprehensive coverage of the web, but also its technology along with great innovations. One of the recent innovations of Google could be observed during the comparisons, and also facts about the Microsoft-Yahoo deal in July this year, meaning Bing powering Yahoo! search. This is the start of the battle between the search giants Bing-Yahoo and Google, their biggest competitor that is doing his best to keep the first position under the major search engines.

After understanding what meta-search engines are and how they function, an analysis helped to figure out if merging the results from several top search engines also provides the best performance.
In dependence on studies of Dogpile, findings show that within the four search engines Google, Yahoo, Live (former Bing), and Ask overlap across their first results page was only 0.6% for a given query. On the contrary, each engine still finds a large amount of unique results search engines what means that they have a lack of duplication and each of them mostly don't see pages as equally important. In other words, the imputation that search engines are the same is a myth.
Meta-search engines collect and thus, cover the best results of the sources, including overlap. This is crucial, since overlapped documents tend to be more relevant.

The question, however, if multiple search engines can outclass single search engines and at the same time optimize the query was critically observed. In spite of the comprehensive search of the web and the timesaving process through reduction of search engines consulted, there is still criticism and improvements needed for meta-search engine.

The last goal was to find the requirements for the ideal search engine. The discussion if a web's best search engine even can exist is arguable for the reason that user's need and intention can't be measured or guessed. It is difficult to measure relevancy that fulfills the user's wishes, because in the end it's the pertinence that is essential.
Some search services try to lead the user to his desired information by suggesting some results. Alternatively, it's being proposed to teach the user ways to search on the internet efficiently, instead of speculating what he might want.

The analysis in this thesis could have been gone more into depth, but overall, the answers to the given research questions at the beginning of the paper could be found in the meanwhile. It would have been interesting, to examine the search engines more thoroughly and reanalyze some findings in order to find contemporary performance, but due to resource it could not be implemented. For future works, a test can be suggested to, for example, measure the exact speed of Bing and Yahoo to see if they are different, even though they're both based on Bing's technology, and compare them with Google search's latency. Furthermore, as far as meta-search engines are concerned, studies can be made to find out, why they are not known or being used like the general-purposed search engines, and how to make them more popular and improve their market share.

The up-to-dateness problem of search engines was directly observed during the research for this paper. A certain weight was put on updated information for the topic, but in the results list were not always current and reliable documents on hand.
At the same time though, surprisingly quite many news of search engines are regularly released, such as Google Instant or the Bing-Yahoo deal, what confirms that they continue to improve and innovate to keep up with the always changing web and its everlasting growth.

While search engines work on enhancing to meet their audience with innovative inputs, they can also try to regularly impart the knowledge to the users in the future, who will benefit more from web search. In this case, web users might be able to seize the chance to be more successful and eventually find their needle in the haystack.

References

[Baeza-Yates & Ribeiro-Neto, 1999] Baeza-Yates, R., & Ribeiro-Neto, B.: *Modern Information Retrieval.* Essex, England: ACM press.

[Bing Webmaster Center, 2010] Bing Webmaster Center. (2010, March 15). *Bing Webmaster Center FAQs.* Retrieved September 17, 2010, from http://download.microsoft.com/download/4/5/4/454C13D4-D94D-4B54-8E46-FE403DF7632B/WMC_FAQ.pdf

[Brutlag, 2009] Brutlag, J. (2009, June 22). *Speed Matters for Google Web Search.* Retrieved September 17, 2010, from http://code.google.com/speed/files/delayexp.pdf

[Clay & Esparza, 2009] Clay, B., & Esparza, S.: *Search Engine Optimization All-In-One For Dummies.* Indiana: Wiley Publishing, Inc.

[comScore, 2010] comScore. (2010, January 22). *comScore Reports Global Search Market Growth of 46 Percent in 2009*. Retrieved September 15, 2010, from comScore: http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009

[DeJarnette, 2009] DeJarnette, R. (2009, September 3). *Search Engine Optimization for Bing*. Retrieved September 17, 2010, from Bing Community: http://www.bing.com/community/blogs/webmaster/archive/2009/09/03/search-engine-optimization-for-bing.aspx

[Dogpile, 2010] Dogpile. (2010). *About Dogpile.* Retrieved September 22, 2010, from http://www.dogpile.com/

[Dogpile, 2007] Dogpile. (2007, April). *Different Engines, Different Results: Web Searchers Not Always Finding What They're Looking for Online.* Retrieved September 12, 2010, from http://www.infospaceinc.com/onlineprod/Overlap-DifferentEnginesDifferentResults.pdf

[Gil, 2010] Gil, P. (2010, September). *The 10 Best Search Engines of 2010*. Retrieved September 5, 2010, from About.com: http://netforbeginners.about.com/od/navigatingthenet/tp/top_10_search_engines_for_beginners.htm

[Google, 2010a] Google. (2010). Retrieved September 18, 2010, from The Official Google Blog: http://googleblog.blogspot.com

[Google, 2010b] Google. (2010). *About Google Instant*. Retrieved September 18, 2010, from http://www.google.com/instant/

[Google, 2010c] Google. (2010). *Google search basic: More search help*. Retrieved September 9, 2010, from Google web search: http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=136861

[Google Inc., 2010] Google Inc. (2010). *Corporate Information*. Retrieved September 15, 2010, from Google: http://www.google.com/corporate/tech.html

[GoogleGuide, 2007] GoogleGuide. (2007, February 2). *How Google Works*. Retrieved September 17, 2010, from GoogleGuide: http://www.googleguide.com/google_works.html

[Langville & Meyer, 2006] Langville, A. N., & Meyer, C. D. (2006). *Google's Page Rank and Beyond.* Princeton: Princeton University Press.

[Lewandowski, 2005a] Lewandowski, D. (2005). *Web searching, search engines and Information Retrieval.* Retrieved September 17, 2010, from

[Lewandowski, 2005b] Lewandowski, D.: *Web Information Retrieval: Technologien zur Informationssuche im Internet.* Frankfurt am Main: DGI. http://arizona.openrepository.com/arizona/bitstream/10150/106395/1/isu_preprint.pdf

[Lewandowski, Whalig, & Meyer-Bautor, 2005] Lewandowski, D., Whalig, H., & Meyer-Bautor, G. (2005). *The Freshness of Web search engines' databases*. Retrieved September 17, 2010, from http://code.google.com/speed/files/delayexp.pdf

[Link-Assistant.com, 2010] Link-Assistant.com. (2010, August 31). *Yahoo Switches to Bing: Check How this Affected Your Site and Turn it Into Advantage*. Retrieved September 21, 2010, from http://www.link-assistant.com/blog/how-to-check-your-old-yahoo-rankings-see-how-the-switch-to-bing-affected-your-site/

[Madhavan, Ko, Kot, Ganapathy, Rasmussen, & Halevy, 2008] Madhavan, J., Ko, D., Kot, Ł., Ganapathy, V., Rasmussen, A., & Halevy, A. (2008). *Google's Deep-Web Crawl.* Retrieved September 25, 2010, from http://portal.acm.org/citation.cfm?id=1454163&coll=GUIDE&dl=GUIDE&CFID=106289065&CFTOKEN=34426368&ret=1#Fulltext

[Manning, Raghavan, & Schütze, 2009] Manning, C. D., Raghavan, P., & Schütze, H.: An Introduction to Information Retrieval. Cambridge, England.

[Microsoft, 2007] Microsoft. (2007, October 24). *Do What I Mean, Not What I Say*. Retrieved September 19, 2010, from Bing Community: http://www.bing.com/community/blogs/search/archive/2007/10/24/do-what-i-mean-not-what-i-say-part-1-of-2.aspx

[Microsoft, 2010] Microsoft. (2010). *About Microsoft*. Retrieved September 16, 2010, from Microsoft: http://www.microsoft.com/about/en/us/default.aspx

[Mohamed, 2004] Mohamed, K. A.-E.-F. (2004, January 29). Merging Multiple Search Results Approach For Meta-Search Engines. University of Pittsburgh.

[Net Applications, 2010] Net Applications. (2010, August). *Search Engine Market Share*. Retrieved September 15, 2010, from Netmarketshare: http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=5&qpct=2#

[Newman, 2010] Newman, J. (2010, February 19). Retrieved September 17, 2010, from PC World: http://www.pcworld.com/article/189801/yahoomicrosoft_deal_makes_bing_better_doj_says.html

[Olsen, 2004] Olsen, S. (2004, March 2). *Yahoo crawls deep into the web*. Retrieved September 2010, from cnet news: http://news.cnet.com/Yahoo-crawls-deep-into-the-Web/2100-1024_3-5167931.html

[SearchEngineWatch, 2004] SearchEngineWatch. (2004, February 18). *SearchEngineWatch.com*. Retrieved September 16, 2010, from Yahoo! Birth of a New Machine: http://searchenginewatch.com/3314171

[Sherman & Price, 2001] Sherman, C., & Price, G.: *The Invisible Web.* Medford, NJ: Information Today.

[The Register, 2009] The Register. (2009, June 4). *Microsoft to Acquire Powerset*. Retrieved September 16, 2010, from Bing Community: http://www.theregister.co.uk/2009/06/04/bing_and_powerset/

[Thurow, 2003] Thurow, S.: *Search Engine Visibility.* USA: New Riders Publishing.

[Web Developers Notes, 2010] Web Developers Notes. (2010). *Best Search Engine of the web - Which one?* Retrieved September 9, 2010, from http://www.webdevelopersnotes.com/basics/best_search_engine.php

[Yahoo, 2010a] Yahoo. (2010, August 30). *Search Tips*. Retrieved September 19, 2010, from Yahoo! Help: http://help.yahoo.com/l/us/yahoo/search/basics/basics-04.html;_ylt=AvjXDq9uArlr2TH9si4KMcMuqCN4

[Yahoo, 2010b] Yahoo. (2010). *Yahoo Press Room*. Retrieved September 16, 2010, from Company Info: http://yhoo.client.shareholder.com/press/

[Yahoo! Help, 2010] Yahoo! Help. (2010). *Yahoo! Search Help Topics: Search Indexing and Ranking*. Retrieved September 20, 2010, from Yahoo! Help: http://help.yahoo.com/l/us/yahoo/search/indexing/;_ylt=AilwLgBe0n4UR9OOCiowOsWutHhG

[Yamada, Craswell, Nakatoh, & Hirokawa, 2004] Yamada, Y., Craswell, N., Nakatoh, T., & Hirokawa, S. (2004, May). *Testbed for Information Extraction from Deep Web.* Retrieved September 25, 2010, from http://research.microsoft.com/pubs/65247/yamada_www2004poster.pdf