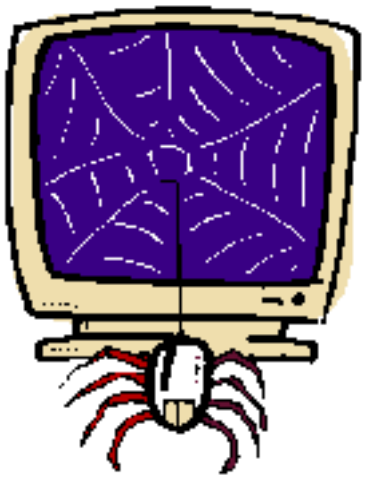
Guiding Children on the Internet

Khanh Vu Koen Vermeer

Master Thesis
Research number: 552

June 30, 2006



IBM
Corporate Community Relations
Supervisor: Warner Dijkhuizen

Radboud University Nijmegen
Information Retrieval and Information Systems
Supervisor: prof. dr. ir. Theo van der Weide
Institute for Science, Innovation & Society
Supervisor: prof. dr. Ben Dankbaar

Radboud Universiteit Nijmegen

Abstract

Children need guidance in their journey on the Internet. An Internet search engine for children is a good way to accomplish this. The requirements for such an application differ from other search engines.

In this thesis the requirements for a children's search engine are described and a technical solution to fulfill these requirements is given. To successfully implement and publish such an application, it has to be profitable. For that purpose a business research on making a search engine for children profitable is included. This thesis contains all aspects for implementing a successful Internet search engine for children.

Preface

The World Wide Web is an extremely large and complex medium, where you can easily get lost or get stuck in its complex web structure. The Internet is used by almost every age group. Children are not capable to surf on the Internet without assistance. As it is normal to send our children to schools to prepare and guide them for the real world, it should also be normal to prepare and guide them for the cyber world, in other words the Internet. The developments on the Internet are extremely fast, so fast that many parents don't seem to be able to keep up with it. Many parents don't have the experience and knowledge to guide their children on the Internet. There is a need for an external guidance method to assist these parents in the Internet education of their children.

In collaboration with IBM we have done research on guiding children on their journey on the Internet. IBM is a worldwide operating corporation specialized in information technology. They are one of the largest corporate contributors of cash, equipment, and people to non-profit organizations and educational institutions across the U.S. and around the world. Since information technology is their core business, they want to use their expertise to address societal issues. One of their key contributions area is public education, from kindergartens to universities. Using technology to make children capable of independently finding information on the Internet clearly fits into the community contributions area of IBM.

Contents

List of Figures

List of Tables

1 Introduction

The Internet has become an important part of our modern society. It becomes more and more difficult to find your way in this continuously growing and unstructured world. The large and unstructured characteristic of the World Wide Web has resulted in a demand of tools to guide the user in its search for information on the Internet. This demand has created a business for this kind of search tools, also known as Internet search engines. These search engines are mostly supplied for free to users by companies who in general have a profit goal. So how do these companies generate a profit? The search engine business is quite different when compared to traditional businesses. In traditional businesses you can talk about a market where products or services are sold to potential buyers. There is a clear seller/buyer relationship. Most search engines are freely accessible for the users, so there is no clear view what is sold and what is bought. So we will talk about the search engine business instead of the search engine market.

Another essential aspect of search engines, besides the financial aspect, are the users. Most search engines focus on an adult target group. But just like adults, children also make use of the Internet, but it is obvious that children may not have access to all information on the Internet like adults. Examples are websites with pornography or violence. Therefore a search engine for children that blocks unwanted information is certainly needed. Besides that children cannot have access to all websites on the Internet we can ask ourselves the following question. Are there any other differences between adults and children concerning Internet search engines?

A good search engine alone is not enough to find the needed information. It is also important that you have the knowledge to judge the relevancy of a search result from a search engine. This information need can be divided into four levels of awareness in seeking information; visceral, conscious, formal and compromised. People go through these four levels of awareness when they seek for information [76].

For adults it can be difficult to find information on the Internet with the help of search engines because they might not have enough knowledge to judge the relevancy of the search results. If you take a look at children, you can ask yourself if their knowledge to judge a search result differs from that of an adult. Do children search in a different way than adults? Do they have different information needs than adults? And if there are differences, will this mean that children need a different kind of search engine than adults?

1.1 Problem Area

Internet search engines help finding information stored on the World Wide Web. There are several general-purpose search engines available on the Internet, and there are also tens of thousands of search engines that are part of a single website [9]. Most of the general-purpose search engines are there to support adults in their search for information.

Since children are also Internet users, they also have information needs like adults. But the difference with adults is that children cannot be exposed to all information that is available on the Internet. A lot of information on the Internet is not suitable for children. Examples are pornography, violence and Internet shops. So parents and teachers have a need to control the information access of the children.

A solution is to use a rating system that relies on websites operators to indicate the nature of their material. Internet browsers can be configured to only allow children to visit sites that are rated at the level that the parents specify. The advantage of this method is that only appropriate websites can be viewed. The disadvantage is that many appropriate websites have not submitted themselves for a rating and will therefore be blocked. This way a lot of relevant information is withhold from children. [48, 64]

This problem was addressed in the project *Giggle: Internet Search Engine for Kids* [78] of IBM. The aim of that project was to successfully design and implement an easy to use tool that supports children in their search for information on the Internet.

This project is a continuation of that project. Within that project an extensive requirements analysis has been done, concerning the demands such a search engine should live up to. The project also delivered a prototype of the search engine, which should match the corresponding requirements. This prototype was considered a study model for the yet to deliver product.

1.2 Research Objectives

This project starts where the previous project has ended. In the previous project a lot of research has been executed on the requirements of a search engine. The result was an extensive requirements analysis. But the research on the technical aspects of a search engine was not extensive. This resulted in a prototype that matched the corresponding functional requirements but did not match the non-functional requirements like a short reaction time, high relevance of retrieved results and filtering.

The aim of this project is to go further with the work of the previous project. This means doing extensive research on the technical aspects of a search engine for children. This research should result in more knowledge on Information Retrieval methods.

An important goal for search engine companies is to earn back their investments and to make a profit out of their services. Most search engines are freely accessible for users, so search engine companies do not earn money in selling their services to the user. The search engine business is therefore different when compared to other businesses where a product or service is sold to the users. Research on the characteristics of the search engine business is needed before you can formulate a business strategy that will help you to earn money out of a search engine. Our goal is to identify these characteristics.

1.3 Research Questions

Main question

What are the requirements to implement a successful Internet search engine for children between 8 and 12 years?

Sub questions

What are the requirements of an Internet search engine for children between 8 and 12 years?
Extensive research on the requirements for an Internet search engine for children has been done in IBM's previous project [78]. These requirements have to be studied and reviewed so that they can be applied in this project.

What is an adequate method to classify documents?
Not all documents on the Internet are suitable for children. On the other hand there are documents that are suitable for children, but that are not found interesting by them. So some documents have to be excluded, and others have to be ranked in some way. For this purpose the documents have to be classified.

What is an adequate method to search for documents?
When searching for information you want to find your required documents as fast as possible. So it is important that relevance of the retrieval results is high and that the response time of the system is low. To achieve this we need a search method that delivers the documents in order of relevance in an acceptable amount of time.

What are the specific characteristics of the Internet search engine business?
To identify the characteristics of the search engine business for children, the research has to be done in a business that is rather similar to the search engine business for children. The existent business of the current public search engines is a good business to do the research on, because the business is mature in comparison to that of the business of search engines for children, which barely exists.

1.4 Research Strategy

This project is a continuation of the project *Giggle: Internet Search Engine for Kids* [78]. To answer the questions described in section 1.3 we will do a thorough review of the previous IBM project, with a special focus on the requirements analysis. These requirements can then be used as input for this project.

After the review we will start with the research of this project, namely the technical research and the business research. We hope that the technical research will aid in a better understanding of the information retrieval methods necessary for a good functioning search engine and the business research in more knowledge on how to make a search engine for kids commercially a success.

Finally the knowledge found in the technical research, together with the requirements analysis from the previous project will be used for a design and implementation of a prototype.

1.5 Thesis Structure

This document describes the process that the project is going through. It will describe chapter by chapter the activities that have taken place and the milestones that have been achieved.

- *Requirements Review:* Chapter 2 will discuss the review of the requirements analysis formulated during the Giggle-project [78]. All the separate requirements will be described and the relevance of each requirement for this project is discussed.
- *Information Retrieval Research:* Chapter 3 discusses research done in the field of Information Retrieval. It describes research done and results achieved in former research and the research that we did to improve the different aspects of the search engine.
- *Business Research:* Chapter 4 discusses the Business Research that has been done to identify the different aspects that are important to successfully commercialize a search engine for children.
- *Development:* Chapter 5 describes the development of the search engine for children that has been developed. It will discuss the design of the tool and the implementation of the prototype that has been done. Finally the test results of the various tests where the product is submitted to will be discussed.
- *Discussion:* Chapter 6 contains the discussion of the results of this project. We will discuss the approach, scientific relevance and the reflection of this project.

2 Requirements Review

The requirements formulated in the project *Giggle: Internet Search Engine for Kids* [78] will be discussed and reviewed in this chapter. The different types of requirements are discussed in the following sections. The functional requirements are discussed in section 2.2, the non-functional requirements are discussed in section 2.3. In section 2.4 we will present our conclusion and prioritization of these requirements.

2.1 Introduction

An Internet search engine is basically a software system. Before you can make a search engine, you need to know the characteristics of this system. A requirement is a description of what a system should do. These requirements were formulated after extensive research in the project *Giggle: Internet Search Engine for Kids* [78] by Weeda and van Gendt. This research was done in collaboration with parents, teachers, librarians and children in the age of 8 to 12 years old. These people were interviewed and the behaviour of the children using the Internet was observed. Their investigation has led to the following general conclusion:

> In order to support a child during their search on the Internet, they not only seek a search engine that will find information appropriate to their needs and interests, but they need a tool they feel comfortable with and attracted to, a tool that has been specifically built for children [78, p.51].

The requirements that were formulated are based on this conclusion. Before the requirements can be used for this research, they have to be reviewed critically. The requirements are divided into two categories, namely functional and non-functional requirements. The requirements are discussed in the following sections.

2.2 Functional Requirements

The functional requirements describe the behaviours (functions or services) of the system that support user goals, tasks or activities [78]. We are going to discuss all 10 functional requirements that Weeda and van Gendt have formulated in their requirements analysis.

F1. Instructions / Help

An instruction or help functionality is very common to computer applications. But most adult users avoid instructions and try to use web sites without having to read about what they are supposed to do [78, p.39]. Something similar also counts for children:

> Generally, if they are given to them, they are willing to read short instructions, however, they will generally not explicitly search for help functions and much more prefer learning by trial and error (source: teacher) [78, p. 209].

Conclusion

An instruction or help functionality is needed, but it should be kept simple for children. It is more important that the user interface will be intuitively easy to learn.

F2. Metaphor

The investigation of Weeda and van Gendt has led to the conclusion that metaphors like a character are needed:

> A child who is emotionally engaged will experience a high level of attention, improved memory, and they may learn more about their emotional being or self. One of the best ways to increase emotional involvement is to link the content to a recognizable, constant character [78, p. 40].

Weeda and van Gendt have formulated some characteristics of such a character:

> Though a character should be very consistent in dialogue, their dialogue should not be limited too much. Occasional use of variations makes the character less static and more believable [78, p. 40].

Conclusion

It is common in applications for children that there is a virtual guide or a character that will guide you with instructions through the program. The character has the function to guide the user through the application, the character should not do more than necessary and should be readily intelligible for children. Children must not be annoyed by the character. A character can be a good aid for children in using the application, but the character must be carefully created [8, 43].

F3. Initiate Searching Process

Sometimes a person doesn't know what to search for, but he knows that he has to find some information on the Internet. This is called a visceral information need [76]. Weeda and van Gendt have reported that there should be some kind of aid to help the user to start the search process. They have come up with two manners of helping young users to get started:

Directory possibility to choose from categories or defined objects

Example search results short list of some previous search results of other users

Conclusion

A directory structure is very common to search engines. Examples of this are Google Directory (`http://www.google.com/dirhp`) and Yahoo! Directory (`http://dir.yahoo.com`). This seems to be a good option in helping people to start with a search process. The categories should be general and broad enough for the specific target group. This will enlarge the effectiveness of the support in formulating the information need.

Example search results are coming from other users. We can ask ourselves the following questions: What is the chance that the search result from a particular user will support another user in formulating his information need? How many search results does the user has to see before he can formulate his information need? Before implementing this method, it needs to be more investigated.

F4. Formulation of Information Need

When searching for information, there is a need to formulate the information need. Weeda and van Gendt have reported the following about the formulation of information need by children:

> Describing their information need and determining which terms would best retrieve relevant information has shown to be problematic. In most cases the childrens searches using queries are too general and result in the retrieval of too many irrelevant documents [78, p. 40].

Weeda and van Gendt have proposed category selection from a directory and a relationship-map as an aid in helping the user to formulate his information need.

Conclusion

Formulating an information need is not a task that can be done in one step. Initiating a search process is the first step, from that point on the user can formulate the information need more precise and specific. This seems to be very difficult for children. There is a need to help them in formulating their information need. The category selection method and the relationship-map method seem to be very promising.

F5. Aid in Query Reformulation

Children don't find it straightforward to formulate alternative search terms when their initial search term doesn't seem to work [78]. Weeda and van Gendt propose that the search engine can help when too many or too few results are retrieved.

When too many results are retrieved, Weeda and van Gendt propose the following:

> In order to trim the results list, the user can narrow down the search by selecting a category or navigating through the relationship-map (including hyponyms and meronyms) [78, p. 41].

In the case that there are too few results:

> When the results list is small, the initial query can be expanded to include spelling variations, term synonyms, or term hypernyms (is a kind of) [78, p. 41].

Conclusion

A searcher needs to reformulate his search query when a search term doesn't give the desired result. Children don't seem to do this. So an aid in query reformulation seems to be necessary for a search engine for children. The functionalities that Weeda and van Gendt have proposed seem to be very promising, but a more advanced procedure is needed. A proposal for this can be found in section 3.4

F6. Following Progress of Search Process

> The query created (due to directory or relationship-map selection) or modified (due to too few or too many results) in order to retrieve the result should be shown to the user. In such a manner the user can both better understand the retrieval result, and also learn, by example, how to create an appropriate Boolean query [78, p. 41].

Conclusion

It is good for the user to see the path that he has taken to get to the search result. In this way the user can better understand the result, but he can also get more insights in searching for information with an Internet search engine. These insights are limited because the user will be a child between 8 and 12 years old who is not capable of learning and understanding all aspects of searching information on the Internet. For example learning how to create an appropriate Boolean query is too difficult for a child.

F7. Giving Relevance Feedback and Understanding Retrieval Result

Weeda and van Gendt argue that there should be the possibility to store previous choices and intermediate results. The argument for this functionality from their requirements analysis is that searching for information is an iterative process, possibly divided over several search sessions. This functionality will allow the searcher to have a choice between different search strategies and it will also reduce the working memory. There is also a need to better understand the retrieval result. Showing all accessible information like document title, URL where it was found, passage selection from the document will help the searcher to better understand the content with the associated link.

Conclusion

The idea to store previous choices and intermediate results is quite new. Other Internet search engines don't have this functionality. It seems to be a useful functionality because it can make it possible to search in many sessions without losing any previous found knowledge.

Showing accessible information like the title of a document is quite standard with search engines. Most search engines also give a short passage of the document. It would be helpful to show, in these passages, in which context the search terms (by highlighting terms for emphasis) are used in the document. The searcher can judge globally and better understand the document and predict the content associated with the link before he actually retrieves the document. This can speed up the searching process.

F8. Reversal of Actions

Weeda and van Gendt report that there is a need for an ”undo” function. This function will allow an easy reversal of actions to return to a previous state.

Conclusion

When the user is searching for information, the user takes some path to get to the desired information. But when the information is not desired, you want to get back and make another choice. An ”undo” function is very useful and needed.

F9. Navigation and Search Strategies

Searching for information is an iterative process. To find the desired information a searcher needs to be able to change the search strategy. The following comes from the requirements analysis of Weeda and van Gendt:

> The capability of formal operations, capable of thinking about ideas, considering various strategies and solutions involved in learning, memory and information processing without having to act them all out evolves around 11 years of age. Because this is the far end of our user group, we must aid the users in these aspects [78, p. 41-42].

There is a need to aid the users in the aspects that were mentioned above. Weeda and van Gendt have proposed the following navigation structures:

- *Query input:* An information query is matched against stored information. The query input must provide assistance with Boolean searches.
- *Directory:* The user can narrow down a search topic by selecting categories.
- *Relationship-map:* The user can navigate through the map by narrowing down or expanding the query at each step.

The above functionalities were already mentioned before. These functionalities can offer support for multiple functional requirements of a search engine for children.

Conclusion

The query input and directory functionalities are very common in Internet search engines. Besides that they are very useful, they are also known to users of Internet search engines. The use of these functionalities corresponds to other search engines.

The relationship-map as described by Weeda and van Gendt is never seen with Internet search engines, but it seems to be very useful for children. Adults usually have more domain knowledge, so they don’t really need a relationship-map. There are some interesting variations [30, 41] on this relationship-map method on the web, they are quite advanced though

F10. Sitemap

Weeda and van Gendt propose a sitemap for the navigation within the Internet search engine website.

Conclusion

A sitemap is very useful when the structure of the website is large and complex. An Internet search engine for kids should be kept simple, so a sitemap is in our opinion not necessary.

2.3 Non-functional Requirements

Non-functional requirements are constraints on the system or on the development process [72].

The non-functional requirements that were formulated by Weeda and van Gendt will be discussed.

NF1. Usability

Weeda and van Gendt have reported the following about usability:

> What is important to the user is the time needed to learn how to use the system, time to achieve goal, chance of doing something wrong, and knowledge-halflife (how long before one forgets it). The tool must be easy to learn, mostly the goal must be achieved within acceptable time frame, and the probability of doing the wrong thing must be limited to a minimum [78, p. 44].

Furthermore their research has indicated that children often play on the computer in pairs. According to them this fact should be considered in design choices.

Conclusion

Usability is very important for software products. A term that is often used with this feature is ”user friendliness”. If a program is not ”user friendly” it will often be seen as a failure, even if the functions that it performs are valuable. [57]

It is not clear to us what effect children often playing in pairs on the computer has on Internet search engines. For the time being we will only consider the one-person-behind-the-computer case.

NF2. Presentation

Presentation of the application is very important. Weeda and van Gendt have reported the following about presentation:

> The interface should be transparent, not requiring or calling for explicit attention, and in this manner, leaving the user to concentrate solely on his or her work, exploration or pleasure. An attractive site will encourage children to stay with the site and return for subsequent visits. Such a site must make them feel comfortable, yet have the feeling they are in control. In essence, it should reflect a home. A place they feel comfortable and in which all necessary facilities are available, yet having a room they can design and use as they wish (obviously, to certain limits) [78, p. 44]

Furthermore Weeda and van Gendt have given 12 guidelines and best practices to design an effective interface.

Conclusion

The appearance of the application should generate a positive feeling. The user is connected with the application through the interface of the application. If the user is not satisfied with the interface, then it is likely that the application will not be used, even when all functionalities are present. Presentation is a very important aspect and it has effect to the system as a whole.

NF3. Relevance of Retrieval Results and Filtering

The main objective to use an Internet search engine is to find needed information. The requirements analysis of Weeda and van Gendt have indicated that the primary aim is to increase precision with respect to retrieval results. This means that this will be a trade off between possibly retrieving inappropriate documents or a smaller set of documents. This trade off is different for adults and children. For adults it is not a big issue if there are some inappropriate documents among the retrieval results. Under the condition that most of the documents are relevant. The safety of children has a higher priority than a wide range of retrieved documents. An important aspect of an Internet search engine for children is filtering. This filtering also brings up some criticism. Some argue that it is a high-tech form of censorship. Others find it a false sense of security because users are relying on software rather than on human judgement. Despite of the criticism, research of Weeda and van Gendt has indicated that many stakeholders find it important that there will be a filtering mechanism.

The problem with automatic filtering is that it will not give a 100% result. It is impossible because the Internet is very unstructured. It is possible that the filter will block information that should not have been blocked, or it can let through inappropriate information. Automatic filtering will not guarantee full protection.

To be able to filter inappropriate documents there is a distinction between different sets of documents. From the requirements analysis [78] we have extracted the following different sets of documents:

- *White-listing:* Documents with benign and safe content with respect to our user group.
- *Black-listing:* Documents with inappropriate (unsafe, unfit, obscene or harmful to minors [39]) content. The research of Weeda and van Gendt has shown to filter on the following aspects:
  - adult content: sexually explicit graphic descriptions or images;
  - child pornography;
  - hate sites: advocating bigotry, hatred, or discrimination;
  - graphic violence: violent images, language, bomb-building, etc.;
  - criminal activity: promoting illegal activity;
  - other offensive sites: alcohol or drug abuse, or promoting suicide and violence;
  - incomplete or poorly maintained sites;
  - pop-ups, advertisements, and banners;
  - Unprocessable sites: Prevent any information to be shown that does not allow processing or content determination (such as encrypted information, or websites not allowing automatic access);

  - Prevention of exchange of personal information;
  - Prevention of online shopping and banking;
  - Prevent harmful doing to software and hardware: viruses, automatic installations (such as toolbars and programs).

- *Gray-listing:* Documents of which the listing category has not yet been determined, but their content could potentially be inappropriate.

- *Favorable document:* A document that adheres to the specifications described in the requirements document, such as reading levels and content topics.

- *Non-favorable document:* Any white-listing document that is not a favorable document according to one or more of the specifications in the requirements document is non-favorable. All black and gray-listings are also non-favorable documents.

- *Relevant document:* A document capable of answering the information need posed by the user. This may be an element of a favorable document set or a gray-listing.

- *Irrelevant document:* Any document incapable of answering the information need posed by the user. We propose that all documents belonging to the set of black-listing also belong to this set.

Conclusion

Relevancy of retrieval results and filtering is one of the most important aspects of an Internet search engine for children. Some kind of filtering is better than no filtering at all, even though automatic filtering does not guarantee full protection.

The classification of the different sets of documents described above will be presumed in our further research on Information Retrieval and Filtering.

NF4. Safety Warning

Research of Weeda and van Gendt has indicated that when a user leaves the website of the search engine, it should be explicitly warned of the dangers when leaving the website. An example of such warning is:

> You have chosen a resource outside the Giggle site. We have provided links to resources that we hope will be helpful to you, but remember, Giggle cannot vouch for their content. Thanks for visiting! [78, p. 50]

Conclusion

It cannot be expected from children that they have the knowledge to avoid all the pitfalls of the world wide web, but giving warnings each time the user leaves the Giggle site might cause irritation with the users. Besides the irritation we have doubts if the warning will be effective. Will it encourage the user to explore websites outside of the Giggle site?

NF5. Hardware, Software and Formats

The application must be compatible with existing hard and software that children use. Research of Weeda and van Gendt has indicated that the application will mostly be used on a Microsoft Windows 2000 operating system (or higher). And all the features should be applicable on an Internet Explorer and Mozilla browser. It must be possible that the application can run alongside a family filter. Furthermore the application must be capable of retrieving documents most frequently consulted and it must be able to view them on general user machines.

Conclusion

These results of the research by Weeda and van Gendt was done over a year ago. But we think that there have not been any major changes on this aspect. Since then Microsoft has not released a new operating system. It can be possible that more children use Windows XP than Windows 2000, but Windows 2000 is still being used by children.

NF6. Reaction Time

Nobody likes to wait, especially children who have a shorter attention span than adults, lack of patience and capacity to concentrate. It is very important that the application has a short reaction time. Weeda and van Gendt propose that some kind of animation can be used to distract children while they have to wait for the system to perform.

Conclusion

It is important to reduce the reaction time to a minimum. This reaction time is dependent on the hardware, the information retrieval methods and filtering methods. To improve the reaction time there will be research on information retrieval and filtering methods.

2.4 Conclusion

After our review of the requirements analysis of Weeda and van Gendt we came to the conclusion that their requirements analysis is in general valid and complete. Besides some small remarks (see 2.4.1) on their analysis we will mostly adopt the requirements from their requirements analysis.

2.4.1 Remarks

Weeda and van Gendt have proposed an advanced search option. In our opinion the user group is not capable of using the advanced search option. Weeda and van Gendt themselves concluded later on, after they have finished the requirements analysis, that this option is not useful for children [78].

We will not adopt requirements F10 Sitemap and NF4 Safety Warning. The website is kept simple, and a sitemap is only necessary with large and complex websites and a safety warning will not be effective, it might cause irritation with the users.

2.4.2 Requirements Overview

There are ten functional and six non-functional requirements. These requirements are enumerated in the tables below for an overview (tables 2.1 and 2.2).

Number	Description
F1	Instructions/Help
F2	Metaphor
F3	Initiate Searching Process
F4	Formulation of Information Need
F5	Aid in Query Reformulation
F6	Following Progress of Search Process
F7	Giving Relevance Feedback and Understanding Retrieval Result
F8	Reversal of Actions
F9	Navigation and Search Strategies
F10	Sitemap

Table 2.1: Functional requirements

Number	Description
NF1	Usability
NF2	Presentation
NF3	Relevance of Retrieval Results and Filtering
NF4	Safety Warning
NF5	Hardware, Software and Formats
NF6	Reaction Time

Table 2.2: Non-functional requirements

2.4.3 Prioritization

In this subsection we will discuss the prioritization of the requirements. We have divided the prioritization into two classes, primary and secondary.

Primary Requirements

The requirements has led to the following list of functionalities that must be implemented:

- *Instructions/Help (F1, NF1)*
  The help page contains an explanation of all the functionalities and tips how to search for information.
- *Query input (F4, F5, F9, NF1)*
  The user can search with the search engine by typing a query in the input field. The user can click on the 'Find' button to start the search task.
- *Relationship-map (F4, F5, F9)*
  The relationship-map is a 3-D map that indicates relationships between the

input keyword and other words. The words in the map can be selected by the mouse to specify the information need and the word where the user has clicked on will be used as query input for the next search task.

- *Directory (F3, F4, F5, F9, NF1)*
  The directory contains categories and words where the user can use the mouse to specify the information need.
- *Query reformulation (F5, NF3)*
  The system will aid the user in his search for information by expanding the query with words when necessary.
- *Results list (F7)*
  The search results will be presented as a list of search results, with a title, a short description and the URL of that website. It will also highlight the query in the short description.
- *Save search (F7)*
  The user can save his search results. With this option the user can divide a search session over multiple time periods.
- *Tab-pane (F6, F8, NF1)*
  The tab-pane makes it possible to navigate between the various search tasks. Every search task is presented in a tab. The user can follow the search process and go back to previous search results.
- *User safety (NF2, NF3)*
  There is a need for a filtering mechanism to block inappropriate content, banners and popups.
- *Time and date (NF2)*
  The website contains an indication of the the time and date.
- *Hardware and software (NF5)*
  The system must be able to run on Windows 2000 or higher in combination with Internet Explorer and Mozilla browsers.
- *Performance (NF6)*
  The system must have an adequate performance, which means among other things that the reaction time has to be as short as possible.

Secondary Requirements

The secondary requirements will be implemented when there are enough resources left to implement them. These requirements are nice to have, not must haves.

- *Document Format Types (NF5)*
  It is preferred that the search engine can give search results in more format types than HTML. Example format types are PDF, Word, Excel, and Powerpoint.
- *Relevance and Filtering (NF3)*
  It is important that a search engine gives search results that are relevant to the user.

- *Natural language query input (F4)*
  Besides short terms in queries it might be useful to have the ability to give as input to the search engine a natural language sentence.

- *Metaphor (F2)*
  A metaphor can aid a child in searching for information with the search engine. It can elevate the user experience of the child, or it will help the child with useful tips.

3 Information Retrieval Research

This chapter covers the research in the field of Information Retrieval. The chapter starts with a brief introduction on information searching and how this is modeled (section 3.2). Subsequently section 3.3 will describe the literature research on the aspects concerning Internet search engines. Section 3.4 describes our research on the use of formal concepts in reformulating user queries. Finally the global design on how a search engine, which covers the requirements, will look like is given. Section 3.5 will describe our model on information searching as a search process and section 3.6 will give the information on how the entire search engine is going to work.

3.1 Introduction

In chapter 1 different problems, questions and goals were identified concerning the field of Information Retrieval. This chapter covers the Computing Science part which is an important aspect of this thesis.

An important aspect within search engines is the retrieval of information. This is basically where Information Retrieval is all about. The Internet is full of information and there is a need to find relevant information. This process must both be efficient and effective. To achieve this goal, retrieval methods have to be developed that deliver a sufficient amount of documents of high quality at a high rate. This is also one of the major requirements mentioned in section 2.3. Another requirement concerns the fact that not all content on the web is appropriate for children, and that they have to be protected against that. To solve this problem, sufficient filtering methods have to be developed.

3.2 Information Searching

The Internet contains a lot of information that is interesting for children. Different researches have been performed on how children learn from the information available on the Web [16, 17, 35]. Children are not capable of using the Internet or search engines to find the information they need in the same way adults can [5, 78]. Therefore tools that help children in this must meet some requirements (chapter 2).

3.2.1 Educational Perspective

Learning begins at birth and continues throughout life. The learning process is very important for humans in their childhood phase. In this phase of human life, children

develop, prepare themselves and gain knowledge to survive in their future adult life in this complex society.

There is a lot of knowledge in this world that would be useful for children in their development. The amount of knowledge is so enormous and complex that children need to be guided in gaining this knowledge. For instance schools have the role in guiding children in gaining useful knowledge. From an educational perspective the Internet contains a lot of useful information for children, but it also contains unappropriate information. Children are not fully capable to judge the usefulness of information on the Internet, they need assistance in gaining this knowledge. This assistance can be shaped in the form of a search engine for children. So a search engine for children is not just a tool to get information from the Internet, it has an educational role in the life of children.

3.2.2 Learning Cycle

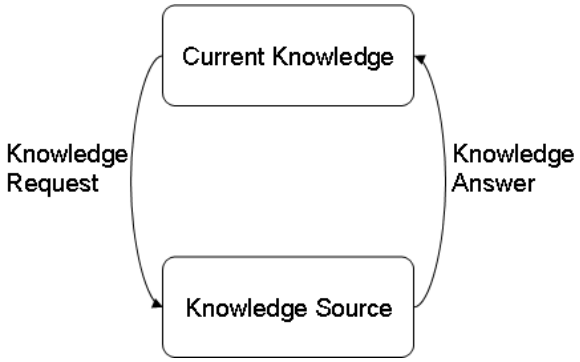


Figure 3.1: Learning Cycle Diagram

We suggest a simple model that describes the learning process (figure 3.1). This learning cycle consists of current knowledge and a knowledge source. The current knowledge represents the limited knowledge known by children. The knowledge source represents all knowledge that could be learned by children. Therefore the current knowledge is a subset of the knowledge source. The relation between these two components consists of a knowledge request, which represents the need for some knowledge, and a knowledge answer, which contains the desired knowledge. The current knowledge will increase each time the learning cycle is walked through. This process can continue as long as the knowledge source contains more information than the current knowledge part.

This model is applicable in any learning environment. For instance the knowledge source could be the books in a library, and the current knowledge would represent the knowledge of the visitor of the library. Library visitors need assistance in finding the needed knowledge, this assistance could be in the form of a library catalogue, or a librarian.

This model of the learning process is certainly applicable to the Internet. The Internet has become an important source of knowledge in the learning process. In this model the Internet stands for the knowledge source and the children represent the current knowledge. We see the search engine as an essential element of the learning process that is centered around the Internet. The knowledge request is the input to the

search engine and the knowledge answer is the output of the search engine. Further information on how the learning cycle can be seen as a search process is given in section 3.5

3.3 Literature Research

This section describes the literature research executed on several fields concerning Internet search engines. It gives a global view of the different retrieval and search methods that exist and are developed.

3.3.1 Global Introduction

Retrieval Relevance

The relevance of a document is the importance of a document to a particular user. The relevancy of a document depends on the information need of the user that is searching for information. Since the quality of pages varies widely, the relevance of pages is not enough. The goal is to return both high-relevance and high-quality pages [36].

As described in NF3 in section 2.3, this is not the only aspect that is important for the search result. The search engine is not allowed to return certain content. Therefore relevant documents are documents that are relevant to the user and do not contain the content as described in NF3.

Retrieval Performance

The performance of an information retrieval system is evaluated on three things: recall, precision and precision at top 10 result pages [36]. Recall is an indication of the coverage of the relevant documents that are retrieved. The precision is an indication of how precise your search was, a low precision means that you have retrieved many irrelevant documents. The precision at top 10 result pages is important because users often do not want to look at all the retrieved results and they want the information they need ranked as high as possible.

- **Recall:** The proportion of all relevant documents in the collection included in the retrieved set:

$$Recall = \frac{|relevant \bigcap retrieved|}{|relevant|}$$

- **Precision:** The proportion of the retrieved set that is relevant:

$$Precision = \frac{|relevant \bigcap retrieved|}{|retrieved|}$$

The goal of information retrieval is to return a set of documents that has a high recall and precision. However recall and precision are each others inverses. So the art is to get both the recall and precision has high as possible. Recall and precision are useful because they allow us to evaluate the quality of the retrieved set and the breadth of the retrieval algorithm quantitatively. [4, 78, 79]

3.3.2 Architecture

The global architecture of a search engine consists of three elements: a crawler, an indexer and a query server. The crawler gathers the pages from the Internet. The indexer processes the found documents and stores them in an efficient data structure. The query server receives the query from the user and returns the relevant result pages by consulting the search data structures. [4, 36]

Crawler

Web crawling can be done by one or more crawlers, Google for example uses several crawlers to collect its data [10]. A crawler should be robust and tested extensively, because a lot can go wrong when crawling the Web. The crawler should visit millions of websites and since every website is different in structure and content, everyone of them could be a thread to the performance of the crawler. Furthermore, the crawler should not overload the websites and network links it visits and it must efficiently deal with huge volumes of data.

One of the most important challenges for crawlers is the selection of URLs to scan from its queue of known URLs. Most crawlers will not be able to visit every possible page. This is because the Web contains a huge amount of data and it is almost impossible to store all that data, and at some point the crawler needs to start revisiting previously visited pages, to check for changes. Therefore it is important to capture the most important pages as early as possible during the crawling process. A solution is to develop a method of crawling that gives important pages a higher probability of being visited first. This way the visited part of the web is more meaningful. [12, 15]

Indexer

Documents consists of many words and phrases. Search engines usually compare the words and phrases in a document with the queries submitted by their users. Comparing query terms with all words of the available documents is not an option, because this would require to much time and resources. Therefore search engines index the documents they found on the Web. The indexer analyses every document and looks for terms that are important for that document. These terms are the information that is saved for that document. Query terms are now compared with that information instead of with all information available in the document. [4]

Query server

The query server is the portal for the user to the information collected by the crawler and the indexer. The user submits a query, which represents his information need, to the query server. The query server will then compare the query with the information available in the database containing the indexed documents. The query server selects the documents that correspond the most to the query and presents them to the user in an ordered list. It is important that the query server delivers documents of which the relevance is high, and that the retrieval performance(section 3.3.1) is sufficient. [4]

3.3.3 Classic Information Retrieval Models

The classic information retrieval models are based on the assumption that each document is described by a set of representative keywords called index terms. An index

term is a word that describes the document by its semantics [4]. This section will describe three classic information retrieval models: the boolean model, the vector model and user relevance feedback.

Boolean Model

The boolean model is a retrieval model based on set theory and boolean algebra. The boolean model is about the presence or absence of index terms in a document. A query consists of index terms linked by the connectives: not, and, or or. An example of a boolean query is: $(books \wedge kids) \vee comics$. This query means that we are searching for documents about books and children or for documents about comics, while the query $books \wedge (kids \vee comics)$ means that we are searching for documents that are either about books and kids or for documents about books and comics. With the boolean model a document is either relevant or non-relevant. There is no notion of a partial match to the query conditions. [4, 66]

Advantages The main advantages of the boolean model are the clean formalism behind the model and its simplicity.

Disadvantages There are various disadvantages concerning the boolean model:

- The exact matching may lead to retrieval of too few or too many documents.
- The output for a query is not ordered, so each retrieved document is assumed to be as important as any other retrieved document.
- All terms in a document are assumed to be of even importance.

Vector Model

The vector model represents documents and queries as vectors in an M-dimensional space. In this case M is the number of terms. With the vector model it is possible to compute the degree of similarity between documents and user queries. Retrieved documents are sorted in decreasing order of this degree of similarity, documents that match the query terms only partially are also taken into consideration. This ordered list of retrieved documents is a lot more precise than the set retrieved by the boolean model. [4]

Advantages

- The term-weighting scheme improves retrieval performance.
- Documents that approximate the query conditions can be retrieved.
- The documents are sorted according to their degree of similarity to the query.

Disadvantages

- Index terms are assumed to be mutually independent. However, in practice, consideration of term dependencies might be a disadvantage.

User relevance feedback

User relevance feedback is an interactive process of obtaining information from the user about the relevance and the non-relevance of retrieved documents. The user is presented with a list of the retrieved documents and, after examining them, marks those that are relevant. The main idea is to select important terms, attached to the documents that have been identified as relevant by the user, and enhance the importance of these terms in a new query formulation. The expected effect is that the new query will be moved towards the relevant documents and away form the non-relevant ones. Examples of user relevance feedback are the Smart system [65] and the probabilistic model [4, 63].

3.3.4 Meta search

In a meta-search engine, you submit keywords in its search box, and it transmits your search simultaneously to several individual search engines and their databases of web pages. Within a few seconds, you get back results from all the search engines queried. Meta-search engines do not own a database of Web pages; they send your search terms to the databases maintained by search engine companies.

A recent study [34] showed that the size of the indexable Web is estimated to be more than 11.5 billion pages. It is almost impossible for a search engine to cover all information available [18, 33]. In fact the coverage of many search engines is decreasing steadily due to the fact that the Web has been increasing at a much faster rate than the indexing capability of any single search engine [45]. On the other hand each search engine covers a different part of the Web [44, 70]. The intersection of the coverages of two search engines is on average only about 55% [34]. By combining the coverages of multiple search engines, a much higher percentage of the Web can be searched [50]. Therefore it is possible to achieve much higher recall with meta-search engines. Another advantage of meta-search engines is that they don't need their own crawler or indexing mechanism, so a meta-search engine can be built with much less resources.

We have limited resources in time, bandwidth, computational power and disk space. These are all important factors in developing and executing a powerful search engine as described in sections 3.3.1 to 3.3.3. Furthermore we want our search engine to have a high performance as described in section 3.3.1. This goal can be reach with meta-searching by choosing the right search engines for high relevance, and high recall is satisfied by combining these search engines. The search engine developed in this project will therefore be a meta-search engine.

3.3.5 Ranking

The ranking mechanism of a search engine orders the search results by importance and/or relevance. Ranking is an important part of a search engine, because users in general only look at the first result page [40, 71]. Therefore a search engine have to make sure that the most relevant pages are ranked as high as possible.

3.4 Query Reformulation

In this section our research on query reformulation is discussed. This research is based on the automatic query expansion by formal concepts. First the theory of

formal concepts will be discussed (section 3.4.2), then section 3.4.3 will discuss our approach using these concepts to reformulate user queries.

3.4.1 Introduction

In section 2.2 the problem of formulating an information need and reformulating a query is addressed. These problems are caused by different aspects: the inherent ambiguity of words in natural language, the difference of interpretation for a query [13] and the fact that many users do not have more than a visceral information need [76]. These are all problems experienced by adults. The problems that children have in formulating and reformulating their queries are obviously bigger. Therefore a tool to aid them in this is very needed.

A method to overcome or reduce these problems is automatic query expansion. With automatic query expansion new words are added to the user's query. The idea is that the expanded query is a better representation of the user's information need than the original query. A lot of researchers have focused their work on this topic[51, 83] and it shows that this method really improves the retrieval effectiveness for short queries, which is the most common type of queries used.

In recent years interesting research has been done on the use of global concepts for query reformulation [31, 32]. This theory will be used in this research to develop a method to reformulate queries used by children in our Internet search engine for children.

3.4.2 Formal Concept Analysis

In this section the theory behind formal concepts will be discussed. Formal Concept Analysis is based on mathematical order theory, in particular on the theory of complete lattices. The basic notions of Formal Concept Analysis are those of a *formal context* and a *formal concept*. These notions are explained in the first parts of this section. The section is concluded with the definition of the concept lattice.

Formal Contexts

A formal context (G, M, I) consists of two sets G and M and a relation I between them. The elements of G are called objects and the elements of M are called attributes of the context. A formal context relates objects to attributes. If an object $g \in G$ has an attribute $m \in M$ then $(g, m) \in I$ [24, 31, 79].

The attributes that belong to all objects in a set are called the common attributes of that set of objects. The common attributes of a set O of objects express the meaning of this collection O in terms of attributes they share. The function to find the common attributes is defined by:

$$ComAttr(O) = \{m \in M \mid \forall_{g \in O}\ (g, m) \in I\}\ with\ O \subseteq G$$

The same counts the other way around. The objects that own all attributes in a set are called the common objects of that set of attributes. The common objects of a set A of attributes express the meaning of A as a set of objects. The function to find the common objects is defined by:

$$ComObj(A) = \{g \in G \mid \forall_{a \in A}\ (g, a) \in I\}\ with\ A \subseteq M$$

Objects \ Attributes	non-alcoholic	hot	alcoholic	caffeine	sparkling
Tea	X	X			
Coffee	X	X		X	
Mineral Water	X				X
Wine			X		
Beer			X		X
Cola	X			X	X
Champagne			X		X

Table 3.1: Context of several drinks with their characteristics [79]

Example 1 Table 3.1 represents a context between several drinks and their characteristics. In this example $G = \{Tea, Coffee, MineralWater, Wine, Beer, Cola, Champagne\}$ and $M = \{non-alcoholic, hot, alcoholic, caffeine, sparkling\}$. An example of a relation between an object and an attribute is $(Cola, caffeine) \in I$.

The only attribute that Wine and Beer have in common is alcoholic, so $ComAttr(\{Wine, Beer\}) = \{alcoholic\}$. The common objects of alcoholic however are Wine, Beer and Champagne, $ComObj(\{alcoholic\}) = \{Wine, Beer, Champagne\}$

Formal Concepts

A formal concept is a pair (D, A) with mutual assignment of meaning. D is the set of all objects having all the attributes of A, and A is the set of all attributes that are valid for all the objects of D:

$$ComAttr(D) = A$$

$$ComObj(A) = D$$

The set D is called the *extent* and the set A is called the *intent* of the concept (D, A). The extent of a concept determines the intent and the intent determines the extent. So a concept is uniquely identified by its set of documents or its set of attributes. [32, 81]

Example 2 We use the context of table 3.1 to show some examples of concepts. An example of a concept here is $(\{Tea, Coffee\}, \{non-alcoholic, hot\})$, because:

- $ComAttr(\{Tea, Coffee\}) = \{non-alcoholic, hot\}$
- $ComObj(\{non-alcoholic, hot\}) = \{Tea, Coffee\}$

The extent of this concept is $\{Tea, Coffee\}$ and the intent of this concept is $\{non-alcoholic, hot\}$. You can find all concepts for this example in table 3.2

concept	objects	attributes
c_1	Tea, Coffee	non-alcoholic, hot
c_2	Coffee	non-alcoholic, hot, caffeine
c_3	Mineral Water, Cola	non-alcoholic, sparkling
c_4	Wine, Beer, Champagne	alcoholic
c_5	Beer, Champagne	alcoholic, sparkling
c_6	Cola	non-alcoholic, caffeine, sparkling
c_7	Tea, Coffee, Mineral Water, Cola	non-alcoholic
c_8	Coffee, Cola	non-alcoholic, caffeine
c_9	Mineral Water, Beer, Cola, Champagne	sparkling
c_{10}	all objects	no attributes
c_{11}	no objects	all attributes

Table 3.2: Concepts for the drinks context

To find all concepts of a context, you first have to find the base concepts. Then the base concepts can be combined to form new concepts. The base concept of an object d or an attribute a is defined by:

$$BaseConcept(d) = (ComObj(ComAttr(\{d\})), ComAttr(\{d\}))$$

$$BaseConcept(a) = (ComObj(\{a\}), ComAttr(ComObj(\{a\})))$$

To combine two concepts you take the intersection of the extensions or the intersection of the intentions. You can combine the new found concepts until you have found all concepts that belong to the context. [24, 31, 79]

Example 3 In this example the construction of base concepts for objects and attributes is shown.

Base concept of Beer:

- $ComAttr(\{Beer\}) = \{alcoholic, sparkling\}$
- $ComObj(\{alcoholic, sparkling\}) = \{Beer, Champagne\}$

So the base concept for Beer is $(\{Beer, Champagne\}, \{alcoholic, sparkling\})$

Base concept of alcoholic:

- $ComObj(\{alcoholic\}) = \{Wine, Beer, Champagne\}$
- $ComAttr(\{Wine, Beer, Champagne\}) = \{alcoholic\}$

So the base concept of alcoholic is $(\{Wine, Beer, Champagne\}, \{alcoholic\})$

Concept Lattices

Concepts may be ordered according to their extensionality:

$$c_1 \leq c_2 \Leftrightarrow ext(c_1) \subseteq ext(c_2)$$

$$c_1 \leq c_2 \Leftrightarrow int(c_1) \supseteq int(c_2)$$

The set of all concepts of (G, M, I) ordered extensional is called the concept lattice of the context (G, M, I). The context in Example 1 has 11 concepts. The line diagram in figure 3.2 represents the concept lattice of this concept. [24]

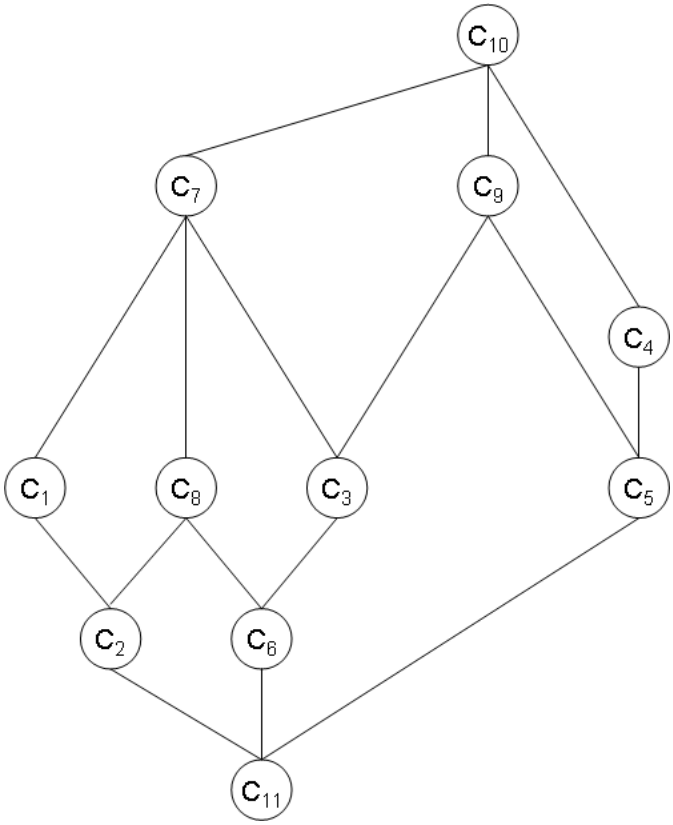


Figure 3.2: Concept Lattice for the drinks context

Example 4 Let's take a look at the ordering of the concepts c_2 and c_8:

- $\{Coffee\} \subseteq \{Coffee, Cola\}$
- so: $(\{Coffee\}, \{non{-}alcoholic, hot, caffeine\}) \leq (\{Coffee, Cola\}, \{caffeine\})$

In figure 3.2 you can see the ordering of all the concepts with concept c_{11} with no objects at the bottom and concept c_{10} with all objects at the top.

3.4.3 Conceptual Query Expansion on Child Queries

In this section we explain how the theory of formal concept analysis(section 3.4.2) can be used in a Web document context. Also our ideas and approach on using formal concepts for query expansion on child queries is discussed.

Formal Contexts in Web Documents

The formal contexts in formal concept analysis consist of objects, attributes and the relations between them. In a Web document context the contexts (D, A, I) consist of a set D of documents on the Web, a set A of those documents' relevant terms (still called attributes) and the relation I that says that a document $d \in D$ is about an attribute $a \in A$.

The formulas of section 3.4.2 still hold in this case. The common attributes of a set D of documents express the meaning of this collection D in terms of attributes

they share and the common documents of a set A of attributes express the meaning of A as a set of documents:

$$ComAttr(D) = \{a \in A \mid \forall_{d \in D}\ (d, a) \in I\}$$

$$ComDocs(A) = \{d \in D \mid \forall_{a \in A}\ (d, a) \in I\}$$

Formal Concepts in Web Documents

Formal concepts for Web documents can be build in the same way as described in section 3.4.2. A base concept can be calculated out of a document d or an attribute a:

$$BaseConcept(d) = (ComDocs(ComAttr(\{d\})), ComAttr(\{d\}))$$

$$BaseConcept(a) = (ComDocs(\{a\}), ComAttr(ComDocs(\{a\})))$$

The remaining concepts can be calculated by joining two concepts. A concept higher in the concept lattice can be calculated by taking the intersection of the sets of attributes of both concepts. Let $D_1, D_2 \subseteq D$ and $A_1, A_2 \subseteq A$. A concept (D_1, A_1) can be joined with a concept (D_2, A_2) as described above. The resulting concept is calculated as follows:

$$(ComDocs(ComAttr(D_1 \cup D_2)), A_1 \cap A_2)$$

A concept lower in the concept lattice can be found similarly, but then by intersecting the sets of documents of both concepts. The same two concepts (D_1, A_1) and (D_2, A_2) can be joined as follows:

$$(D_1 \cap D_2, ComAttr(ComDocs(A_1 \cup A_2)))$$

When you keep on joining concepts until no new concepts can be found, you find the complete concept lattice. [31]

	games	computers	Internet	animals
d_1				X
d_2	X	X		X
d_3		X	X	
d_4	X	X	X	

Table 3.3: Example Context

Conceptual Query Expansion

Formal concepts can be used to expand short queries that are submitted by users of search engines. The concepts constructed represent two things: the documents that are alike and the attributes that are commonly found together in documents. This latter fact can be used for query expansion.

concept	objects	attributes
c_1	d_1,d_2	animals
c_2	d_2	games, computers, animals
c_3	d_3,d_4	computers, Internet
c_4	d_4	games, computers, Internet
c_5	d_2,d_4	games, computers
c_6	d_2,d_3,d_4	computers
c_7	D	$\emptyset$
c_8	$\emptyset$	A

Table 3.4: Example Concepts

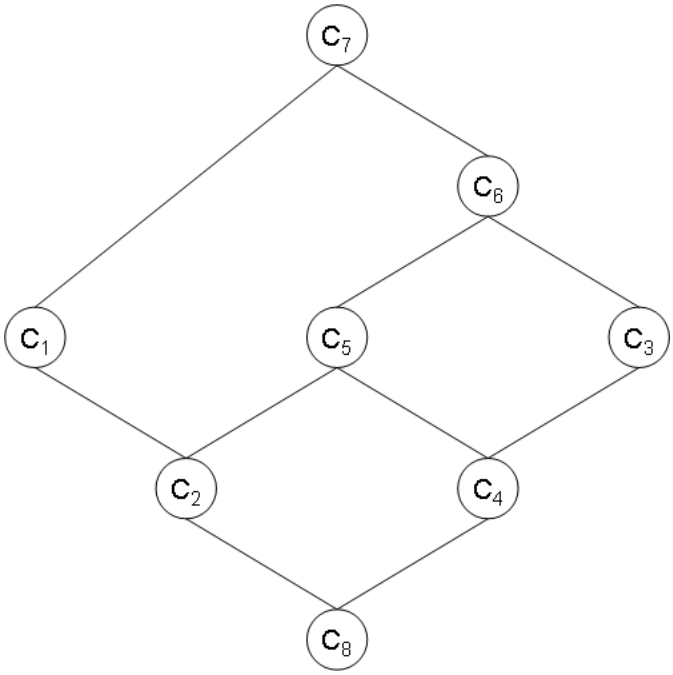


Figure 3.3: Concept Lattice

Building a concept lattice for all documents on the web is a very expensive and time consuming activity. But finding an alternate term to expand a query can be done by just building a sublattice. When a user submits a short query, the base concept for each attribute in the query can be calculated. Subsequently all concepts are joined until no new concepts can be found. Then you have constructed the complete sublattice. In the top node of the sublattice you can find attributes that are candidates for query expansion. If the top node does not contain any new attributes, then the query can't be expanded. If the query is expanded this way, the query is probably a better representation of the user's information need and thus the search result will better fulfill the user's information need. [31]

Example 5 Take the context of table 3.3 as the complete information space. All concepts of this context are listed in table 3.4. The complete concept lattice of this context will look like figure 3.3.

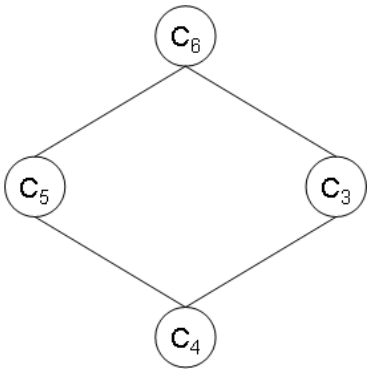


Figure 3.4: Sublattice of figure 3.3

A user submits a query containing the terms $games$ and $Internet$. The base concepts for these terms are c_5 and c_3 respectively. Round these two concepts a sublattice can be build, this is shown in figure 3.4. The top node, representing concept c_6, of this sublattice contains the attribute $computers$. Probably $computers$ is a good expansion on the user query, because documents about $games$ and documents about $Internet$ commonly contain the word $computers$.

Conceptual Query Expansion in a Search Engine for Children

Most documents on the Web are not interesting for children. Therefore the concepts build for all documents on the Web will not correspond with the concepts concerning documents and attributes interesting for children. A solution is to construct only concepts containing documents interesting for children. This way queries can be formulated that enhance the probability of finding documents interesting for children. This method is extremely useful when for example homographs (words spelled the same, but with a different meaning) are used. A good example is the word *settlers*. This query could deliver all sorts of boring documents about settling, but all the user could want was information about the game *The Settlers of Catan*. With formal concept analysis could be encountered that all child documents with the term *settlers* also contain the term *catan*, so the query could be expanded to *"settlers, catan"*. The query result will then better represent the user's information need.

The problem is that we don't have all documents interesting for children. If we would have all documents interesting for children, then a search engine for kids wouldn't be that much of a problem. Besides the fact that the content on the Web is constantly changing, the interests of children are also constantly changing. Therefore we have to develop a method that only builds concepts from child documents and that keeps the concepts up to date to the interest of the day.

Our suggestion is to build the concepts from the search results of the children. We assume that the top documents found by children, during a search session, are documents that are relevant and interesting for children. So the terms used in these documents are related and meaningful to children. We use the set of all top relevant documents found by children for the calculation of concepts over these documents and their corresponding attributes. These concepts can be used to expand user queries in the way described above. The collection of documents will be updated with every

query submitted. This way the concepts will be up to date with the users' interests.

A disadvantage of this method is that the query expansion will only work when enough queries are submitted to the search engine. When only a few users have used the search engine, the context of documents interesting for children is not big enough to calculate a set of concepts that represents the real information space. Therefore the query expansion method can only be initiated when a particular threshold of unique user queries is submitted.

There is a solution to overcome this problem. We could initiate the search engine with a large set of documents interesting and appropriate for children. This way there is enough data available to build a reasonable set of concepts. An example of a data source where these documents could be found could be Yahooligans (http://yahooligans.yahoo.com). This search engine has a very large directory structure containing a great amount of websites. All these websites could be indexed. An advantage of this method is the great diversity of websites for children, which are both appropriate and interesting for children. A disadvantage is the volume of the set of documents to calculate concepts. The information space for selecting expansion options is very static. A solution is to combine this initial set of documents with the method of using the top relevant documents to calculate concepts. This way the query expansion module could be operational from the beginning and the concept space is up to date with the current interests of children.

3.5 Search Process

This section describes how the learning cycle described in section 3.2.2 and shown in figure 3.1 can be seen as a search process. The learning cycle consisted of Current Knowledge (CK) and a Knowledge Source (KS). These two components of the learning cycle will be described in respectively section 3.5.1 and 3.5.2.

3.5.1 Current Knowledge

The CK can be seen as a user describing it's information need to increase his current knowledge. The user uses a user interface which helps him formulating this information need. The output of this user interface is a query describing the user's information need. This query is submitted to the KS-component. A visual representation of the process is shown in figure 3.5.

3.5.2 Knowledge Source

The KS can be seen as a module fulfilling the user's information need based on the query provided by the CK-component. The KS-component globally consists of three components:

The Query Reformulation module based on the theory of section 3.4. This module enhances the user's query to better represent the user's information need.

The Document Gathering component which really contains and delivers the knowledge desired by the user.

The Presentation module which transforms the information into a form readable for the user and returns it to the user.

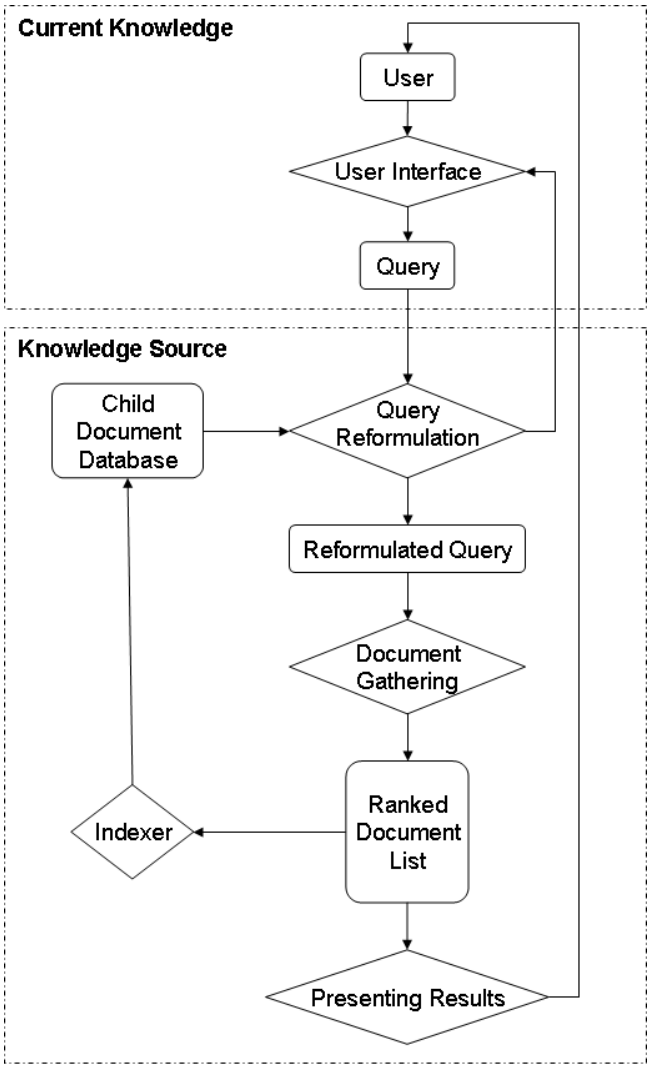


Figure 3.5: Search Process Diagram

The remaining components of the KS-component can be considered as part of the Query Reformulation module. Detailed information on all components of the search process of figure 3.5 will be given in the next section.

3.6 Specified Search Process

This section gives a description on how the entire search engine is going to work. Section 3.6.1 describes the overall structure of the search engine. Subsequently, the sections 3.6.2 to 3.6.7 will describe the different components of the search engine.

3.6.1 Introduction

The overall structure of the search process is shown in figure 3.6. The search process starts with the user formulating his information need. The formulation of the user's information need results in a query. This process is described in section 3.6.2.

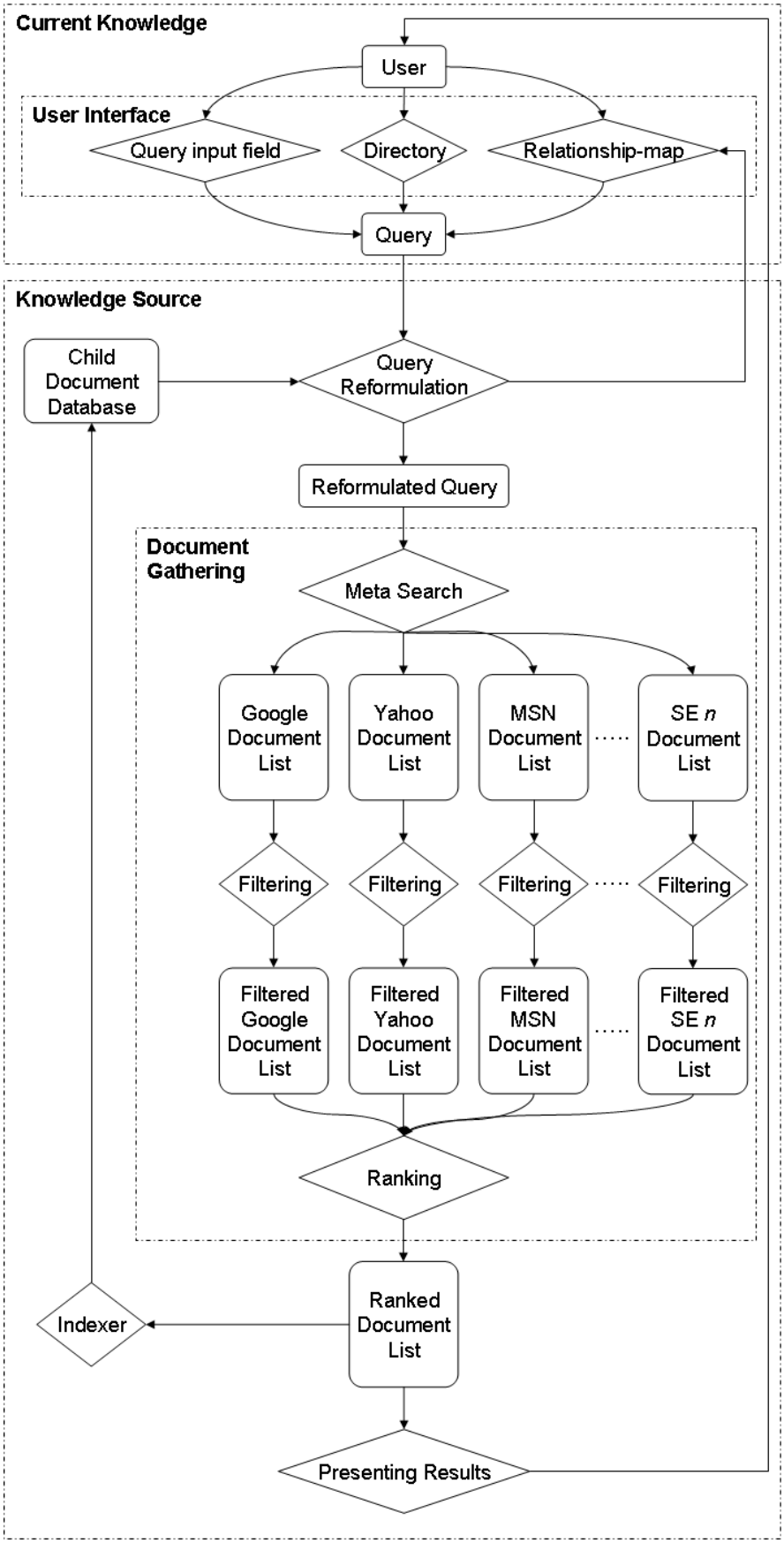


Figure 3.6: Specified Search Process Diagram

The query formulated by the user is then reformulated by the query reformulation module. More information about this module can be found in section 3.6.3. The new query that is calculated by this module is used for the gathering of documents.

The document gathering part described in section 3.5.2 is subdivided in three parts: meta search for retrieval of documents, filtering for restriction on documents and ranking for ordering of documents.

The reformulated query is used in the meta search process to search for relevant documents with other search engines. This will be described in section 3.6.4. The meta search process will deliver a list of unfiltered documents for each search engine it uses.

It is important to filter the lists of documents derived from the different search engines on documents not meant for children. The filtering method will therefore deliver a filtered list of documents for each used search engine containing only documents meant for children. More about this method can be found in section 3.6.5.

All documents found by the different search engines have to be ordered by relevance for the user. This is done in the ranking module (section 3.6.6). The ranking module constructs an ordered list of documents which can be presented to the user.

The presentation of the results is of crucial importance for the user friendliness of the search engine. A description of the criteria and ideas to enhance the user friendliness in this search engine are described in section 3.6.7.

3.6.2 User Interface

The user has three ways of formulating his query: by typing keywords in the query input field, by selecting subjects in the directory structure or by the relationship-map. All these actions result in the formulation of a query used in the search process.

Query Input Field

The query input field can be used by the user to formulate his information need. The information submitted to the input field consists of one or more keywords that are, according to the user, a good representation of that information need. The query formulated by the user is submitted to the query reformulation module.

Directory

The high-level intuitive directory allows the user to specify his information need by selecting hierarchical subjects. Since formulating your information need is not an easy task, especially for children, this is a good method to help them with that. [78]

Relationship-map

A relationship-map is a map indicating relationships (synonyms, hypernyms, etc.) between the input keyword and other words in documents predicted to be relevant. Words in the map can be selected by the mouse to (further) specify the user's information need [78]. The word selected is submitted to the query reformulation module. New keywords and relationships are displayed in the map on every submitted query by either the relationship-map itself or the directory and input field.

The relationship-map enhances the understanding of the user towards the search process. By providing the relation of the query term with other keywords, the user

will better understand why some documents appear in the results. This will aid the user in reformulating his query.

3.6.3 Query Reformulation

The Query Reformulation module is the first part of the KS-component. The theory of formal concepts and concept lattices is used for the reformulation of queries. This method can add keywords to a user query to enhance the query in such a way that it becomes a better representation of the user's information need. This method was discussed in section 3.4. The documents and corresponding attributes required for this method can be found in the Child Document Database. This database consists of documents interesting and appropriate for children. The database is updated every cycle of the search process by adding the top relevant document, found in this process, to the database.

Indexer

To make the documents suitable for formal concept analysis they have to be indexed first. All documents that are submitted to the database are analyzed by the indexer. The indexer scans the document for terms relevant for that document. These terms are saved in the database as attributes that have a relation with the document.

3.6.4 Meta Search

As already described in section 3.3.4, meta-search engines forward search queries to other search engines and combine their results. That is exactly what the meta search process in this search engine does. It uses the most used search engines. At this moment the most used search engines are Google and Yahoo [73]. Therefore at least these two search engines are used by the meta search process. It is not preferable to use the web interface of these search engines and then parse the result pages. This would take a lot of time, and when one search engine would decide to change his interface, the application would not work anymore. Fortunately, search APIs are provided by Google [29, 49, 53] and Yahoo [87]. MSN also has a search API [52], so that would be good third option. Using these APIs the application can access the results directly and can save valuable time and resources.

3.6.5 Filtering

The quality of the search results is of great importance for a search engine for children. Many documents on the Web should not be shown to children by the search engine. The requirements on which documents should and which should not be shown were discussed in section 2.3. The search engines used in for meta-searching each have filters to remove offensive content from their results. Google and Yahoo for example offer the possibility to turn on SafeSearch [27, 85]. There are some drawbacks for these methods [20], but in general it works well. Nevertheless the system will still filter the results on offensive content that might have slipped through the SafeSearch filters.

Offensive content is not the only aspect which needs filtering. Many documents on the Web do not contain any content not suitable for children, but they are still not interesting for children. Documents with high reading levels or documents which lay-out is not interesting(text-only) fall under this category.

3.6.6 Ranking

The results derived from the different search engines have to be combined into a single list of documents that is ordered by relevance and importance. To achieve this we make use of two assumptions:

- The results derived from a single search engine are already ordered by relevance and importance.
- Documents found by more than one search engine are more relevant and important than documents found by only one search engine, so the more search engines find a document the more relevant an important it is.

The features of these two assumptions are combined as follows: first the entire list of documents is divided and ranked by the number of search engines they are found with. Subsequently the groups are ranked by the rankings in the search engine results. If the document appears in more than one search engine the rankings from the different lists are added, and the document with the lowest sum ends the highest.

For example, consider the following three lists as being the results of three search engines:

1. tea	1. cola	1. wine
2. coffee	2. wine	2. milk
3. cola	3. water	3. water
4. beer	4. tea	4. cola
5. champagne	5. coffee	5. fanta

This would then result in the following ranked list:

	# SEs	Score
1. cola	3	8
2. wine	2	3
3. tea	2	5
4. water	2	6
5. coffee	2	7
6. milk	1	2
7. beer	1	4
8. champagne	1	5
9. fanta	1	5

3.6.7 Presenting Results

This is the last part of the KS-component. The user must have a good understanding of the results presented to him. If the presentation of the results is not clear to the user, he can not classify the resulting documents as relevant or non-relevant. As already addressed in the requirements in section 2.2 as much information as possible should be shown about a particular document. This information includes the document title, a passage selection showing in which context the search terms are used in the document, highlighting of those search terms and the URL where the document was found. This information helps the user understand and predict the content associated with the link. Furthermore, a new relationship map centered around the search

term will be shown. This will aid the user in understanding the relation of his search term to other keywords, which probably appear in one or more of the documents in the result.

Other requirements addressed in section 2.2 are the search history and the reversal of actions. Since searching is an iterative process it is an advantage, for orientation purpose, to give an indication of the path that has been taken to get to the current situation. It is also important to include an ”undo” function to return to a previous state on that path. This is done by showing the current and previous search keywords in tabs. By clicking on a tab you return to that search session. This way it is easy to see which search actions have been taken and to return to such a search action.

3.7 Conclusion

In this chapter the information retrieval aspects of this project were addressed. The chapter began with a global introduction on information searching in which the learning cycle was introduced. This learning cycle represents the main idea behind information searching and search engines, and is especially applicable on children.

The main part of this chapter was about finding and developing technologies for a search engine for children that covers the learning cycle. General information about search engines was given as well as some interesting technologies that will be used in the search engine, which we will develop, such as meta search. A very interesting part of this chapter was the use of formal concepts for query reformulation. This newly developed method reformulates queries from users in a rather intelligent way such that the query will better represent the users information need.

Finally, the learning cycle was represented as a search process. This search process is the basis for the search engine, which will be build. A thorough description of this search engine was given in the last part of this chapter.

4 Business Research

The business research on Internet search engines will be discussed in this chapter. The research objectives and the research process are discussed in the introduction (section 4.1). Following the introduction will be the literature research and the survey. These are respectively discussed in sections 4.2 and 4.3. This chapter will conclude with the results (section 4.4) and the conclusion (section 4.5).

4.1 Introduction

With the public introduction of the World Wide Web in the 1990s there emerged a need for a portal that will help the user to find information stored on this medium. The idea of a search engine already existed before the launch of the World Wide Web. The first search engine created was Archie, a search engine designed to index FTP archives. FTP was the main way to share data back in those days. The first Web search engine was Wandex, and after Wandex other search engines were launched such as Aliweb and WebCrawler. The competition in the search engine business began in 1994 with the entry of Yahoo and Lycos. Since then a lot of search engines were founded that competed with each other. The search engine business grew spectacularly with many search engine offerings and many users. Until, mid-1998, the search engine business was dominated by Excite, Infoseek, Altavista, Lycos and Yahoo. After that some new entrants came into the business gaining a significant business share by offering high quality products (relevant search results) [23]. Nowadays the business is dominated by three big players: Yahoo, Google and Microsoft (MSN Search) [75]. Together with these three big players there is still room for smaller search engines in this strong competing business.

In the beginning of the search engine business there were companies behind search engines like Infoseek that thought that people were willing to pay for their services. They underestimated how quickly competitors would provide these services for free, this development had forced them to follow these websites and to find other ways of generating money [42]. There are still companies that sell searching services, for instance Northern Light has taken down their public search engine and are now marketing their Enterprise-only editions [80]. Nowadays most of the search engines are public search engines and are freely accessible for everyone.

As most search engines are offered for free, there are of course costs to keep these search engines operational. Companies behind these search engines don't offer them without making money out of it. Like any other businesses they want to gain a profit out of their business. So where does the money come from for keeping these search

engines operational? How do these search engines generate revenue?

4.1.1 Research Objectives

The objective of this business research is to define a method for search engines for children to earn money and to sustain themselves. We want to achieve this by identifying the characteristics of the Internet search engine business. This is done by investigating how current search engines on the World Wide Web earn money.

4.1.2 Research Setup

The research will be a survey of Internet search engines. In this research setting a survey means a detailed critical inspection [82]. The first part of this research is to find scientific literature on aspects of the methods by which Internet search engines (the companies behind these search engines) earn their money. The expectation is that scientific literature on this specific topic is scarce.

The main part of this research will be a field survey on Internet search engines. Search engines (information about search engines) are mostly easy and freely accessible, this makes it very economical to get the data for the survey. There are a lot of search engines available on the World Wide Web, which can be examined. Based on various characteristics we have distinguished a number of categories (taxonomy) for search engines. The setup of the survey is globally as follows:

- Searching for Internet Search Engines
- Classifying the Internet Search Engines according to their category
- Examine the Internet Search Engines on methods of earning money

4.2 Literature Research

In this section we will discuss the literature on how Internet search engines earn their money. Scientific literature on search engine is very scarce, so we have used literature that covers electronic businesses, because search engines are a class of electronic businesses.

We are basically trying to find information in the literature on how search engines and other search engine related Internet ventures earn their money. A description of how Internet ventures earn money is often denoted in the electronic business literature as 'business models'. Business models tell in short what the company will do to achieve its profitability objective. Business models will be discussed in this section together with the market for Internet search engines, in other words, the information market.

4.2.1 The Business Model Concept

A very important aspect of entrepreneurship is generating revenue and profits. Senior staff within corporations will most likely ask themselves and each other: "How are we going to make money with our business?". A term that is often used to answer the previous question is 'business model'. This term is coming from the discipline of entrepreneurship and management. In other words the business model concept

is used in this discipline to describe how a company makes money and can sustain itself.

The digitalization of the world has caused major changes in doing business in many industries. An example is the music industry. Before the digitalization it was only possible to obtain and distribute music by buying L.P's and C.D's, which could only be made by manufacturers. The digitalization has made it possible for consumers to obtain and distribute music through other ways, like the Internet where consumers can download the music for free illegally. So obtaining music no longer needs to go through a record company. The digitalization has radically changed the way of doing business in the music industry, it has forced the music industry to develop new methods of earning money. In other words they had to develop new business models.

The same applies for the information market. Before the digitalization we were dependent on publishers for access to information. The Internet has resulted in less dependency on traditional publishers because everyone can publish and access information on the Internet. The diverse and huge amount of information that is available on the Internet has created a business opportunity for search engines to help users and publishers in finding each other. (See section 4.2.6)

The Internet is a product of the digitalization. The Internet has created new business opportunities for entrepreneurs (e.g. the search engine business), which need new business models. Lee and Vonortas [46] see e-commerce as a 'disruptive' innovation. According to them it is disruptive to the traditional way of doing business in that it is transforming the rules of competition and inventing new value propositions and business models. Lee and Vonortas have discussed in their article several disruptive attributes that are brought by the introduction of e-commerce. The disruptive attributes that are relevant to the search engine business are [46]:

- *Open platform.* Internet provides an open and nonproprietary platform for communication and collaboration. The open source movement in software development [59] has contributed to Internet-enabled collaboration and free information sharing.
- *Exchange and sharing of Information.* In the digital economy, the traditional trade-off between richness and reach in information exchange no longer exists [22]. In e-commerce, information can reach many customers or business ecosystem partners through the Internet without sacrificing the richness of the contents.
- *Digital assets as the input into the business transformation process.* A firm that exploits the Internet should build and utilize its digital assets [61], which is all the information about its customers, in order to provide value across many different and disparate markets. In the digital economy, information is a source of revenue and every business is an information business [19]. Therefore, a firm should use information to create new businesses and/or reinvent customer relationships through the implementation of virtual value chain [61].
- *Speed and frequency of changes.* Change is fast and frequent in the digital economy. Firms in every industry must learn to adapt quickly to changing business and economic environments. Adaptation in a turbulent environment means watching for the next wave

> that is coming, figuring out what shape it will take, and positioning the company to take advantage of it [1].

The above described attributes of the Internet have invented new business opportunities like the search engine business. A new business results in new business models.

In the next sections we will discuss the aspects a company has to consider when it wants to earn money. We will discuss the business models in the management and electronic commerce literature.

4.2.2 Business Models in Management Literature

The term 'business model' is often used these days, but it is seldom defined explicitly. This term is widely discussed by practitioners and investors of e-commerce, but not yet prominent in academic literature. Chesbrough and Rosenbloom [14] have searched on the Internet for references to 'business model'. They have found 107,000 references with Google in May 2000 (we have found 46,800,000 references with Google in March 2006). The academic literature, on the contrary, remains slow to reflect new terminology from practice. A search of a database of academic journals in economics found only three citations for the term 'business model'. Chesbrough and Rosenbloom think that a reason why academic scholarship has not focused on the concept may be that it draws from and integrates a variety of academic and functional disciplines, gaining prominence in none.

Chesbrough and Rosenbloom [14] depict the business model as a provider of a coherent framework that takes technological characteristics and potentials as input, and converts them through customer and markets into economical outputs. Thus they present the business model as a construct that mediates between technological inputs and economic outputs. In their view technological change requires new business models. The failures of firms to manage effectively technological change can be understood as the difficulty these firms have in perceiving and then enacting new business models. They have presented six functions of a business model [14]:

- articulate the value proposition
- identify a market segment
- define the structure of the value chain
- estimate cost structure and profit potential
- describe the position of the firm within the network linking suppliers, customers, complementors, and competitors
- formulate a competitive strategy

Eisenmann [21] sees the business model as a hypothesis about how a company will make money over the long term. Questions that a business model would answer are:

- What will the company sell, and to whom?
- How will the company collect revenue?

- What technologies will be employed?
- When will the company rely on partners?
- How will the cost scale with growth?

It appears that all definitions include a description of how a company makes money and can sustain itself by providing more value to its clients than the competitors.

4.2.3 Internet Business Models

The objective of every venture is to be successful. What success means to each venture can be different. So what should a company do to be successful? The preceding sections have introduced and explained the business model concept. This concept acts as a model to aid the entrepreneur in making the venture a success. Every business needs a business model. As Rayport has written in his article[60]: "In the end, e-commerce is just, when all is said and done, another kind of business."

Researchers of e-commerce have came up with several definitions of the Internet business model. "Business Models for Electronic Commerce" of Timmers [77] is the article that started the discussion of business models in electronic commerce. Timmers had noticed that the term 'business model' was not used consistently in the literature on electronic commerce. Authors who used this term often did not give a definition of the term 'business model'. Timmers gives the following definition of the term **'business model'**:

- An architecture for the product, service and information flows, including a description of the various business actors and their roles; and
- A description of the potential benefits for the various business actors; and
- A description of the sources of revenues.

There are many depictions of the variety of Internet business models. In the end the essence of the idea is 'how you get paid', or 'how you make money', with a taxonomy of alternative mechanisms [14].

Most of the writings of business models for the Internet focus on producing taxonomies or categorizations, or on stating what business models include or exclude. Timmers offers a taxonomy of 11 business models for e-commerce. Rappa [58] has extended that taxonomy of Timmers. Rappa has identified 29 different types of business models, in 9 categories. The taxonomies of Timmers and Rappa are based on the Internet businesses that are observable on the Internet. All observable Internet businesses can be classified into one of their categories. There are a few Internet business models of Timmers and Rappa that are relevant to the search engine business. We will discuss them in the next two sections.

4.2.4 The Internet business models of Timmers

Timmers has a taxonomy of eleven business models. These business models are listed and briefly discussed in table 4.1.

Ten business models of Timmers are not relevant to the search engine business. E-shop, e-procurement, e-auction and e-mall are concerned with selling and buying

E-shop	This model stands for an Internet expansion of a traditional company or shop.
E-procurement	Companies can seek suppliers through the Internet.
E-auction	Auctions of products on the Internet.
E-mall	A collection of e-shops.
3rd party marketplace	Outsourcing the web marketing to a third party.
Virtual communities	Users share information to a virtual community company. For instance at Amazon.com buyers give book reviews.
Value chain service provider	Support a part of an electronic commerce value chain. For example electronic payment through a credit card.
Value chain integrator	Integrating multiple steps of an electronic commerce value chain.
Collaboration platforms	Providing a set of tools and an information environment for collaboration between enterprises.
Information brokers	Information brokerage is the service of adding value to the huge amounts of data available on the open networks.
Trust services	Supports the assurance, integrity or security of electronically executed activities.

Table 4.1: Business model categorization of Timmers [77]

products and services on the Internet. A search engine doesn't belong to that class of models. The two value chain models support electronic commerce business like e-shops, e-mall, etcetera. Collaboration platforms are specific for enhancing collaborations between enterprises. So the value chain models and the collaboration platform are also not relevant to the search engine business.

The only business model from Timmers that is relevant to the Internet search engine business is information brokerage.

- **Information Brokerage:** Timmers has defined information brokerage as follows :

  > Information brokerage is the service of adding value to the huge amounts of data available on the open networks or coming from integrated business operations, such as information search, e.g. Yahoo (http://www.yahoo.com/), customer profiling, business opportunities brokerage, investment advice, etc [77].

The huge amounts of data available on the Internet have opened the market for a whole range of new information services. According to Timmers information and consultancy resulting from these services have to be paid directly either through subscription or on a pay-per-use basis. Search engines on the other hand are a special category of information services, with advertising as the main source of revenue.

4.2.5 The Internet business model taxonomy of Rappa

Rappa has made a categorization of nine business models for electronic commerce. These models are listed and briefly explained in figure 4.2.

Brokerage	Brokers in the brokerage model gather buyers and sellers together and ease transactions.
Advertising	Generating revenue through advertising.
Infomediary	Collecting and selling information form the Internet to other businesses.
Merchant	This model stands for the classic wholesalers and retailers of goods and service, which are also known as e-tailers.
Manufacturer	Manufactures can reach their clients directly and thereby bypassing the retailer.
Affiliate	Collaborations between websites.
Community	Users have a high investment in both time and emotion in the site. They often come to a particular site. This gives opportunities for advertising.
Subscription	Generating revenue through subscription fees.
Utility	The utility or ”on-demand” model is based on metering usage, or a ”pay as you go” approach. The more you make use of a service, the more you have to pay.

Table 4.2: Business model categorization of Rappa [58]

There are six business models of Rappa that are not relevant to the Internet search engine business. The source of revenue for the brokerage model is a commission for each transaction. Furthermore Rappa sees this model as a way to gather buyers and sellers together. This model is more relevant to buying and selling of products. The merchant and manufacturer model are concerned with buying and selling products. The community model is about users who often come to a website and make a content contribution to the website. The utility model is based on the meterage usage of services and the infomediary model is about consumer data and their buying habits.

The following three business models of Rappa [58] are relevant to the Internet search engine business.

- **Advertising Model:** Radio and television programming has been broadcast for free over the airwaves the past century. But somehow the broadcaster always made money when it programming attracted many people. The advertising model is an extension of the traditional media broadcasting model. The broadcaster is in this case a website that offers content like news or services like e-mail. Usually this will be offered for free. The content and the services on the web site are mixed with advertising messages in the form of banner advertisements. This banner advertisement is probably the most important source of revenue, if not the sole source of revenue. The advertisement model only works when the volume of viewer traffic is large or highly specialized.

- **Subscription Model:** With this model the users pay a fee to get access to a website. An example of such a website is the search engine MEDLINE (`http://medline.cos.com`). MEDLINE is a source of life sciences and biomedical bibliographic information, with nearly eleven million records. The user can get access by subscribing to the website with a fee. This model seems to work only with high value-added content. Content that is important enough for people that they are willing to pay a fee to get access to.

- **Affiliate Model:** The affiliate model provides purchase opportunities wherever people may be surfing. The affiliates provide a purchase- point-click-through to the merchant. The merchant returns a percentage of the revenue to the affiliate when a purchase is made through this affiliate. An Internet search engine can generate revenue by being an affiliate of a merchant like Amazon.com.

4.2.6 The information market

The Internet is a place where information suppliers and information seekers can meet each other for the exchange of information. This exchange of information on the Internet is seen as an information market, where demand and supply of information meet [6, 7, 25]. The information market can be seen as the business opportunity of search engine companies. The strategy of how to make money by these companies depends on the information market, in other words the business models of search engines are based on the information market. The enormous size of the Internet has made it difficult for the information seeker to find the appropriate information and for the information supplier to supply the information to the appropriate seeker. The difficulty of finding each other on the Internet information market calls for an information broker to assist both parties in finding each other.

An Internet search engine can be seen as an intermediary between information providers and information seekers. The business opportunities of a search engine is centered around the information market. The following definitions were found [7]:

- **Information market:** The information market is the market where resources are exchanged between searchers and publishers, possibly by means of brokers.

- **Resource:** Resources are the 'entities' on the Web that make up information supply.

From the above definition of the information market you can conclude that the broker could play an important role with the exchange of resources on the Internet. Examples of these resources are Web pages, newsgroups, mailing-list archives, networked databases, applications, business services, as well as indexing services. These resources are there for doing business, search other information, educational purposes, or relaxation [6].

The information market model consists of a certain number of stakeholders. We have identified the following stakeholders (see figure 4.1):

- **Information brokers:** The broker is described as a participant in the transaction of resources between the seeker and the supplier. The broker does not alter a resource that is exchanged and furthermore the broker has a value adding role for the supplier and the seeker[7].

- **Information suppliers:** In the above definition of the information market the information supplier is called publisher. The information supplier wants to reach information seekers who might be interested in the information provided by the publisher. The supplier knows that being accessible through a broker will enlarge the chance to be found because information seekers will probably consult a broker to satisfy their information need.

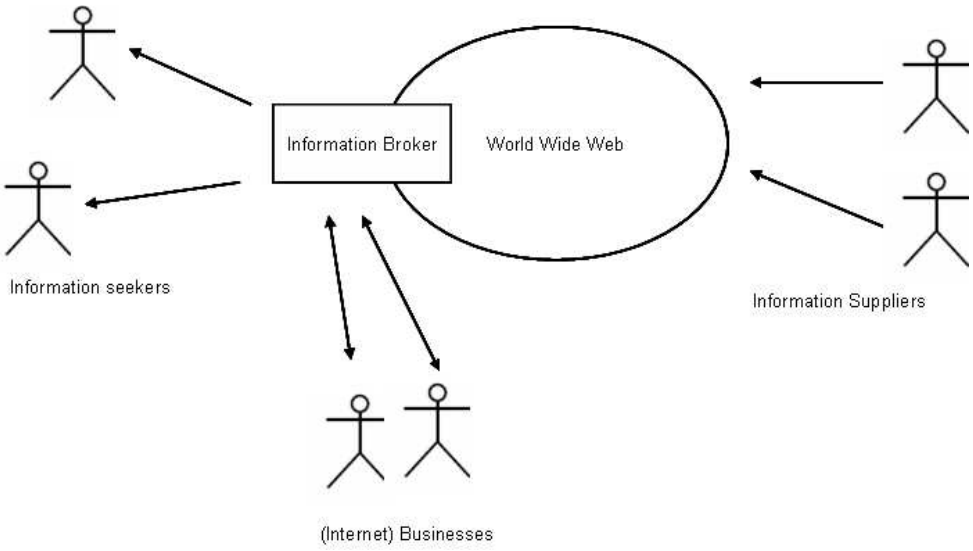


Figure 4.1: The Information Market

- **Information seekers:** The objective of the information seeker is to find the needed information that is available on the Internet. Without a broker it will be very difficult and time consuming to find the needed information. Furthermore the broker can give the information seeker more confidence about the quality of the information.

- **(Internet) Businesses:** Businesses realized that the World Wide Web is a portal to reach customers. Reaching these customers is very difficult since the World Wide Web is extremely large. These businesses can find customers with the help of brokers.

The information market is the business environment of search engines. Search engines are dependent on the information market for the success of their business. The growth of the Internet in recent years has resulted in the increase of the demand of information by Internet users, which means a growth of the information market. The growth and development of the information market has given search engines more business opportunities. Search engines are developing more services based on a specific demand of information need by Internet users. Examples of services based on specific information needs are news search engines, image search engines, scientific articles search engines, and many more.

4.3 Survey

The objective of this survey is to identify the methods of how search engines earn their money and how these methods can be applied to search engines for children. In other words we are trying to find out what kind of business models are suitable for search engines for children.

4.3.1 Set up

The requirements analysis of Weeda and van Gendt [78] has shown that there is a need for search engines for children. In spite of this need there are very few search engines for children. This makes it difficult for us to investigate the business models of search engines for children. For this reason we will investigate the business models of several types of search engines. With this investigation we hope to get more insight into the search engine business, which can be of input in formulating a business model for search engines for children.

There are many search engines available on the Internet. But what makes a search engine a search engine? Is a Internet bookshop like Amazon just an e-shop, or can you also call it a search engine for books? The borderline between search engines and not search related web services is not always clear. In some cases it depends on how you use the service. For instance a person who wants to buy a book goes to Amazon to see how much it costs, this person uses Amazon as an e-shop. But Amazon can be used as a search engine for a person who wants to buy a book on a certain subject, but does not know the title of the book.

The website Search Engine Watch (www.searchenginewatch.com) provides all kinds of information about search engines. For instance they provide analysis of the search engine industry. This website has identified the following categories of search engines that we will use as the basis for our own categorization of search engines in this survey [69]:

- Major search engines and directories
- Metacrawlers and metasearch engines
- News search engines
- Pay per click search engines
- Shopping search engines
- Multimedia search engines: Image, Audio & Video Searching
- Search Toolbars & utilities
- Kids search engines
- Specialty search engines
- Country-specific search engines

The above categories are not all relevant for our research. Some categories contain overlap with other categories. So we have created our own categorization of search engines that will be relevant for our research. Because of the large number of

search engines on the web, it is not possible to investigate all search engines that are available. To keep our research feasible, we have decided to investigate two search engines from each category of search engines that we find relevant for search engines for children. These search engine classes will be discussed in the next section (section 4.3.2).

4.3.2 Search Engine Classes

We have chosen a subset of four classes of search engines: generic search engines, Dutch search engines, search engines for children and specialty search engines.

The categories that were enumerated in section 4.3.1 form the basis of the four classes of search engines that we will use in this survey. The category major search engines & directories and the class metacrawlers & metasearch engines are translated into our classification of generic search engines. The difference between a regular search engine and a metacrawler is the method of retrieving the search results. A regular search engine crawls the Internet to retrieve documents, a metacrawler sends the search requests to several search engines all at once. The results are then blended together onto the page. For our research we see this category as part of the generic search engines, because the type of retrieved information is the same for both types of search engines. So we will not make a difference between these two classes.

News search engines, shopping search engines and multimedia search engines belong in this research to the specialty search engines. These search engines cover a special part of the information on the Internet. The specialty search engine only has to focus on a smaller information area than for instance the generic search engine. This specialty search engine can put more effort in the presentation of the information coming from that information area. The presented information might be of more value to the user than the information resulted from the same search task with a generic search engine. You can find that specific information easier with that special search engine than with a generic search engine. For instance the Dutch specialty search engine Lens Express (www.lensexpress.nl). This is a search engine for finding and comparing the prices of different brands of contact lenses and buying them from this search engine. Besides that this search engine is an e-shop, it can be used by users to see the prices and compare them with the prices of other contact lens suppliers and shops. People who are searching for contact lenses can probably find more useful information with Lens Express than with a generic search engine. The specialty search engine category is covered in this survey.

Search toolbars & utilities is about searching and navigating the Internet without going directly to a search engine. Our focus in this project is to create a search engine for children with functionalities that cannot be realized by a search toolbar that the user installs on the desktop of its computer. The user has to go the website of the search engine. So the category of search toolbars & utilities is not relevant for our survey.

Pay-per-click search engines are about a specific method of advertising. This method is becoming more popular. There is some overlap with other categories. For instance Google belongs to this category, but it also belongs to the category of major search engines. For this survey it is not important to have a special focus on the pay-per-click search engines, because there is overlap with other search engine categories. We can include the pay-per-click aspect by including search engines in our survey that also belong to other categories.

Generic search engines

The most used and popular search engines on the Internet are the generic search engines. At this moment there are three big players active in the search engine business: Google, Yahoo and MSN Search. These search engines are quite successful and it is interesting to see how they became so successful.

The two search engines that will be investigated are Google and Yahoo. These two search engines have the largest market share of all search engines [75]. Therefor it will be interesting to include these two search engine in our survey.

Dutch search engines

These search engines fall under the category country-specific search engines (see section 4.3.1). This category of search engines will be very interesting for us because we reside in the Netherlands. It is possible for us to get information about them through direct contact with these search engines.

Dutch search engines that we will investigate are Ilse and Ixquick. Ilse is the first Dutch search engine and it has become the market leader among the Dutch search engines. Ixquick was originally an American search engine and it was acquired by a Dutch company in 2000. It sees itself as the worlds most powerful meta search engine and it has the objective to be used worldwide, it can search in 18 languages. These two search engines have different goals, Ilse is active in the Netherlands and Ixquick wants to be active worldwide. It will be interesting for our survey to include these two search engines.

Search engines for children

A search engine for children is the main subject of this thesis. It is obviously that it will be of interest to investigate these kind of search engines. These search engines are designed to serve the information needs of children. Furthermore these search engines will filter out sites that parents and teachers might find inappropriate for children. At the moment there are some search engines for children available like Yahooligans, Ask Jeeves for Kids, KidsClick! and Looksmart's Kids Directory.

We are going to investigate two search engines for children: Yahooligans and Ask Jeeves for Kids. These two are major search engines for children. Yahooligans belongs to Yahoo! and is the oldest major directory for children. It was launched in 1996. Ask Jeeves for Kids is a unique service where you enter a question, and Ask Jeeves for Kids tries to point you to the right web page that provides an answer.

Specialty search engines

These search engines will help the user to find specific information on a certain topic or class of information. Examples of these kind of search engines are shopping search engines, science search engines, medical search engines and travel search engines. This category is interesting because search engines for children can also be seen as specialty search engines.

There are many specialty search engines available, but due to time constraints we cannot investigate all of them. We have to make a choice which two search engines we will investigate. We have chosen for the search engines/Internet portals Expedia.com and FindArticles.com. Expedia.com provides users information on traveling products and services. Traveling products and services are very popular on

the Internet and Expedia.com is a big player in this area. We have chosen a different information area for the second search engine. In our search for search engines we have found FindArticles.com, a search engine for publications in magazines and other sources. They claim that they have 10 million articles. This website belongs to LookSmart, another generic search engine. It is very interesting to investigate this search engine for publications, and also because it belongs to another search engine, we therefor get the chance to see how another search engine company does business.

4.3.3 Research Method

From each class we will investigate two search engines and try to find out how they generate revenue. What we have to do is find out what the different business models are of the search engines. We will be using the business model definition of Timmers [77]. This definition is often used by e-commerce researchers and it is the best known definition for Internet business models. By using the business model definition of Timmers we have to give a description of the following aspects of each search engine (see section 4.2.3):

- business actors and their roles
- benefits for the various business actors
- sources of revenue

The process of finding the appropriate information will be divided into three activities:

- Gathering information from the website of the search engines
- Gathering information from scientific articles, Internet articles and other media
- Contact with search engines & interviews

At the end we will analyze the information and try to formulate a business model for each investigated search engine.

The business actors can be divided into the four groups of stakeholders that were discussed in section 4.2.6. But still in each group there might be essential differences between the business actors that might influence the method of doing business.

Probably most search engines have in common that advertisement is their main source of revenue. But the advertisement policies of these search engines are probably not all the same. Advertisement can be used in different ways, which might affect the method of doing business.

4.4 Results

4.4.1 Introduction

The results from the investigation of the eight search engines will be presented in this section. From each search engine we will give the business model, which consists of the actors and their benefits and the sources of revenue. The information from the different sources, that was used for this investigation can be found in appendix A.

4.4.2 Google

Google Inc. is the developer of the Google Search Engine. This company was founded in 1998 by Larry Page and Sergey Brin [10, 55]. Since then Google has become the worlds largest search engine. Google is a generic search engine.

Actors & Benefits

- *Investors*
  Google needs money to finance its business. A source of financing its business is investors. Investors can buy shares of Google Inc. If Google does well than the value of the shares will increase. The objective of investors is to gain profit with their shares, meaning the value of their shares should become higher than the value of the shares at the time when they bought them. They are of course also interested in getting dividends from their stock. Google Inc. is doing very well at the moment, which has resulted in an increase of value of its stock. The dividend policy of Google is that they have never declared or paid any cash dividend on their common stock. They currently intend to retain any future earnings and do not expect to pay any dividends in the foreseeable future. The only possible benefit for the stockholders is the possible increase in value of their shares.

- *Advertisers*
  Google has a very large user base, which means that advertisers can reach a very large audience through Google. The ads are expressed in words as sponsored links. The service of Google towards its advertisers is automated with the program AdWords.

  With AdWords the advertiser has control over most of the elements of their ad campaign. The advertiser can design his own ad, specify on what search terms his ad is shown and define how much he is willing to pay. The advertiser only has to pay when a user clicks on the ad. Designing the ad is bound to restrictions. The advertiser can only specify the title and a short description. The user can click on the title to go to the web page of the advertiser. The idea behind AdWords is to advertise with ads that only consists of words, it is not possible to include images or other visual components. Google has proven that this method is very successful. AdWords shows only ads to people seeking information on topics specified by the advertisers. This way advertisers reach their targeted audience and users only see ads related to their interests. When several ads on a specific topic are available, only the ads of the highest bidding advertisers are shown. This auction-based system forces advertisers, who want their ads shown on common search terms, to raise the price for their ads.

  For advertisers with more budget Google is offering services through an account team of professionals. Basically the advertiser only needs to specify the ad and determine the budget, the rest is done by Google.

- *Website publishers*
  Besides displaying ads on its search engine, Google is also providing user friendly ads to third-party websites. User friendly means no pop-up ads or ads that take over websites. Internet users won't be annoyed. These third-party websites make up the Google Network. The third-party websites can generate

revenue by placing ads on their website. These ads only contain text, the ad has a title and a short description. The title is also a link to the website of the advertiser. The advertiser only has to pay when a user clicks on this link. By cooperating with Google these websites have the advantage that they can reach the large base of advertisers of Google. The websites don't have to find advertisers for themselves, the advertisers are supplied by Google. Furthermore Google is supplying ads that are targeted to the visitor of a specific website.

- *Business Clients*
  Companies can use the search technology of Google to search for corporate documents. Another business possibility is that Google executes Internet search jobs for other websites. The advantage of this is that these companies and websites don't need to develop their own search technology, the search technology is provided by a specialist like Google.

- *Users*
  The main objective of Google is to attract as many users as possible. Their business value is centered around the interest of the user. For instance according to Google they deliver relevant and useful information to their users. The high quality search technology makes this possible. According to Google they are objective, in the sense that they don't accept money for search result ranking. Furthermore they provide global access and ease of use to their users. The services of Google are based on the information interests of the users.

Sources of revenue

The main source of revenue of Google is advertisement. They generate revenue by delivering relevant, cost-effective online advertising. In other words you can say that Google acts as a middleman for online advertsing. Google connects advertisers with websites and vice versa. The advertisement revenue will increase if the number of users increases.

Another source of revenue is the supply of their search technology to companies and websites that don't have the resources to develop their own search technology.

4.4.3 Yahoo

Yahoo is an Internet company that was founded in 1994 by two Stanford Ph.D. students, David Filo and Jerry Yang. They started as a guide to keep track of the personal interests of its founders on the Internet. Soon they evolved in keeping track of the interests of other Internet users. Nowadays they have evolved themselves into a major Internet company that is offering many services to Internet users. Among those services is a search engine. Their search engine belongs to the class of generic search engines.

Actors & Benefits

- *Investors*
  Yahoo is a public company that sells stocks on the Nasdaq stock market. Their goal is to attract money that is needed to sustain their business. Investors gain advantage when the value of the stocks increases. Yahoo is a very successful Internet company. They generate high volumes of revenue, which results in

an increase of the value of the stocks. Yahoo has not paid or declared any cash dividend on their common stock. They have no plans for the foreseeable future to pay any dividend. All of their earnings will be retained to fund the development and growth of their business. The only possible benefit for the stockholders is the possible increase in value of their shares.

- *Advertisers*
  The services of Yahoo are attracting users worldwide. Their user base is very large and diverse, which makes it very interesting for advertiser to advertise on Yahoo websites and services. Also like Google, Yahoo is offering advertisements to third-party websites. Furthermore Yahoo offers advertisers the possibility to place sponsored links on the search results page of the Yahoo search engine. Furthermore Yahoo also offers advertisers to place banner ads on the Yahoo pages.

- *Website publishers*
  Third-party websites that want to place ads on their website can find advertisers through Yahoo. Yahoo has a large base of advertisers who, besides placing ads on Yahoo websites, also like to place ads on the third-party websites of Yahoo, the so called Publisher Network. Websites earn revenue by placing these ads, that are supplied by Yahoo.

- *Business Clients*
  Yahoo is attracting many users which makes it interesting for companies to cooperate with Yahoo. For instance companies can sell their products and services online through the service Yahoo! Shopping.

- *Users*
  Yahoo's top priority is to provide users with relevant information. This is ensured by a wide range of products and services. The Yahoo search engine is one of those services. These users are essential for the business of Yahoo. Yahoo will continually maintain and develop services to keep users satisfied. The benefits for the users is that Yahoo is providing many Internet services that these users need.

Sources of revenue

Yahoo recognizes two sources of revenue: marketing services revenue (advertisement) and fees revenue. Marketing services is the main source of revenue, it generates 88% of the total revenue of Yahoo. Advertisers are attracted by the large user base of Yahoo. One of the advertisement possibilities is the search engine of Yahoo. The other 12% are generated by the fees revenue. Users can get premium services from Yahoo by paying a fee. For instance besides the free e-mail service, Yahoo has a premium e-mail service with more options for the user, like bigger storage space.

4.4.4 Ilse

Ilse is a Dutch search engine that was founded in 1996 by Wiebe Weikamp, Merien ten Houten and Robert Klep. Ilse was the first search engine of the Netherlands. Since then it has evolved into the market leader of Dutch search engines. Currently Ilse belongs to Sanoma Uitgevers, and it is placed within the company Ilse Media

B.V. Ilse is focusing on the Dutch market and is trying to gain back market share from its competitor Google. Ilse belongs to the category of Dutch search engines.

Actors & Benefits

- *Investors*
  The search engine Ilse is part of Ilse Media B.V. that on its turn is part of Sanoma Uitgevers. Sanoma Uitgevers belongs to the Finnish public company Sanoma. Investors can buy shares of Sanoma on the Finnish stock market. Because Ilse is a small part of Sanoma, it does not have a large influence on the value of the stocks of Sanoma. The investors are certainly actors of Ilse, and their influence on Ilse can be in potential very large, but the value of Ilse is quite small to Sanoma. This means that Sanoma can let go Ilse very easily. Sanoma pays dividend to its stockholders. Investors can have a direct financial benefit when Sanoma does well, in other words the investors have a share in the profits of Sanoma.

- *Advertisers*
  Ilse offers advertisers the opportunity to advertise on the result page. These ads are in the form of banners that are placed on the top of the result page. The search engine Ilse is generating approximately 16 million page impressions per month. This means that the advertiser can reach quite a large group of Internet users. Ilse has investigated that most of the users of its search engine are women of the age category 30+. Advertisers can focus on this group.

- *Overture/Yahoo! Search Marketing*
  The results pages of Ilse display sponsored links. What kind of links are displayed is dependent on the information users are searching for. These sponsored links are handled and delivered by Yahoo! Search Marketing, formerly known as Overture. The advantage of Yahoo! is that they enlarge their publishing network with Ilse. Ilse shares the revenue from these sponsored links with Yahoo! Search Marketing.

- *Commercial Partners*
  Ilse has cooperations with several other partners like Gouden Gids (yellow pages), Achmea (insurance company) and Postbus 51 (government's centre for public service information). Information from these partners are displayed on the result page when a user searches for information that is similar to the information of the partner. For instance when a user searches for a plumber, Gouden Gids will supply links to plumbers that can be found in the Gouden Gids. Gouden Gids will pay Ilse for linking users to their website. The advantage of this method is that these partners can enlarge their reach by cooperating with Ilse.

- *Users*
  Ilse is focusing on Dutch Internet users. According to themselves their strength lies in the fact that they are located in the Netherlands, so they can adapt their information supply quickly to the needs of the Dutch Internet user. Furthermore the users can have access to specific Dutch information, like information from Dutch partners like Gouden Gids.

Sources of revenue

Ilse has three sources of revenue: banners, sponsored links and commercial partners. Banners and sponsored links belong to the category of advertisement. Ilse generates revenue from commercial partners by linking the users on the result page to these commercial partners.

4.4.5 Ixquick

Ixquick is a meta-search engine. A meta-search engine is a search engine that sends user requests to other search engines and presents the results from these search engines to its user. Ixquick is originally an American search engine, but it was incorporated in 2000 by the Dutch company Surfboard Holding B.V. Ixquick belongs to the category of Dutch search engines.

Actors & Benefits

- *Search Engines*
  Ixquick does not have its own search engine, it is dependent on other search engines. Ixquick has an agreement with several search engine companies to use their search engines with no charge. On the Ixquick result page it is shown from which search engine the result comes from. You could say that Ixquick serves as a promotion channel for these search engines.

- *Overture/Yahoo! Search Marketing*
  Ixquick displays sponsored links on top of the result page. These sponsored links are supplied by Yahoo! Search Marketing, formerly known as Overture. Supplying sponsored links is the business of Yahoo! Search Marketing. By supplying to Ixquick means that Yahoo! Search Marketing enlarges it market share in the sponsored links business.

- *Users*
  Ixquick profiles itself as the most powerful meta-search engine in the world, and that it develops its services with the users interest in mind. In their view that means a user friendly search engine with a simple user interface and no irritating banner and pop-up ads. Furthermore users can choose which search engines are used when they search with Ixquick.

Sources of revenue

The only source of revenue comes from the sponsored links provided by Yahoo! Search Marketing. They generate enough revenue to be profitable, but their budget does not allow them to promote their search engine. They are dependent on word-of-mouth promotion among Internet users.

4.4.6 Expedia.com

Expedia.com is an Internet portal/search engine that provides services to plan a complete traveling trip. Examples of these services are search engines for flights, hotels, and rental cars. Expedia.com is a U.S. corporation with activities in the United States, United Kingdom, Canada, Germany, France, Italy, the Netherlands and Australia. Expedia.com is classified into the specialty search engines category.

Actors & Benefits

- *Investors*
  Expedia.com used to be a part of IAC Travel, which is a segment of the public traded company IAC/InteractiveCorp. On December 21, 2004 IAC announced that all the traveling services will be put into a new public traded company called ”New Expedia”. This has resulted that the performance of Expedia.com is of more influence to the company where it belongs to. The money of the investors is needed to finance the business. The objective of the investors is of course to gain profit with their shares, meaning an increase in value. The other possible objective of investors is to acquire dividend on their stocks, but so far they have not paid any cash dividends on their common stock. Future dividend payments will depend on, among other things, their results of operations. At the moment they do not have plans for dividend payments, but it is not ruled out that it won't happen in the foreseeable future.

- *Advertisers*
  Expedia.com makes a distinction between travel advertisements and non-travel advertisements. Advertisers can list their travel products and services on Expedia.com. Expedia.com offers these advertisers an advertising campaign for more exposure. Expedia.com is a very interesting advertising place for travel service and product suppliers.

- *Affiliate Websites*
  Websites can partner with Expedia by selling travel products and services from Expedia.com. These websites have to link their users to Expedia.com. The benefit is that these websites earn a commission when their users buy travel services or products on Expedia.com.

- *Travel Supplier Partners*
  Expedia.com can be seen as a sales channel for companies who sell travel services and products. These companies can partner with Expedia.com. Expedia.com offers their partners a wide range of innovative, targeted merchandising and promotional strategies designed to increase their revenues, while simultaneously reducing their marketing transaction and customer service costs.

- *Users*
  Users who are interested in traveling products or services can find it all in one place. They don't have to surf on the Internet to find the suppliers of travel services. Everything that is needed to plan a trip can be found in one place.

Sources of revenue

Expedia.com generates revenue through advertising on their website and through two business models: merchant business model and agency business model. The revenues from these two business models are coming from their travel supplier partners. With the merchant model Expedia.com buys travel services from the suppliers and subsequently sells it to the consumer. Revenue is derived from the difference between amounts paid to the travel suppliers and the amounts paid by the consumer. With the agency model Expedia.com acts as an agent in the transaction between travel supplier and consumer. Revenue is derived from commissions and ticketing fees from travel suppliers.

4.4.7 Findarticles.com

FindArticles.com is a search engine for publications. They claim that they have 10 million articles from magazines and journals. FindArticles.com is a service provided by LookSmart, Ltd. We have classified this search engine into the specialty search engine category.

Actors & Benefits

- *Investors*
  FindArticles.com is a service provided by LookSmart, which is a public traded company. LookSmart defines itself as a provider of products for advertisers who wish to pay to be included in relevant web search results. FindArticles.com is one of those products that LookSmart offers to its advertisers. The money coming from investors is needed to finance the business of LookSmart. If LookSmart does well, than the value of the stocks will increase. LookSmart has not declared or paid any cash dividends on their common stock, and they intend to retain their future earnings to fund the development and growth of their business, therefore they do not anticipate paying any cash dividends in the foreseeable future. The only benefit for stockholders is the possible increase of value of the stocks.

- *Advertisers*
  FindArticles.com offers advertisers the possibility to place ads on the result page in the form of banners and pop-up ads. The objective of these advertisers is to reach as many people as possible.

- *Publishers*
  The publishers can partner with FindArticles.com to be included into the search results. FindArticles.com acts as an agent to guide Internet users to their publications.

- *Google*
  Google provides FindArticles.com sponsored links. These links are displayed on the result page of FindArticles.com. Providing sponsored links is a very important business objective of Google, so by providing sponsored links to FindArticles.com, Google expands its customer base.

- *Users*
  Users who want to find publications can find it with FindArticles.com. This service is developed to make it easier for users to find the needed publications. FindArticles.com claims that they have access to 10 million articles.

Sources of revenue

One of the sources of revenue is commissions for purchases of premium content. The user needs to pay a fee for this premium content. FindArticles.com gets a commission from the publisher if the user has been linked to the publisher through FindArticles.com. The other sources of revenue are advertisements from banner and pop-up ads, per-click fees on their own paid listings, and revenue sharing from clicks on paid listings provided by Google, Inc.

4.4.8 Yahooligans!

Yahooligans is a web guide for kids in the age of 7 to 12 years. This service is provided by Yahoo. The websites that are included in the directory are carefully checked by educators to ensure that the content and links are appropriate for children. We have classified this search engine into the category of search engines for children.

Actors & Benefits

- *Investors*
  Yahooligans is part of Yahoo, Inc. There is very little information in the annual report of Yahoo about Yahooligans.

- *Advertisers*
  The result page of Yahooligans displays banner ads, but these ads are all ads for other Yahooligans services. The only external ad that we have seen so far is the Cartoon Network ad.

- *Users*
  The users of this service are children in the age of 7 to 12 years. Yahooligans returns results that are interested and appropriate for children.

Sources of revenue

There is very little information about Yahooligans. The result page of Yahooligans displays banner ads, but these are all ads for services of Yahooligans. For instance Yahooligans Music. The only external ad that we have seen so far is from Cartoon Network.

According to Nils Rooijmans of Ilse search engines for children are also used as a promotion tool for Internet branding. Yahooligans is an example of this kind of promotion tools. By attracting young children, Yahoo hopes to maintain these Internet users for Yahoo when these children get older. So Yahooligans does not have any sources of revenue, it has purely a promotional objective.

4.4.9 Ask Jeeves Kids

Ask Jeeves Kids is a service of Ask Jeeves Inc. It is a search engine for children. This search engine is different from other search engines, because children can ask questions to the search engine just like as they would ask their parents or teachers.

Actors & Benefits

- *Investors*
  Ask Jeeves Inc. is a public traded company. The objective of investors is that the company is doing well, so their stocks will increase in value. Generally investors are interested in the methods how the company tries to achieve its business goals. We think that Ask Jeeves Kids is a promotional tool to attract Internet users at a young age.

- *Advertisers*
  We have found an advertisement policy on the website of Ask Jeeves Kids. This policy states that Ask Jeeves Kids only accepts ads that are appropriate

for children under the age of 13. So far we have not seen any advertisement on the result page of Ask Jeeves Kids.

- *Users*
  Aks Jeeves Kids offers children a unique user friendly method of searching on the Internet. Children can ask the search engine questions just like as they would ask their parents or teachers.

Sources of revenue

The annual report of Ask Jeeves states that they earn revenue primarily as users click on advertisements that are displayed when they navigate the Internet. Since we have not found any advertisement at all on Ask Jeeves Kids, we can conclude that this search engine has a promotional objective for Ask Jeeves Inc.

4.5 Conclusion

The demand and supply of information on the Internet, the information market, has created the business opportunity for companies to supply Internet search engines to assist in the exchange of information on the Internet. A very important aspect in sustaining these search engines is money.

The management literature has introduced the concept of business models. From this literature it became clear that business models are essential for the success of a business. The business model describes in short what a company does to achieve the objective of generating revenue. This can help us in answering how search engines earn money. The best known and used definition of the term 'business model' in the electronic commerce literature is defined by Timmers [77]. We have used this definition in our survey to identify how search engines generate revenue.

Table 4.3 gives a enumeration and a short description of the sources of revenue that were found.

Advertisement	Advertisers place ads on the search engine. This method of generating revenue is very common in the search engine business.
Searching fees	The user has to pay a fee before he can use the search engine.
Promotional	The search engine is used to promote a brand, the service is freely available for everyone.
Partners	A partner can cooperate with the search engine company in offering services and content to users.
Search services	The search engine company can offer its search services to other web sites.
Commission	The search engine company can get a commission from a retailer when a user, who is guided by the search engine to that retailer, buys a product.

Table 4.3: Sources of revenue

We have identified six sources of revenue with the current search engines. Only three of them might be interesting for search engines for children. Some of these methods need to be adjusted to make them appropriate for the search engine business for children.

Partners can have far stretching influence on the business of the search engine. The partner and the search engine cooperate with each other to create a win-win situation, but both parties could have conflicting objectives. These conflicting objectives can affect the quality of the content that is offered by the search engine. A search engine can easily drop an advertiser without far stretching influence on the business and image, but the consequences of dropping a partner can be very big. It is better for a search engine for children not to cooperate with a (commercial) partner, because you don't know what the possible negative effects might be for the future. The search engine will lose the total independency when it cooperates with a partner.

A very significant aspect of search engines for children is the method of presentation. The content has to be presented in a way that children can better understand the information. Doing search services for other websites means that the search engine has to drop the possibility of presenting the content. Providing search results without the presentation aspect means that the search engine loses its identity.

The commission method is not suitable for search engines for children, because it is not desirable that children can buy products from the Internet. If a child wants to buy a product, the child can ask his parent to buy the product. The commission method is classified by Rappa as the affiliate model (see section 4.2.5).

We will discuss the three appropriate methods in the subsequent subsections.

4.5.1 Advertisement

The most used method of generating revenue by search engines is advertisement. This method is identified by Rappa as the advertisement model (see section 4.2.5). The advertisement makes it possible to offer search engines to Internet users for free. There are several methods of applying advertisement to the search engine business. Some search engines display banners and pop-up ads on their results page, others offer advertisers the possibility to pay for being enlisted in the search results. A very successful method of advertisement is displaying sponsored links on the results page. The benefit of this method is that the user will see sponsored links that have a relationship with the search results, which means that the user will probably see useful advertisement.

When you talk about search engines for children there are major restrictions on advertisement. Basically the ads have to be appropriate for children. An option might be that these ads cannot contain any link to the website of the advertiser. Another option is that these ads do contain links, but advertisers have to create special websites for children which only contains information that is appropriate for children. Children are not allowed to buy the product or service through the website.

4.5.2 Search fees

Another option for generating revenue is that search engines charge a fee for the use of their service. Rappa has classified this method as the subscription model (see section 4.2.5). In section 3.2 we have introduced the educational view of the search engine. The Internet is a knowledge source for children, but these children don't have the ability to judge the relevance of the knowledge on the Internet for their

development. These children need guidance in gaining knowledge from the Internet, this can be in the form of a search engine for children. These search engines have the same role as schools. Parents have to pay tuition fees to schools for the education of their children. From an educational perspective, parents (or schools) can also pay a fee to search engines for letting their children gain knowledge from the Internet on a safe and responsible way.

4.5.3 Promotional

The third option is the use of search engines with a promotional purpose. In this case the search engine for children is part of another major search engine. These parent search engines want to make children get used to the brand. They hope that when these children get older, that they will make use of the parent search engine.

We have investigated two search engines for children, but we have not found any source of revenue. For instance the policy of Ask Jeeves Kids, which is a subsidiary of the search engine Ask Jeeves, says that they accept advertisement on their result page with the condition that these ads are appropriate for children under the age of thirteen years. But we have not found any ad at all on Ask Jeeves Kids. The same applies to Yahooligans, which is a search engine for children of the parent search engine Yahoo. They also don't charge fees for their services. We conclude that these search engines have a promotional objective for the parent search engine.

5 Development

This chapter describes the technical part of the software engineering process of our search engine for children, Giggle. The technical part consists of three technical activities - design, code generation, and test. These activities are required to build and verify the software. Each activity transforms information that results in validated computer software [57].

We have chosen for an object-oriented development approach. This method of developing software was first proposed in the late 1960s. It became popular and common among software developers throughout the 1990s. Object-oriented software engineering has several benefits. This method leads to reuse of program components, and reuse leads to faster software development and higher quality programs. Furthermore object-oriented software is easier to maintain because its structure is inherently decoupled, which leads to fewer side effects when changes have to be made. In addition, object-oriented systems are easier to adapt and scale. [57]

The activities that are required to build and verify the software will be addressed in the next sections of this chapter. We will discuss the design of Giggle in section 5.1. The implementation and testing of Giggle will be discussed respectively in section 5.2 and section 5.3.

5.1 Design

Designing is one of the three technical aspects of software engineering. Software design is an iterative process where requirements are translated into a ”blueprint” for constructing software. The requirements for Giggle were formulated by Weeda and van Gendt in the project *Giggle: Internet Search Engine for Kids* [78]. These requirements are discussed in chapter 2 of this thesis.

The design begins at a high level of abstraction. This level can be directly traced to the software requirements. The objective is to translate the requirements into a representation that can be used by implementers to generate code. This translation consists of several iteration steps. Initially, the blueprint depicts a holistic view of software, but as the design iterations continue, the design representations become more detailed and depict much lower levels of abstractions. These iteration steps are discussed in the subsection Design Process (subsection 5.1.1). The subsequent subsections present the design representations of each iteration step.

5.1.1 Design Process

The process of designing Giggle consists of several activities. It begins with the requirements of Giggle as input into the design process and it ends with a representation that can be translated into Java code. These activities are discussed in this subsection.

Application Model

The designing begins with the requirements of Giggle as input into the design process. These requirements were categorized into primary and secondary requirements in section 2.4.3. The first step is to translate the requirements into features that the search engine will offer to the user. This will lead to a view on application level of Giggle, which describes the functionalities and a global view of the behavior of the final application. This representation of Giggle can be used as input for the analysis model.

Analysis Model

The level of abstraction of the application model is too general to use it directly as a recipe for implementation. It needs to be further analyzed before implementation can begin. The analysis model begins with an understanding of the intended behavior of the system. Once these scenarios have been defined, the modeling of the software can begin.

The behavior of the system will be represented by a use-case view of the system. This view will be represented with UML notation, a diagrammatic representation of a use-case, called a *use-case diagram*. Once the use-case view has been developed, it is time to identify the classes of objects. These classes are identified together with their responsibilities and collaborations. A method to identify these classes is *class-responsibility-collaborator* (CRC) modeling. This method provides a simple means for identifying and organizing the classes that are relevant to system or product requirements. The CRC model encourages an intuitive understanding of the importance of the logical separation of the various components, helping to emphasize the cohesion and coupling [11]. After identifying the classes and objects using the CRC model, we will focus on defining structures and hierarchies of these classes. The structures and hierarchies are represented by a *class diagram* in UML notation. Another aspect of the analysis model is the dynamic behavior of the system. It is important to identify and model the response of the system to external events or stimuli. This modeling starts with use-cases as input to the event identification process. The event identification generates multiple class diagrams and sequence diagrams. These diagrams will be translated into Java code. [57]

To summarize the activities of the analysis model approach, this approach consists of the following activities:

1. Identifying the behavior of Giggle by means of a use-case view.
2. The CRC model is used to identify the classes with their responsibilities and collaborations.
3. Defining structures and hierarchies by means of class diagrams.
4. Identifying the dynamic behavior of Giggle by creating multiple class diagrams, state-chart diagrams and sequence diagrams.

5. Translate Java code from multiple class diagrams and sequence diagrams.

Architectural Design

The architectural design represents the structure of data and program components that are required to build a computer-based system. It shows the structure and properties of the components of the system on a high level view. It also considers the interrelationships that occur among all architectural components of a system [57]. The product of the architecture design will be an architecture model that describes the data architecture and program structure of the application.

User Interface Design

Giggle is developed for human users. These users need a communication medium to interact with the computer. The user interface design creates this communication medium between a human and a computer. The quality of the user interface is of essential importance for the success of Giggle. The user interface design begins with an identification of user, task and environmental requirements. The results of these activities form the basis for the creation of a screen layout. The final deliverable of the user interface design is a prototype of the interface. [57]

5.1.2 Application Model

The application features that are derived from the requirements are for the greater part coming from the requirements analysis of Weeda and van Gendt [78]. These features are presented in table 5.1. Each feature has a development priority level, primary or secondary.

Feature	**Priority**
Instructions/Help (F1, F3, F9, NF1)	Primary
Query input (F4, F5, F6, F9, NF1)	Primary
Directory (F3, F4, F5, F9,NF1)	Primary
Relationship-map (F4, F5, F9)	Primary
Query reformulation (F5, NF3)	Primary
Results list (F7)	Primary
Save search (F7)	Primary
Tab-pane (F6,F8, NF1)	Primary
User safety (NF2, NF3)	Primary
Time and date (NF2)	Primary
Hardware and Software (NF5)	Primary
Performance (NF6)	Secondary
Document Format Types (NF5)	Secondary
Relevance and Filtering (NF3)	Secondary
Natural language query input (F4)	Secondary
Metaphor (F2, NF1, NF2)	Secondary

Table 5.1: Prioritization of Giggle features

Primary Features

- *Instructions/Help (F1, NF1)*
  The help page contains an explanation of all the functionalities and tips how to search for information.

- *Query input (F4, F5, F9, NF1)*
  The user can search with the search engine by typing a query in the input field. The user can click on the 'Find' button to start the search task.

- *Relationship-map (F4, F5, F9)*
  The relationship-map is a 3-D map that indicates relationships between the input keyword and other words. The words in the map can be selected by the mouse to specify the information need and the word where the user has clicked on will be used as query input for the next search task.

- *Directory (F3, F4, F5, F9, NF1)*
  The directory contains categories and words where the user can use the mouse to specify the information need.

- *Query reformulation (F5, NF3)*
  The system will aid the user in his search for information by expanding the query with words when necessary.

- *Results list (F7)*
  The search results will be presented as a list of search results, with a title, a short description and the URL of that website. It will also highlight the query in the short description.

- *Save search (F7)*
  The user can save his search results. With this option the user can divide a search session over multiple time periods.

- *Tab-pane (F6, F8, NF1)*
  The tab-pane makes it possible to navigate between the various search tasks. Every search task is presented in a tab. The user can follow the search process and go back to previous search results.

- *User safety (NF2, NF3)*
  There is a need for a filtering mechanism to block inappropriate content, banners and popups.

- *Time and date (NF2)*
  The website contains an indication of the the time and date.

- *Hardware and software (NF5)*
  The system must be able to run on Windows 2000 or higher in combination with Internet Explorer and Mozilla browsers.

- *Performance (NF6)*
  The system must have an adequate performance, which means among other things that the reaction time has to be as short as possible.

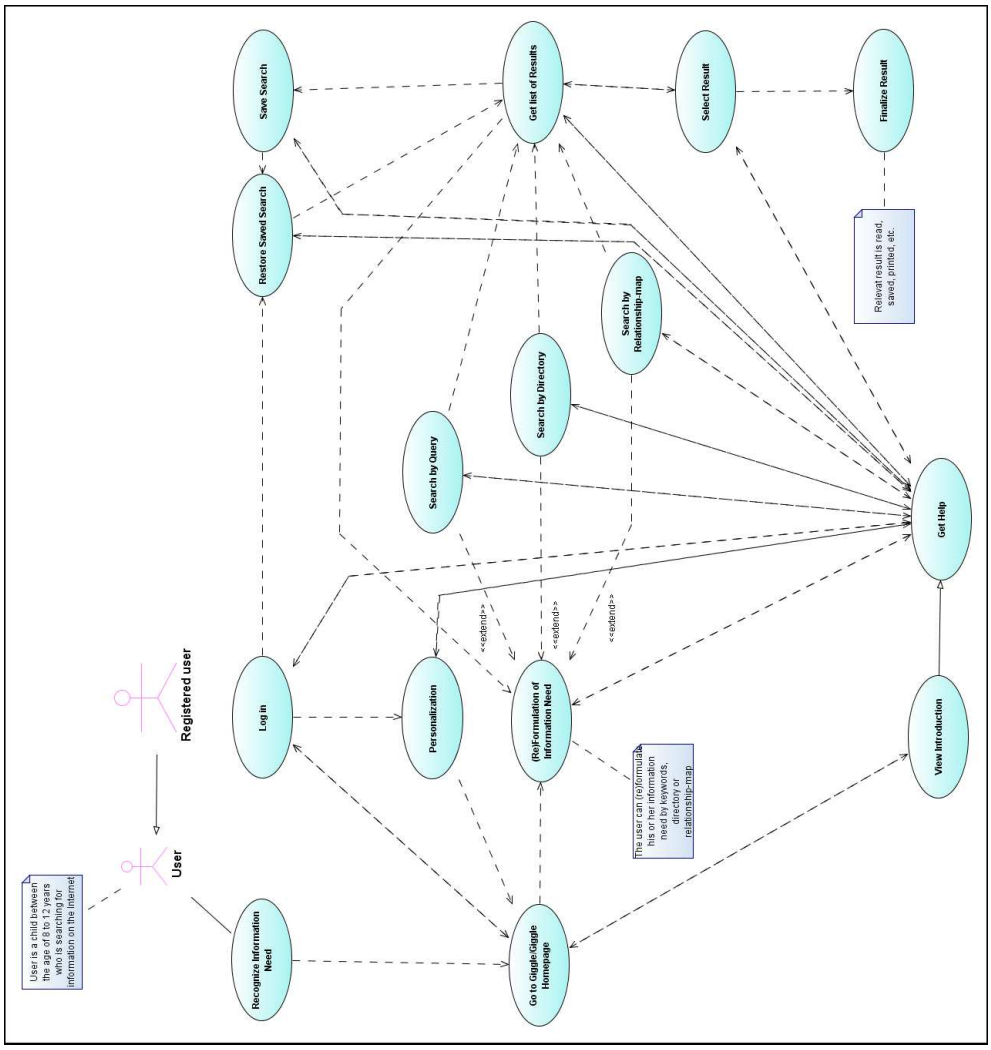


Figure 5.1: Use Case View

Secondary Features

- *Document Format Types (NF5)*
  It is preferred that the search engine can give search results in more format types than HTML. Example format types are PDF, Word, Excel, and Powerpoint.

- *Relevance and Filtering (NF3)*
  It is important that a search engine gives search results that are relevant to the user.

- *Natural language query input (F4)*
  Besides short terms in queries it might be useful to have the ability to give as input to the search engine a natural language sentence.

- *Metaphor (F2)*
  A metaphor can aid a child in searching for information with the search engine. It can elevate the user experience of the child, or it will help the child with

useful tips.

5.1.3 Analysis Model

Use-Case View

The behavior of Giggle is identified by means of a use-case view. This view is presented in a use-case diagram (figure 5.1). Each ellipse represents a user action. The arrows indicate the direction of the dependencies between the actions.

Class Responsibility Collaboration

The classes are identified with the CRC method. This subsection contains the CRC cards.

Class:	**Homepage**
Responsibility:	Log in Personalization Formulation of information need View introduction (help) See metaphor See results
Collaboration:	Login, Personilization, SearchInterface, Help, Metaphor

Class:	**Login**
Responsibility:	Allow users to log in Allow users to register Recovering lost password
Collaboration:	Homepage, PersonalizationDB

Class:	**Personalization**
Responsibility:	Save search Restore saved search
Collaboration:	Homepage, PersonalizationDB

Class:	**SearchInterface**
Responsibility:	Search by query Search by directory Search by relationship-map Providing search Presenting results
Collaboration:	Homepage, Search, Presentation, DirectorySearch, RelationshipSearch, QuerySearch

Class:	**Search**
Responsibility:	Reformulate query Post (re)formulated query for results
Collaboration:	SearchInterface, QueryReformulation, DocumentGathering, Indexer, PersonalizationDB

Class:	**Presentation**
Responsibility:	Presenting results
Collaboration:	SearchInterface

Class:	**Help**
Responsibility:	View introduction Contents and index
Collaboration:	Homepage

Class:	**Metaphor**
Responsibility:	Entertain users Give tips
Collaboration:	Homepage

Class:	**DirectorySearch**
Responsibility:	Show directory categories on screen Allow navigation through the categories (iterative search) Return query
Collaboration:	SearchInterface

Class:	**RelationshipSearch**
Responsibility:	Show word relationships in a 3D map on screen Allow navigation through the 3D map (iterative search) Return query
Collaboration:	SearchInterface

Class:	**QuerySearch**
Responsibility:	Allow user to input query Return query
Collaboration:	SearchInterface

Class:	**Query Reformulation**
Responsibility:	Reads query Post reformulated query
Collaboration:	Search, DocumentContextsDB

Class:	**Document Gathering**
Responsibility:	Send query to Meta Search Filter documents Rank documents Return ranked document list
Collaboration:	Search, MetaSearch, Ranking, Filtering

Class:	**Indexer**
Responsibility:	Index documents on important terms Store these documents and their terms in the database
Collaboration:	DocumentContextsDB, Search

Class:	**PersonalizationDB**
Responsibility:	Store and retrieve searches Store and retrieve login information Store and retrieve personalization information
Collaboration:	Login, Personalization, Search

Class:	**DocumentContextsDB**
Responsibility:	Store and retrieve terms from child documents
Collaboration:	Indexer, QueryReformulation

Class:	**Meta Search**
Responsibility:	Send query to the search engines return document lists from search engines
Collaboration:	Document Gathering

Class:	**Filtering**
Responsibility:	Filter document lists on inappropriate content Return filtered document list
Collaboration:	Document Gathering

Class:	**Ranking**
Responsibility:	Rank documents of search engines Return ranked list
Collaboration:	DocumentGathering

Class Diagram

Based on the CRC cards we have created a class view of Giggle. This class view is represented with a class diagram (figure 5.2).

Sequence Diagrams

We have used the use case view to identify the dynamic aspects of Giggle. These aspects are represented with a sequence diagram (figure 5.3).

5.1.4 Architectural Design

Figure 5.4 shows the high-level overview of the architecture of Giggle. The architecture of Giggle is based on the client/server model. Giggle runs on a computer (server) and the user can call on the service provided by Giggle through their own computer (client). Giggle will run on a web server which makes it possible for the user to access Giggle via the Internet with a web browser. The proposed client/server architecture for Giggle is a three-tier system. It contains a presentation layer (the client-side user computer), a processing layer (the web server) and a database layer (database server) to store document and user information. [57]

The architecture of Giggle contains on a high level the following components:

- *Browser:* The user can access Giggle via the Internet with a browser. The user interface of Giggle is made using Java Applets and HTML code.
- *Web Server:* The Giggle application runs on the web server. This application is a J2EE application which is build on the J2EE platform. The webpages of Giggle are powered by Java Servlets technology.
- *Database:* The application processes data that is retrieved from and stored on databases.
- *Internet:* The Internet is a data source of Giggle.

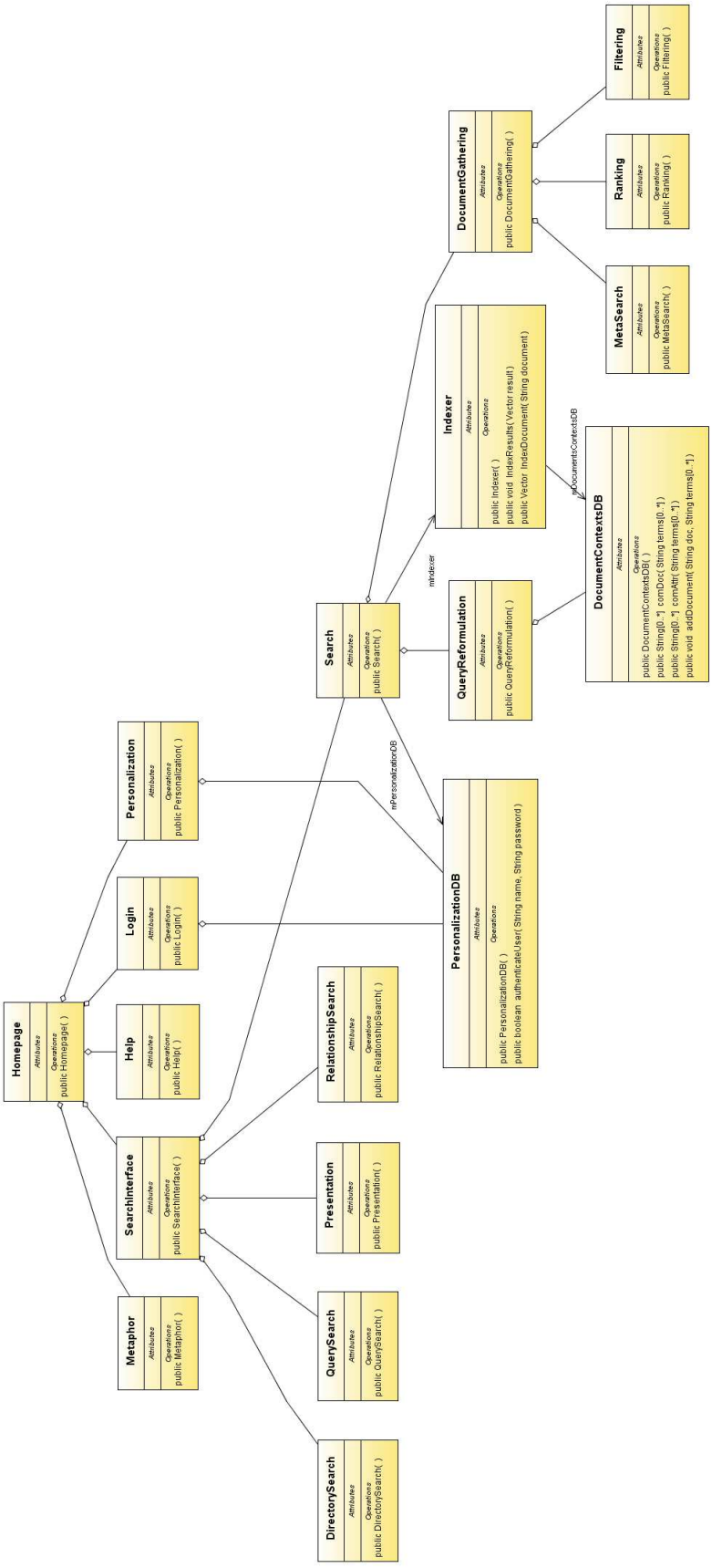


Figure 5.2: Class Diagram

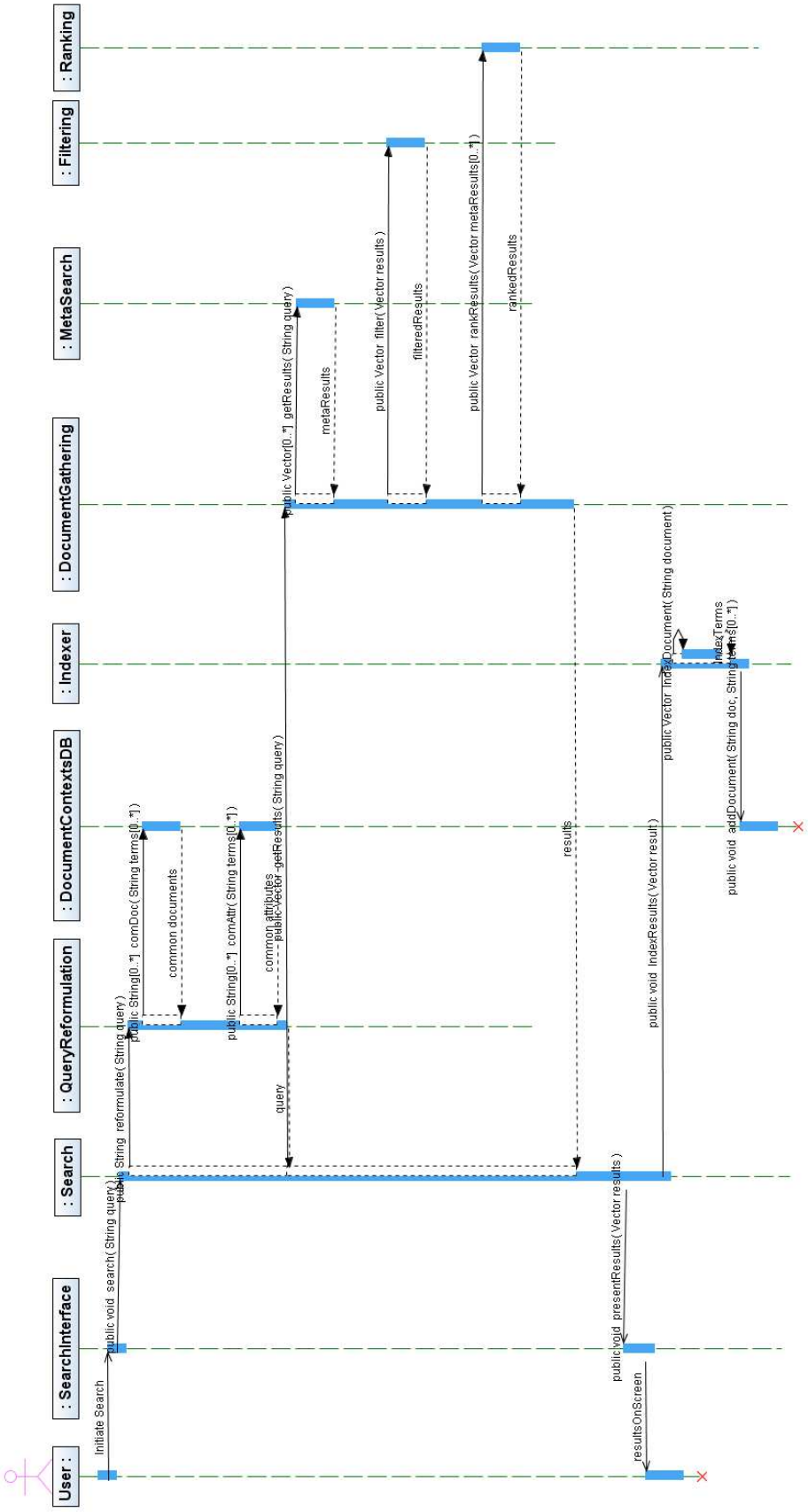


Figure 5.3: Search Sequence Diagram

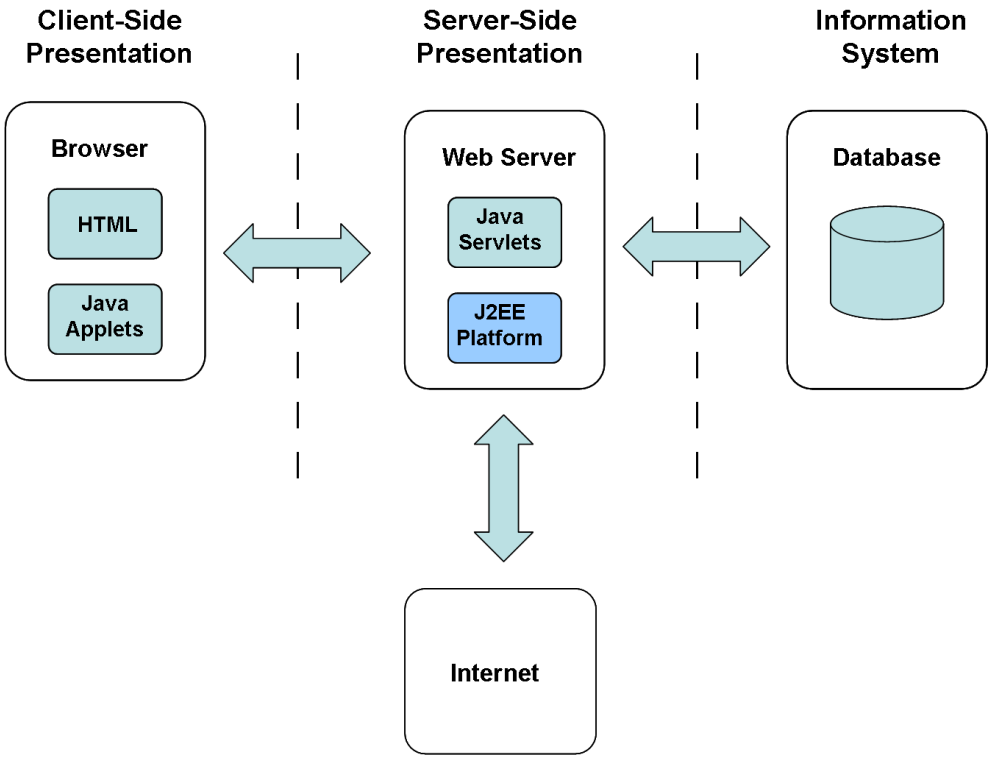


Figure 5.4: Architecture Overview

5.1.5 User Interface Design

The objective of the application is to retrieve search results for the user. The user can search via three methods, and the subsequent search results will be presented on screen. The most important things that need to be shown on screen are the directory, the relationship map, the query input, and the search results. These elements are globally depicted in figure 5.5.

During the implementation we will further work out the visual layout of the user interface by trying out several colors, letter types, banner designs and other visual components like buttons. A global user interface design with the most important components is sufficient for now to begin with the implementation. During the implementation the details of the user interface will become more clear to us and we will then be able to add them in our user interface design.

5.2 Implementation

This section will describe the implementation decisions that were made during the implementation phase. The design models (see section 5.1) have been used as input for the implementation phase. During the implementation some aspects of the design needed to be adjusted, but the requirements of the application were unchanged. This section will conclude with the results of the implementation.

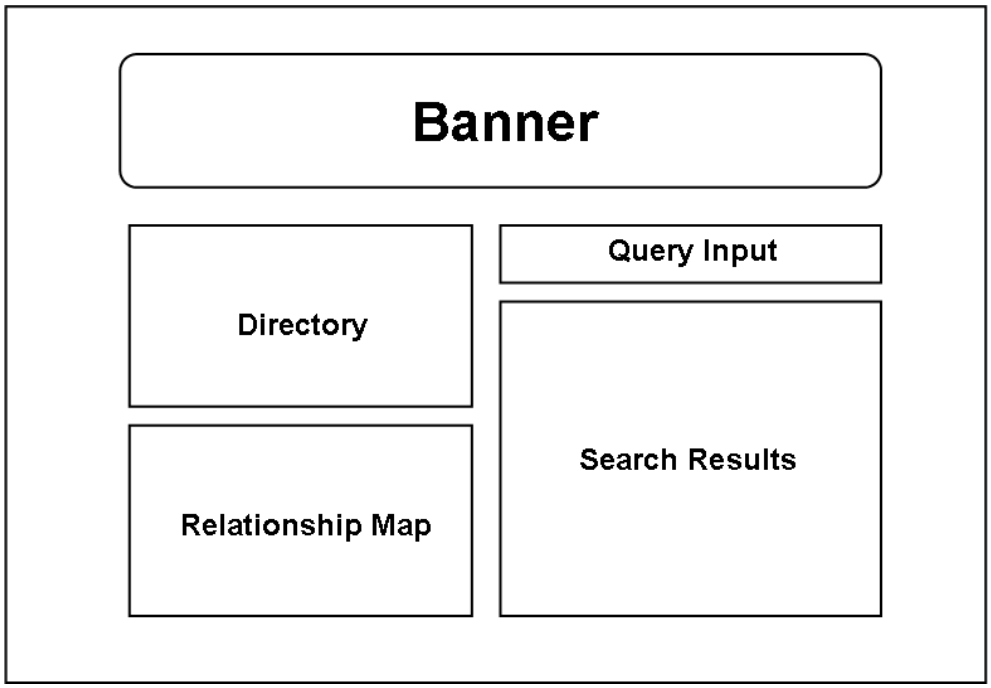


Figure 5.5: Global view User Interface

5.2.1 Implementation Decisions

Integrated Development Environment

To develop our application in an easy and fast method we needed an Integrated Development Environment(IDE). We have searched for an IDE which could be used for the Java language and especially for J2EE. We decided to use Java Studio Enterprise of Sun. This tool has supported us in the design and the implementation of J2EE applications.

Class Implementation

We implemented our classes according to the design described in figure 5.2. Some aspects of the class design have been changed during implementation. The main difference with the original class design is that the class Presentation has disappeared. The tasks of this class were to present the search methods and the results to the user. These tasks are now integrated in the SearchInterface class.

Package Implementation

We have divided our application into two packages: Interface and Search. The Interface package contains all the classes for the representation of the search engine. The Search package contains all the classes needed to reformulate queries and to return search results.

System Implementation

The application is implemented with the J2EE platform. It is a web application that runs on System Application Server Platform Edition 8.2, a Java server program. The application also needs a database to store and access data. We have chosen for the Oracle 9i database. Oracle is a database system that is used by many companies. It is a database system with many possibilities and it has proven to be very stable in combination with large software systems. The Oracle database system is sufficient enough for our application.

5.2.2 Results

We have implemented a meta search engine from which the user can start an Internet search in three ways : directory, relationship map and query input. The search engine has an automatic save option. When a user logs in, the search engine will automatic save the last five search tasks. Whenever the user logs on, the last five search tasks will be restored.

We have implemented a query reformulation function that will reformulate the queries provided by the user by one of the three methods described above. This functionality can help children in formulating their search tasks.

5.3 Testing

Before the application is delivered to the customer we need to be sure that it works properly according to the requirements. This is where testing comes into the picture. Testing is the process in software engineering in which we check if the product is built correctly, but also to check if we have built the correct product. Testing means we are going to execute the application before it gets to the customer with the intention to find and remove all errors. The tests must be conducted systematically to be able to find the highest possible number of undiscovered errors. Testing begins in the small and it will progress towards testing in the large. Our application consists of components and these components will be integrated into clusters of components and finally these components will be put together as the whole system. Therefore we will begin testing the components and expand these tests towards testing the system as a whole.

The next sections will discuss the process of testing and how they apply to our application. This section ends with the results and the conclusion [57].

5.3.1 Testing Process

The process of testing exists on a global level of the following activities: Unit Testing, Integration Testing, System Testing and Validation. These activities will be discussed in this section.

Unit Testing

Unit testing means verifying the correctness of the smallest units of the application, software components or modules. With unit testing the code is tested on its correctness within the class or module. Things that can be thought about are executing every

line of code, checking the boundaries of conditional statements and checking the use of variables.

Unit testing of the different classes was performed in line with implementation. Every class and function was tested after it was implemented to ensure that all units work correctly.

Integration Testing

Separate components will form the whole system. Integration testing will test if these components work properly when they are integrated with each other.

Our application consists of two packages which all together form the application. The communication between these packages was limited to a minimum to avoid integration faults. We have tested the packages separately during the implementation.

System Testing

The application is an element of a larger computer-based system. Our application will run on an operating system and the user can have access to the application through a web browser. With system testing we will try to test whether the application works correctly with its environment.

We have tested our application on several operating systems and web browsers.

Validation

A software application is developed to serve the user. A developer can develop a perfectly working product, but if it doesn't work the way the user expected it, the application will be worthless. The validation testing phase will verify if the application conforms with the user requirements. In other words the purpose of this phase is to check if we have built the correct product.

To validate our product we will create several use-cases to test the application.

5.3.2 Results

The application serves as a prototype for our ideas that were the result of the research of this project. It is important for us to see how our ideas work and if they conform to the user requirements. Therefore the main focus went to the validation and system testing phases. We have chosen to develop use cases and apply those use cases to our application. Besides our main focus on the validation phase we didn't neglect the other phases. We have done some smaller tests within the other phases. For instance during the implementation of the classes we have tested the classes and the packages separately before we integrated them into one system. We also did some integration testing before we focused on the system tests and the validation tests.

Unit Testing

During the class tests in the implementation phase some coding errors were discovered. These errors varied from wrong usage of variables to errors in conditional statements. Once these errors were discovered, they were fixed without any major problems.

Integration Testing

The classes were integrated step by step. We have used stub classes instead of the real classes when they were not ready for integration. We have encountered some errors and we were able to solve them very fast.

System Testing

We have tested the application with three operating systems: Solaris, Linux(RedHat) and Windows. We have also tested the application with the web browsers Internet Explorer, Mozilla Firefox and Netscape. Most of our users will be using Windows as operating system. Therefore our focus goes to combinations of different web browsers with the Windows operating system.

The application seems to work well with the several web browsers on Windows. We have tested the application on several computers and not all computers are standard enabled with Java. In those cases we needed to install Java to be able to run the applet in the application.

We encountered a small problem with Internet Explorer. It seems that before the user can use an applet, the user has to click on it to get the focus on the applet. This problem only occurs with Internet Explorer version 6 or higher. It turns out that Microsoft had a disagreement with a company named Eolas. This company has a "Applet" patent for automatically invoking external application providing interaction and display of embedded objects within a hyper media document. Microsoft did not buy a license from Eolas, so they had to install a patch into Internet Explorer to prevent it from automatically invoking the applet and other objects that belong to the patent [54].

We have also done some stress testing, which means that we have subjected the application to controlled amount of stress. This means using the application under stressful conditions like doing a search task multiple times, or doing multiple search tasks as fast as possible. The application passed those tests very well. The relationship map and the directory structure seem to work very well under stressful conditions. During the system tests we have encountered a problem with the login functionality, but it was a small error that was fixed very fast.

Validation

We have tested the application with several use-cases. These use-cases are based on the usage purposes of the user. The common aspect of all the purposes is that the user wants to find information, but information might be needed for school, or the user might want to find games. We have formulated the following types of use-cases:

- School
- Entertainment
- Infotainment

All use-cases were used with the application with different variables applied. No major problems were encountered, therefore the application passed the validation phase.

5.3.3 Conclusion

The unit testing and integration testing revealed several coding errors in an early state. The early discovery of the errors made it possible for us to solve these errors quickly before the impact of the errors could become large. The unit and integration level also made it easier for us to locate the error.

The system tests revealed that our application does not work optimal with all operating systems in combination with several web browsers. The most important thing for us is that it works well with web browsers in combination with the Windows operating system, because most of our users will be using this operating system. We encountered a small problem with the applet, and this problem only occurs with the browser Internet Explorer.

The results of the execution of the use-cases are quite good. The main problem is the content of the relationship map. Some relations don't seem to be relevant with the users interest. We need to take a closer look at that. But when searching with queries the results seem to be good.

6 Discussion

The goal of this research was to identify the requirements needed to successfully implement a search engine for children. This chapter concludes the thesis with the results of this project with respect to this goal.

6.1 Approach

The first step taken was to identify the requirements which a search engine for children has to possess. Most of the data for this requirements analysis came from the preceding project of IBM [78]. This analysis contains all the functional and non-functional requirements the search engine has to satisfy.

Second we needed methods to classify and find documents for our target group. We developed methods to extract search results from existing search engines and to combine them into one ordered list of search results. This way we saved ourselves much bandwidth, computational power and disk space, which is all required if your search engine contains its own spider and database. Besides that these search engines all have very sophisticated indexing, filtering and ranking methods, our combining technique takes advantage of all these methods and delivers results that exceeds them. To meet our target group we developed methods to assist them in formulating their information need (query reformulation, directory search, relationship map) and handling the results (former query tabs, save search).

To make the search engine a success we had to investigate how to make money with the search engine. We did this by finding the characteristics for the market for Internet search engines. We made a representing selection of search engines and uncovered the ways in which they made money. This gave us a good vision on the possibilities in the search engine market.

Finally we developed the search engine itself. We translated the requirements developed methods into a formal design model. This design specification was systematically translated into code. All units of code were tested separately and the whole implementation was tested based on use cases.

6.2 Scientific Relevance

This thesis was written as a part of our master graduation project. Our master education focuses on two fields of science, namely Computing Science and Management & Technology. In our research proposal we stated why this research is relevant for

both fields of science. The following sections will recap how these aims have been achieved.

6.2.1 Computing Science

For computing science purposes we introduced a model to represent a search engine as a learning method for children. To help them in asking questions (submitting queries) to learn, we developed the query reformulation technique. This technique revises the children's queries to better represent their information need. For this technique we made use of the formal concept theory. We had already encountered the formal concept theory during our education, but in this research we were able to really apply this theory on something practical.

For the retrieval of results we designed a meta-search engine. It was interesting to develop mechanisms to extract and combine results from several other search engines. Also working with new developing methods and tools like J2EE and Java Studio Enterprise was challenging, but that is something we should be able to learn easily as master graduands and we did, without major problems.

6.2.2 Management & Technology

The management & technology part of our research consisted of a survey on several Internet search engines. This survey increased our insight and experience in performing a structured business research. We interviewed managers of search engines, read annual reports, analyzed websites and read articles about our selected search engines. All these activities gave us insight in the way these companies do their business. We were able to formulate business models of these search engines and to describe the way how they are making money.

6.3 Reflection

This section gives an indication of the division of tasks in the project by both project owners. In our opinion the contributions of both project owners were comparable and many tasks were executed together. We worked together most of the time and checked each other's work on content and quality. However, for some aspects of the projects the emphasis lay primarily with one of us. The following self-reflections have been written by each project owner individually and describe their input and role in the project.

6.3.1 Reflection Koen

Responsibilities

Within this project we were both responsible for different parts of the project. I will describe my tasks for the different phases of the project:

- *Initialization Phase:* Contact with client. Wrote part of the research proposal. Designed the lay out for the thesis.
- *Research Phase:* Conducted research on various information retrieval techniques and search engine technology, studied and applied the formal concept

theory, and documented the results. Was responsible for the information retrieval chapter. I also wrote part of the business research chapter.

- *Design Phase:* Designed the meta-search and query reformulation techniques. Worked on the class diagrams.

- *Implementation Phase:* Implemented the search part of the search engine, including the meta-search, query reformulation, indexing, filtering and ranking classes. Was responsible for the database related classes. Designed part of the interface, including the login and savesearch classes. Collected data for the directory and relationship map interfaces.

- *Evaluation Phase:* Performed tests on the final product and fixed all the bugs in the application. Wrote the discussion.

Experiences

Looking back at this project it was a great experience. A global goal of the graduation project is to show my academic capabilities which I learned in the past years. But this project was a learning experience in it self.

I had never worked on a project of this scale before. Because it was a big project everything had to be done a lot more systematic than with smaller projects I was used to. We had to plan activities in time and were forced to perform each task after another to keep the overview. Another consequence of the size of this project was that it was performed by two people. Therefore we could support each other in many tasks, but it was also necessary to communicate very well. We sometimes worked on the same documents and shared files both of us were working on, but since we made good agreements this never led to any problems.

Writing this thesis was also a learning experience to me. I learned to systematically set up a research and to document it in line with all the activities. Thereby I also learned writing reports in LaTeX, which is in my opinion a great progression in comparison with for example Word.

The research on the various information retrieval techniques was a very interesting experience to me. I especially learned a lot from my study on formal concepts. I already known this theory from some lectures during my education, but to really investigate the subject yourself and to develop a practical usage for it was an enlightenment.

The development phase was the last part of the project. We were aloud to design and implement everything we had devised to satisfy the requirements, this was a lot of fun. I learned some new techniques to build web-applications, for example J2EE. To create a design around our ideas, translate this design into code, build the complete application and to see it come out of the tests well gives a good feeling. It really gave me the feeling we created something useful next to our scientific research we described in the thesis.

Conclusion

I can conclude that in my opinion, and I think Khanh feels the same, this project was a great success. We did some interesting research in two different fields of science and brought it together. Above that we developed an application that can really guide

children in there journey to discover the Internet. I can only hope that everyone who will use it in the future will have as much fun using it as we had developing it.

6.3.2 Reflection Khanh

Responsibilities

Within this project we were both responsible for different parts of the project. I will describe my tasks for the different phases of the project:

- *Initialization Phase:* Contact with client. Wrote together with Koen the research proposal. Was responsible for the requirements review.
- *Research Phase:* Conducted research on business models, the information market and survey method. Searched for information on search engines and analyzed the collected information. Was responsible for the business research chapter and co-wrote the information retrieval chapter.
- *Design Phase:* Designed the user interface. Worked on the use-case view.
- *Implementation Phase:* Implemented the interface classes (directory, relationship-map).
- *Evaluation Phase:* Performed tests on the final product. Was responsible for the development chapter.

Experiences

It was the first time that I had to think, plan and execute a project of this size. It was very difficult in the beginning because I didn't know how to handle such a project. But step by step and with the collaboration with Koen we managed to formulate an idea and project plan. Formulation is one step, but execution is the other important step. It became clear to me that it is not always possible to execute the activities according to plan. Because of several reasons we had to adjust our plan during the project. I didn't see this as something negative, but it gave me more experience on how to handle a project of this size.

I gained more experience on setting up a research in the business research part of this project. Doing research and collecting information from people who you had to find, make contact with and convince them to talk to you is not an easy task. It was a great experience for me to actively approach people and to convince them to give you information. I have followed some management courses, but the research field of management was still new for me. It was very interesting and instructive to see the characteristics of an academic discipline different from computing science.

For other courses I have written course reports, but this was my first experience in writing a thesis. I have experienced periods when it was very difficult for me to write something on paper, but I have learned to stay focused and to be persistent. The fact that we had chosen to write this thesis in English instead of Dutch did not make things easier, but both of us wanted to learn something new. We had also chosen to use a different text processing program, which would also be new for us. I had learned to write reports in LaTeX, which is commonly used by the members of computing science.

I have also got more experience on how to design, implement, and test an application. It was interesting for me that next to the theoretical work, I got the chance to do something practical like developing a software application. I already got some experience in developing a software application from other courses, but I have never developed a web application before. I have learned new methods and techniques for developing a web application, like J2EE.

Finally I want to say that for myself I have learned a lot about working together with a partner, in this case Koen. It is common to do a graduation project by yourself, but in the future it will be very common to work together with colleagues on a certain project. The choice to already experience working together with someone in a project of this size was very useful and good for me.

Conclusion

To summarize the whole project: Interesting, instructive, and a lot of fun. It was a very interesting and fun project for me, because it was more or less the first project where I had to actively think, plan and execute a project of my own. Together with Koen I was responsible for the whole project. Our supervisors have given us advice, but in the end we had to do everything for ourselves. This was very instructive for me and I think that the experience I have gained in this project will certainly help me in my future work environment. I am very pleased with the result. Besides the important theoretical research part of this project, I got the chance to try out my ideas from that research part into an application. This gave the project an extra interesting and fun dimension. From beginning until end, it was a lot of fun for me.

References

[1] W.B. Arthur. Increasing returns and the new world of business. *Harvard Business Review*, pages 100–109, July–August 1996.

[2] AskJeeves. Ask jeeves annual report 2004. `http://www.hoovers.com/free/co/secdoc.xhtml?ID=59572&ipage=3334989&doc=0&attach=on`, December 2004.

[3] Eric Auchard. Yahoo seeks growth by uniting ad strategies. `http://today.reuters.com/news/newsArticleSearch.aspx?storyID=141455+02-Dec-2005+RTRS&srch=Yahoo`.

[4] Ricardo A. Baeza-Yates and Berthier A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[5] Dania Bilal and Joe Kirby. Differences and similarities in information seeking: children and adults as web users. *Inf. Process. Manage.*, 38(5):649–670, 2002.

[6] P. van Bommel, B. van Gils, H.A. (Erik) Proper, Th.P. van der Weide, and Mario van Vliet. The information market, its basic concepts and its challenges. LNCS, Berlin, Germany, EU, 2005. Springer. Accepted for the 6th International Conference on Web Information Systems Engineering (WISE'2005).

[7] P. van Bommel, B. van Gils, H.A. (Erik) Proper, Th.P. van der Weide, and Mario van Vliet. Value and the Information Market, 2005. Submitted to: Data & Knowledge Engineering.

[8] J. Borse, E. Robles, and N. Schwartz. Designing for kids in the digital age: Summary of research and recommendations for designers. 2002.

[9] boutell.com. How many search engines are there? last checked: October 18, 2005. `http://www.boutell.com/newfaq/misc/howmanysearch.html`.

[10] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[11] T. Budd. *Understanding Object-Oriented Programming with JAVA*. Addison-Wesley, 2000.

[12] C. Castillo. Effective web crawling. *SIGIR Forum*, 39(1):55–56, 2005.

[13] C.H. Chang and C.C. Hsu. Integrating query expansion and conceptual relevance feedback for personalized web information retrieval. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, pages 621–623, Amsterdam, The Netherlands, 1998. Elsevier Science Publishers B. V.

[14] H. Chesbrough and R. Rosenbloom. The role of the business model in capturing value from innovation: Evidence from xerox corporation's technology spinoff companies. *Industrial and Corporate Change*, 11(3):529–555, 2002.

[15] J. Cho, H. García-Molina, and L. Page. Efficient crawling through URL ordering. *Computer Networks and ISDN Systems*, 30(1–7):161–172, 1998.

[16] M. DeBell and C. Chapman. Computer and internet use by children and adolescents in 2001. Technical report, National Center for Educational Statistics, 2003.

[17] P. Dool and A. Brummelhuis. Compileren van kennis voor ict-rijk leren. Technical report, Ict op school, 2003.

[18] D. Dreilinger and A.E. Howe. Experiences with selecting search engines using metasearch. *ACM Transactions on Information Systems*, 15(3):195–222, 1997.

[19] M.J. Earl. *Rethinking Management Information Systems*, pages 161–174. New York, NY: Oxford University Press, 1999.

[20] B. Edelman. Empirical analysis of google safesearch. 2003.

[21] T.R. Eisenmann. *Internet Business Models*. McGraw-Hill/Irwin, 2002.

[22] P.B. Evans and T.S. Wurster. Strategy and the new economics of information. *Harvard Business Review*, pages 70–82, September–October 1997.

[23] N. Gandal. The dynamics of competition in the internet search engine market. *International Journal of Industrial Organization*, 19:1103–1117, 2001.

[24] B. Ganter, R. Wille, and C. Franzke. *Formal concept analysis: mathematical foundations*. Springer, 1999.

[25] B. van Gils, H.A. (Erik) Proper, and P. van Bommel. A conceptual model for Information Supply. *Data & Knowledge Engineering*, 51:189–222, 2004.

[26] Google. About google. `http://www.google.com/intl/en/about.html`.

[27] Google. Safesearch. `http://www.google.com/safesearch_help.html`.

[28] Google. Google annual report 2005. `http://investor.google.com/pdf/2005_Google_AnnualReport.pdf`, December 2005.

[29] Google Web APIs. Develop your own applications using google. `http://www.google.com/apis/`.

[30] Grokker. A new way to look at search. last checked: October 18, 2005. http://www.grokker.com.

[31] F. Grootjen and Th.P. van der Weide. Conceptual relevance feedback, 2002.

[32] F. Grootjen and Th.P. van der Weide. Conceptual query expansion, January 2004.

[33] A. Gulli and A. Signorini. Building an open source meta-search engine. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 1004–1005, 2005.

[34] A. Gulli and A. Signorini. The indexable web is more than 11.5 billion pages. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 902–903, 2005.

[35] S.G. Hirsh. How do children find information on different types of tasks? children's use of the science library catalog. *Library Trends*, 45(4), 1997.

[36] L. Huang. A survey on web information retrieval technologies, 2000.

[37] IAC. Interactivecorp annual report 2004. http://library.corporate-ir.net/library/19/190/190013/items/158619/IAC_10K2005.pdf, December 2004.

[38] Ilse Media. http://www.ilsemedia.nl/.

[39] P.T. Jaeger, J.C. Bertot, and C.R. McClure. The effects of the childrens internet protection act (cipa) in public libraries and its implications for research: A statistical, policy, and legal analysis. *Journal of the American Society for Information Science and Technology*, 55(13):1131–1139, 2004.

[40] B.J. Jansen and A. Spink. An analysis of web documents retrieved and viewed. *Internet Computing Conference*, 2003.

[41] KartOO. Last checked: October 18 2005. http://www.kartoo.com.

[42] S. Kirsch. Infoseek's experiences searching the internet. *SIGIR Forum*, 32(2):3–7, 1998.

[43] A. Large, J. Beheshti, and C. Cole. Information architecture for the web: the ia matrix approach to designing children's portals. *J. Am. Soc. Inf. Sci. Technol.*, 53(10):831–838, 2002.

[44] S. Lawrence and C.L. Giles. Searching the World Wide Web. *Science*, 280(5360):98–100, 1998.

[45] S. Lawrence and C.L. Giles. Accessibility of information on the web. *Intelligence*, 11(1):32–39, 2000.

[46] C.-S Lee and N.S. Vonortas. Business model innovation in the digital economy. *Social and economic transformation in the digital era*, pages 164–181, 2004.

[47] LookSmart. Looksmart annual report 2005. http://www.shareholder.com/Common/Edgar/1077866/1193125-06-54740/06-00.pdf, December 2005.

[48] L. J. Magid. Child safety on the information highway. last checked: October 18, 2005. http://www.safekids.com/child_safety.htm, 2003.

[49] P. Mayr and F. Tosques. Webometrische analysen mit hilfe der google web apis, 2005.

[50] W. Meng, C. Yu, and K. Liu. Building efficient and effective metasearch engines. *ACM Computing Surveys*, 34(1):48–89, 2002.

[51] M. Mitra, A. Singhal, and C. Buckley. Improving automatic query expansion. In *Research and Development in Information Retrieval*, pages 206–214, 1998.

[52] MSN Developer Center. Msn search. http://msdn.microsoft.com/msn/msnsearch/.

[53] J.P. Mueller. *Mining Google Web Services: Building Applications with the Google API*. 2004.

[54] R. Naraine. Eolas:changing ie user experience a shame. last checked: June 9, 2006. http://www.eweek.com/article2/0,1895,1944867,00.asp, 2003.

[55] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[56] F. Pereira and E. Fife. Meeting consumer needs on the internet: Successful business model. http://www.marshall.usc.edu/ctm/publications/busmodels.pdf, 2000.

[57] R. Pressman. *Software Engineering, A Practitioner's Approach, European Adaptation*. 5th edition, 2000.

[58] M. Rappa. Managing the digital enterprise: Business models on the web. http://digitalenterprise.org/models/models.html, 2000.

[59] E.S. Raymond. *The Cathedral and The Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. OReilly and Associates, Inc., 1999.

[60] J.F. Rayport. The truth about internet business models. Strategy & Business, 1999.

[61] J.F. Rayport and J.J. Sviokla. Exploiting the virtual value chain. *Harvard Business Review*, pages 75–85, November–December 1995.

[62] Reuters. Yahoo using online behavior to target ads. http://today.reuters.com/news/newsArticleSearch.aspx?storyID=16113+02-Dec-2005+RTRS&srch=Yahoo.

[63] S.E. Robertson and K.S. Jones. Relevance weighting of search terms. *Journal of the American Society for Information Sciences*, 27(3):129–146, 1976.

[64] Salfeld. Child control 2005. last checked: October 18, 2005. `http://www.salfeld.com/software/childcontrol`, 2005.

[65] G. Salton. *The Smart Retrieval System–Experiments in Automatic Document Processing*. Prentice Hall Inc., Englewood Cliffs, New Jersey, 1971.

[66] G. Salton, E.A. Fox, and H. Wu. Extended boolean information retrieval. *Communications of the ACM*, 26(11):1022–1036, 1983.

[67] Sanoma. Sanomawsoy annual report 2004. `http://www.sanomawsoy.fi/kuvat/images/attachments/SanomaWSOY_Annual_report_2004.pdf`, December 2004.

[68] Rich Scholes. Uniquely google. `http://www-db.stanford.edu/pub/voy/museum/google.htm`.

[69] SearchEngineWatch. Search links. last checked: November 15, 2005. `http://searchenginewatch.com/links/`, 2005.

[70] C. Sherman. Exploring search engine overlap. `http://searchenginewatch.com/searchday/article.php/3344081`, April 2004.

[71] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999.

[72] I. Sommerville. *Software Engineering*. Addison-Wesley, 6th edition, 2001.

[73] D. Sullivan. comscore media metrix search engine ratings. `http://searchenginewatch.com/reports/article.php/2156431`.

[74] D. Sullivan. Yahoo renews with google, changes results. `http://searchenginewatch.com/sereport/article.php/2165081`, October 2002.

[75] D. Sullivan. Who powers whom? search providers chart. last checked: November 15, 2005. `http://searchenginewatch.com/reports/article.php/2156401`, 2005.

[76] R.S. Taylor. Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29:178–194, 1968.

[77] P. Timmers. Business models for electronic markets. *Electronic Markets*, 8(2):3–8, 1998.

[78] R. Weeda and J. van Gendt. Giggle: Internet search engine for kids. Master's thesis, Radboud University Nijmegen, 2005.

[79] Th.P. Weide. Lecture on information retrieval. RU Nijmegen, 2005.

[80] Wikipedia. Wikipedia, the free encyclopdia: Search engine. last checked: October 19, 2005.
http://en.wikipedia.org/wiki/Search_engine.

[81] K.E. Wolff. A first course in formal concept analysis - how to understand line diagrams. *Advances in Statistical Software 4*, pages 429–438, 1994.

[82] WordNet 2.0. survey.
http://www.wordreference.com/definition/survey, 2003.

[83] J. Xu and W.B. Croft. Query expansion using local and global document analysis. In *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4–11, 1996.

[84] Yahoo. Company information.
http://docs.yahoo.com/info/.

[85] Yahoo. Safesearch.
http://help.yahoo.com/help/in/ysearch/basics/basics-16.html.

[86] Yahoo. Yahoo! 2005 annual report.
http://www.shareholder.com/shared/dynamicdoc/YHOO/1183/YHOO.pdf, December 2005.

[87] Yahoo! Search Web Services. Create applications using yahoo! search.
http://ws1.inf.scd.yahoo.com/search/index.html.

A Business Research

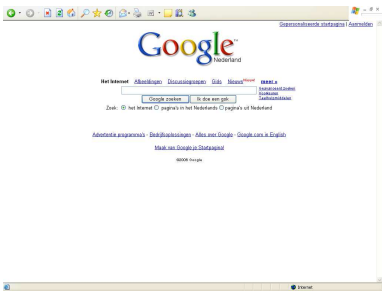
The relevant information about the search engines that are included in our business research is presented in this appendix. The information about each search engine is grouped by its source.

A.1 Google

Google Inc. is the company behind the Google Search Engine. This company was founded in 1998 by two Ph.D. students at Stanford, Sergey Brin and Larry Page. Since then Google has become the worlds largest search engine. They have indexed more than 8 billion web pages [28, p.2]. Their core business is their web search engine and their advertising network. Their mission is to organize the world's information and make it universally accessible and useful. Google believes that offering a high-quality user experience will lead to increased traffic and a strong word-of-mouth promotion. [26, 28]

A.1.1 Google Website

The layout of the Google website (www.google.com) is very simple. There are no ads. Centrally placed on the website are the logo of Google and the search box where the user can input his search query. Google has placed links around this search box to a number of services like image search and news search. Under the search box there is a link to About Google, where the user can find information about Google.



On the website Google gives an answer to the question: "How does Google make money? (investor.google.com)

> We make most of our money from the sale of advertising, which appears adjacent to our own search results and on sites maintained by a large network of publishers across the web that participate in our AdSense program. These ads, known as AdWords, are automatically matched to the content of the page on which they appear by our proprietary software. On Google properties, ads are always clearly identified as such and are physically separated from the algorithmically derived search results.

> Google does not sell placement within our actual search results. We believe paid advertising can provide information as useful as search results, but that users should always be able to distinguish between the two.

Advertisers can advertise on Google through Google AdWords, an auction-based advertising program that enables advertisers to deliver relevant ads targeted to search results or web content. This means AdWords will only show ads to people seeking information related to what the advertisers are selling. AdWords will automate the process of putting an ad on the Google Network. The advertiser can create an AdWords account, where it can specify the ad that Google will place immediately on its network. Google does not accept pop-up ads or ads that take over the web pages. Ads provided by Google are a combination of words and a link on which the user can click to go to the website of the advertiser. This form of ads are called sponsored links. The advertiser can define its own ad, choose the keywords associated with that ad, choose where the ad will appear, and define the price that it is willing to pay. The price that the advertiser pays is of influence when ads compete with each other to be placed on a certain website. This auctioning will place the ad of the advertiser who pays the highest price. The more an advertiser pays, the larger the chance of being placed on a website. The advertiser only pays for the ad when a user clicks on it. This is called Pay-Per-Click (PPC).

Google is offering websites the opportunity to generate revenue by displaying relevant ads on their website's content page. Google delivers these ads to the websites. These ads are coming from their network of advertisers. Google will share the revenue that comes from the advertiser with the website where the ad was displayed. These websites make up the Google Network. Websites can register to the Google Network via the program Google AdSense. This program will automate the process of providing ads that are relevant with the website's content. The purpose is that users will find these ads useful.

The core business of Google towards its users is search technology. They provide their users with search services on several areas like web, books, news and images. These services are freely accessible for everyone in the world, and it is available in more than 100 languages and there are 112 international domains besides the Google.com domain. Google also sells its search technology to businesses. Companies can use the search technology of Google to search for corporate documents.

A.1.2 Google Annual Report 2005

From the annual report of Google [28], Google gives the following overview of their business:

> Google is a global technology leader focused on improving the way people connect with information. Our innovation in web search and advertising have made our web site a top Internet destination and our brand one of the most recognized in the world. We maintain the world's largest online index of web sites and other content, and we make this information freely available to anyone with an Internet connection. Our automated search technology helps people obtain nearly instant access to relevant information from our vast online index.
>
> We generate revenue by delivering relevant, cost-effective online advertising. Businesses use our AdWords program to promote their products and services with targeted advertising. In addition, the thousands of

> third-party web sites that comprise our Google Network use our Google AdSense program to deliver relevant ads that generate revenue and enhance the user experience.

The mission of Google is to organize the world's information and make it universally accessible and useful. To achieve this mission they think that the needs of the users should be seen as top priority in their business strategy. Their experience has learned them that a high-quality user experience will lead to increased web traffic and strong word-of-mouth promotion. In their annual report they tell the investor how they provide value to their most important stakeholders: users, advertisers and websites. [28]

Users

The key benefits Google offers to its users:

- *Relevant and Useful Information*
  According to Google, their search technology is of high quality, they deliver relevant and useful search results in response to user queries.They have indexed more than 8 billion pages.
- *Objectivity*
  Google produces results with only the interests of the user in mind. Google does not accept money for search result ranking or inclusion.
- *Global Access*
  Google is accessible for everyone in the world, it is available in more than 100 languages and there are 112 international domains besides the Google.com domain.
- *Ease of Use*
  In their opinion they have created an easy-to-use interface that provides a simple, intuitive way to get the information that the user wants.
- *Pertinent, Useful Commercial Information*
  The search for information often involves an interest in commercial information, for example researching a purchase, comparing products and services or actively shopping. Google provides search services to find products for sale online.

Advertisers

The following quote about advertisement is from the annual report of Google [28]:

> As more people spend additional time and money online, advertisers are increasingly turning to the Internet to market their products and services to consumers. For these advertisers, we offer Google AdWords, an auction-based advertising program that enables them to deliver relevant ads targeted to search results or web content. Our AdWords program provides advertisers with a cost-effective way to deliver ads to customers across Google sites and through the Google Network.

The benefits of AdWords for the advertisers are:

- *Effective Return on Investment*
  Advertisers can achieve greater cost-effectiveness with their marketing budget for two reasons: AdWords shows ads only to people seeking information related to what the advertisers are selling, and advertisers choose how much they pay when a user clicks on their ad.
- *Access to the Google Network*
  Ads are placed on thousands of third-party websites that make up the Google Network, this gives advertisers an increased exposure to people who are interested in the products of the advertiser.
- *Precise Campaign Control*
  The advertiser can have control over most of the elements of their ad campaign. The advertiser can for example target their campaign by city, country, regional area or language.
- *Global Support*
  Google provides customer care service to their advertisers. The service can be via email support, but advertisers with more budget can have an account team of experienced professionals to help them set up, manage and optimize their ad campaign.

Websites

For websites Google is offering the Google AdSense program. This program enables the websites that make up the Google Network to display AdWords ads that are relevant to the search results or content on their pages. The revenue that is generated by Google from ads that are displayed on a third-party website will be shared with these websites.

The key benefits AdSense offers to third-party websites are:

- *Access to Advertisers*
  Websites can easily have access to advertisers that are part of the Google advertiser client base. Websites don't have to maintain advertisement relationships, these are all handled by Google.
- *Improved User Satisfaction*
  There are no pop-up ads or ads that take over web pages. Internet users won't be annoyed. The ads are targeted to the visitor of that website. The ad will be of surplus value to the website.

A.1.3 Internet

On the Internet there are all kinds of websites that focus on the businesses of Google. Not all Internet sources are appropriate as an information source for this business research.

Rich Scholes [68] has written an article named Uniquely Google that was published in the newsletter of Stanford's Office of Technology Licensing. In his article he discusses how Google earns its money. Its approach is two-tiered. First, Google earns its money by posting search-specific advertisements. Secondly, it generates revenue

by searching for other companies that don't have the technology and resources to search the Internet themselves. For instance, the searches on Netscape are executed by Google. Google is paid for its service and receives exposure by being listed on the website of Netscape (http://home.netscape.com/escapes/search).

A.2 Yahoo

Yahoo was founded in 1994 by two Stanford Ph.D. students, David Filo and Jerry Yang. Yahoo started as a guide to keep track of the personal interests of its founders on the Internet. Since then Yahoo Inc. has evolved into a major Internet company. They are providing Internet services to Internet users like the search engine Yahoo! Search, the email service Yahoo! Mail, the news service Yahoo! News and many more. Yahoo! Search is Yahoo's own proprietary algorithmic search technology, which was launched in February 2004. Before that Yahoo partnered with Google for search technology [74]. Furthermore they are also providing tools and marketing solutions for businesses to market to the community of Yahoo users. Their mission is to be the most essential global Internet service for users and businesses. [86, 56]

A.2.1 Yahoo Website

The Yahoo website of the U.S. (www.yahoo.com) contains many links to all kinds of services of Yahoo, for instance Yahoo! Autos and Yahoo! Travel. The website contains banner ads to advertisers and to services of Yahoo itself. Below there is a link to the Company Info.

Yahoo offers all kinds of services to its users. These services are centered around the objective of providing information to users:

> Yahoo!'s top priority is to provide users with the information they want, when they want it, how they want it and on any device. To ensure this, Yahoo! offers a broad and deep array of products and services that provide our users with the power to connect, communicate, and create, access, and share information online. [84]

The website gives information to the possible business partnerships with Yahoo. The most important partnerships are advertising and the affiliate programs. There are also other options like selling your products online with Yahoo and linking users to Yahoo by adding a search box of Yahoo on your own website.

A.2.2 Yahoo! 2005 Annual Report

The following quotes are from the section Business Overview from the annual report of Yahoo [86]:

> Yahoo! Inc., together with its consolidated subsidiaries, (Yahoo!, the Company, our, we or us) is a leading global Internet brand and one of the most trafficked Internet destinations worldwide. We seek to provide

> Internet services that are essential and relevant to users and businesses through the provision of online properties (the Yahoo! Network) to Internet users and a range of tools and marketing solutions for businesses to market to that community of users. Yahoo! was developed and first made available in 1994 by our founders, David Filo and Jerry Yang, while they were graduate students at Stanford University. We were incorporated in 1995 and are a Delaware corporation. We are headquartered in Sunnyvale, California, and have offices in North America, Europe, Asia, Latin America and Australia.
>
> Many of our services are free to our users. We generate revenue by providing marketing services to businesses across the majority of our properties and by establishing paying relationships with our users for premium services. We classify these revenues as either Marketing Services or Fees. Our offerings to users and businesses currently fall into four categories Search; Marketplace; Information and Entertainment; and Communications and Connected Life. The majority of our offerings are available globally in more than 15 languages.

Yahoo recognizes two groups of important stakeholders: Businesses and Users. Both groups are sources of revenue of Yahoo. The businesses are the source for the marketing services revenue(88%), in other words advertising. And premium service users are the source for the fees revenue (12%).

Marketing Services to Businesses

Yahoo says the following about these services to businesses:

> As part of our strategy to provide the most efficient and effective marketing services for businesses, we are committed to providing end-to-end internet marketing solutions for advertisers. There is ongoing growth in the advertising market and an increasing shift in advertisers use of media towards the Internet and away from traditional media.
>
> Our offerings enable marketers to display their advertisements in different formats and in different locations on the Yahoo! Properties and our affiliates websites.

There are three stakeholders involved with marketing services for businesses: advertisers, websites (Yahoo! Network and affiliate websites) and businesses that want to list information on the Yahoo! Network. Yahoo offers the following to its advertisers [86]:

> Advertisers can display graphical advertisements across the Yahoo! Properties on the pages that are viewed by our users. Yahoo! offers a broad range of tools available for online display advertising, including rich media, video and targeting. We work with our advertisers to maximize the effectiveness of their campaigns by optimizing advertisement formats and placement on the network. We also use our targeting capabilities to help advertisers reach their desired audiences by placing contextually relevant advertisements on our pages. For these advertising services, we earn revenue as ”impressions” are delivered. An ”impression” is delivered when an advertisement appears in pages viewed by users.

For the websites Yahoo offers the following [86]:

> Advertising is also provided through a series of search offerings that enable advertisers to display text based links to their websites on the Yahoo! Properties as well as on our affiliates websites. These advertisements are displayed in response to different user actions when a keyword is used in a search query initiated by a user or when specific content is being viewed by a user on the Yahoo! Properties or on the networks of our affiliates. For example, if a user searches using the keyword ”television” in the Yahoo! Search box or the search box on the website of one of our affiliates, links to websites for advertisers selling televisions will appear alongside the algorithmic search results. As another example, if the user is reading an article about interest rates, he or she may be presented with advertising links to websites for mortgage-related advertisers. For these advertising services, we earn revenue when ”click-throughs” occur. A ”click-through” occurs when a user clicks on an advertisers listing. Such advertising services will be referred to as ”search marketing.”

Yahoo also generates revenue through businesses that want to list information on the Yahoo! Network [86]:

> We also provide services that allow businesses to list information on the Yahoo! Properties, which can be accessed by our users. For example, a company can include job postings on our Yahoo! HotJobs site or include information about its business in our Yahoo! Local Enhanced Listing Service. We earn revenues as these listing services are provided.

Offerings to Users

The offerings to users fall into four categories: Search; Marketplace; Information and Entertainment; and Communications and Connected Life.

The Search offerings are services that enable users to search for information, whether from their computer or mobile device. Their vision is to enable users to find, use, share and expand all human knowledge. The search engine Yahoo! Search belongs to this category.

The Marketplace offerings are often the starting point for users who seek to purchase products and services on the Internet. Offerings that belong this category are: Shopping, Travel, Autos, Personals, Jobs and Small Business.

The Information and Entertainment offerings deliver information and entertainment to users. Yahoo offers basic content with no charge to their users, but they also offer some content on a fee or subscription basis. Examples of these offerings are: Yahoo! News, Yahoo! Music and Yahoo! Movies.

The Communications and Connected Life provide users and small businesses a wide range of communication and content services. Some of these services are provided free of charge, but others are on a fee or subscription basis. Some services that belong to this category are: Yahoo! Mail (email service), Yahoo! Broadband (Internet provider) and Yahoo! Mobile Internet.

Yahoo is global player in the Internet business, which means that among their user group there are geographically differences. They have launched numerous geographically and demographically- targeted Web offerings. They provide services in more than 15 languages in over 20 countries.

A.2.3 Internet

On December 1, 2005, Reuters [62] has reported that Yahoo will be using online behavior to target ads. Yahoo tries to boost up the effectiveness of its advertising by targeting ads to users based on their surfing behavior on its site. Yahoo will track down a few types of behavior by its users, including search queries, movement through Yahoo sites and the specific ads clicked. These types of behavior will decide on the fly what ads are most appropriate for a users. This method is different to the Google (and Yahoo) method, which bases the placements of the ads on words searched.

On December 2, 2005, Reuters [3] reported that Yahoo is going to unite its ad strategies to seek growth in advertisement sales. They want to offer major corporate advertisers the possibility to combine text-based keyword search advertisements and graphical, brand-reinforcing ads. Yahoo expects that their rival Google will offer brand-based advertising in the future.

A.2.4 Scientific Articles

Pereira and Fife [56] have done research on successful Internet business models. They have examined the business models of three large and successful brand names on the Web: America Online, Yahoo and Amazon.com. They have said the following about Yahoo and its business model:

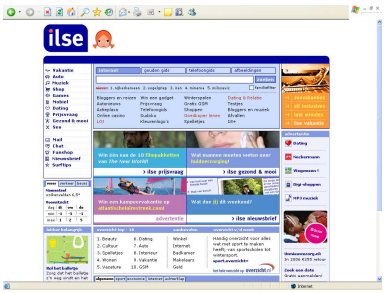
> The business model of Yahoo is based on organizing a massive inventory of pages each of which present a revenue opportunity to acquire sponsorship or advertising.
>
> Since the service is free to users, and depends upon advertising, online commerce must grow along with Internet usage or Yahoos business model will not succeed.
>
> Yahoo's pages are available to users free of charge, with revenue generated mainly through the sale of advertisements, merchandising, sponsorships, direct marketing and promotions.
>
> Attaining scale is vital to Yahoos business model-attracting large numbers of users gives it the power to charge advertisers high prices to display banners.In addition, taking a fee from electronic commerce transactions conducted through its site is a quickly growing source of revenue.
>
> An advantage that Yahoo possesses over most e-commerce sites, such as Amazon, is that it has no inventory, since its focus is on electronics services supported by advertising revenue.
>
> Marketing expenses are the companys largest expenditure-and are expected to remain so. Like all on-line companies that lack a physical presence, Yahoo must keep its name alive through all forms of media.

A.3 Ilse

Ilse was founded in 1996 by Wiebe Weikamp, Merien ten Houten and Robert Klep. Ilse was the Netherlands first Internet search engine. Since then it has evolved into the market leader of Internet search engines in the Netherlands. In 2000 Ilse was acquired by Sanoma Uitgevers (formerly known as VNU Uitgevers). The search engine Ilse

was placed within the company Ilse Media B.V. The strategy of Ilse Media is focused on three areas: search engines and portals, news sites, and youth communities. At this moment Ilse Media is generating approximately 500 million page impressions per month, from which around the 16 million page impressions are generated by the search engine Ilse.nl. [38, 67]

A.3.1 Ilse Website



There is a search box on top of the website of Ilse (http://ilse.nl/). The search box has 4 tabs. These tabs stand for searching in the following fields: internet, gouden gids (yellow pages), telefoongids (phone book) and afbeeldingen (images). On the website there are many links to services of Ilse and its partners. Furthermore Ilse displays banner and textual ads on their website. At the bottom of the website there are links to information about Ilse. Among those links is a link to Ilse Media, where Ilse belongs to.

Ilse Media is active on the Internet and as a publisher of magazines. Among the Internet activities is the search engine Ilse. According to Ilse, Ilse has a high brand awareness (96%) and thousands of loyal fans in the Netherlands. Ilse has 1,425,000 users aged 13 years or older in October 2005. In that periode they received over 16 million page impressions. According to STIR (Stichting Internetreclame, http://www.internetreclame.org/), a foundation that has the objective to promote and stimulate digital marketing by giving insights into online reach of the participants of STIR, Ilse is reaching 14.9% of the Dutch Internet users. Most of the users of Ilse are women of the age category 30+. According to Ilse, their user group is very curious and they are open to advertisements.

On the website of Ilse Media there is information on the advertisement possibilities of its websites. The search engine Ilse displays different types of ads: banners, buttons, text links and skyscrapers.

A.3.2 Press Releases

Ilse has announced in 2003 a cooperation with Gouden Gids (yellow pages). When a user searches on Ilse, Ilse will return relevant results of Gouden Gids in a separate window. The results of Gouden Gids are related to the IP-address of the user, which means that the user will get Gouden Gids results that are in the region of the user.

In 2004 Ilse announced a partnership with Overture (now known as Yahoo! Search Marketing) for the supply of textual ads on the search results page of Ilse. The textual ads are selected based on the search query of a user. When a user searches for coffee, Overture will display coffee ads on Ilse.

A.3.3 Interview

Nils Rooijmans is the head of the Search & Research department of Ilse Media. He is involved in the development of the search engines Ilse.nl and Vindex.nl of Ilse Media.

Who are the stakeholders of Ilse.nl and what are the benefits for these stakeholders?
Ilse.nl has it own team of developers. They have developed their own search technology. The team of Nils Rooijmans is continuously developing and improving the search methods. Besides the technical aspects of a search engine, there is also a marketing aspect. Ilse.nl cooperates with Overture for the supply of sponsored links that are displayed on the results list of Ilse.nl and Vindex.nl. Ilse.nl also displays banners on their website, these banners are coming directly from advertisers. Advertisers can contact Ilse Media for banner space on the websites of Ilse Media. Ilse has commercial partners like Gouden Gids, Achmea and Postbus 51. These partners supply information that Ilse will display to its users dependent on the query of the user.

What are the sources of revenue?
Ilse.nl has three sources of revenue: banners, sponsored links and commercial partners. Advertisers can contact with Ilse Media to place banners on the websites of Ilse Media. Ilse.nl is one of these websites. Many search engines don't display banner ads, they only display sponsored links. An example of such a search engine is Google. Ilse's opinion is that banners are more effective than sponsored links if an advertisers is focusing on branding. This is the reason why Ilse.nl also offers banner ads on their website. The sponsored links are delivered by Overture. What kind of sponsored links are displayed is dependent on the search query of the user. Overture is also responsible for selecting the sponsored links. The third source of revenue are the commercial partners like Gouden Gids (Yellow Pages), Achmea (Insurance company) and Postbus 51 (government's centre for public service information). For example if a user searches for a plumber, Gouden Gids will supply links to plumbers that can be found in the Gouden Gids. Gouden Gids will pay Ilse.nl for linking users to their website.

What are the plans for the future?
Ilse wants to gain market share back from their competitor Google. Their plan to achieve this is to be a typical Dutch search engine. They are focusing on the Dutch market and expect that Ilse can better find information that is relevant to the Dutch Internet user than Google, which is globally active. The advantage of Ilse is that they are based in the Netherlands and they have contacts with Dutch partners like Gouden Gids.

Is there a promotion campaign for Ilse?
A few years back there was a promotion campaign to increase the brand awareness of Ilse.nl. Currently there is no promotion campaign and Ilse.nl relies on word-of-mouth promotion.

What are the differences between Ilse.nl and Vindex.nl?
The search technology behind these search engines are similar to each other. The differences are the layout and services that these search engines provide. Ilse.nl is targeted at a general audience, and Vindex is targeted at a business audience. On Vindex the user can find traffic and stock market information and in comparison with Ilse there are more search services on Vindex.

Is Ilse.nl profitable?

Ilse.nl is a profitable search engine.

Ilse.nl has lost web traffic in comparison to other Dutch websites. According to Multiscope they fell from position 5 in 2004 to position 14 in 2005. What is the reason for this downfall?
Since 2002 Google has gained popularity among Dutch Internet users. The reason for this popularity was that the quality of the search results of Google were better than the search results of Ilse. According to Nils Rooijmans the quality gap between these two search engines has vanished. The popularity of Google is still growing because of the former quality gap. Nils Rooijmans stated that Ilse did not loose web traffic. They still generate as much web traffic as last year. There is only a relative loss of web traffic.

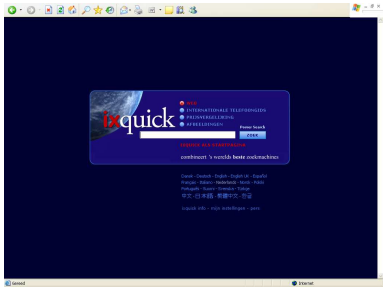
Is Ilse involved in the search engine business for children?
At the moment Ilse is doing research on search engines for children. Their main challenge is to find a model to filter out all the offensive content for children. The problem is that with the current search techniques it is not possible to guarantee that offensive content will not be returned by the search engine. In their opinion there is a business opportunity for children search engines. Advertisers are indicating that they are interested in children as a target group. Children cannot directly buy products and services, but these children are in contact with their parents who can buy these products and services for their children. According to Nils Rooijmans search engines for children are also used as a promotion tool for Internet brands. For example Yahooligans belongs to Yahoo Inc. By attracting young children, they hope to maintain these Internet users for Yahoo when these children get older.

A.4 Ixquick

Ixquick is a meta-search engine that sends user requests to several other search engines and/or databases and returns the results from these search engines. Ixquick is originally an American meta-search engine, but since 2000 Ixquick is owned by the Dutch company Surfboard Holding B.V.

A.4.1 Ixquick Website

The layout of the website of Ixquick (http://www.ixquick.com/) is very modest. There are no ads displayed. The search box is placed in the center of the website. There are 4 radio buttons that represent the search options: web, international phone guide, price comparison and images. At the bottom of the website are links to information about Ixquick.



Ixquick profiles itself as the most powerful meta-search engine in the world. When a user sends a search query to Ixquick, they will send this request to minimal 10 other popular search engines. The user will search in a much bigger part of the Internet than with one individual search engine. The search results are ranked according to their top ten rankings at the different search engines. A web page that has the most top ten rankings at the different search engines will be

returned at the top position of the Ixquick results list.

On the website there is a Question & Answer document. A few interesting questions and answers from this document:

> Q: What is Ixquick? A: Ixquick is a metasearch engine: when you search Ixquick, you are searching many popular search engines, such as Yahoo/AltaVista, Gigablast, Ask Jeeves/Teoma, Open Directory and many others, at the same time. Combined, these engines cover more of the Internet than any one search engine.
>
> Q: What is Ixquicks market share? A: We are a stable, growing and profitable company. In 2004 we handled over 120 million searches. Our focus is growth of search queries, where other search engines focus on commercialization and maximizing profit.
>
> Q: What do Ixquicks financials look like? A: Ixquick is a privately held company, so we do not report financials. We are a financially strong company and profitable for the last 2 1/2 years. It is Ixquicks profitability that enables us to invest heavily in software development.
>
> Q: How does Ixquick make money? A: Contrary to search engines that strongly commercialize results, Ixquick has opted for a clean and simple user-interface with a minimum of advertisements. Some result pages show sponsored results, when relevant, which are clearly indicated at the top of the page and limited to three.

A.4.2 Interview

Alex van Eesteren is vice president of marketing & sales of Ixquick. He is the contact person of Ixquick.

Who are the stakeholders of Ixquick and what are the benefits for these stakeholders?
Ixquick is a meta-search engine that is owned by the Dutch company Surfboard B.V. Ixquick was originally an American meta-search engine that went live in 1998. Surfboard acquired Ixquick in 2000. On March 25 of 2005 Ixquick has launched a complete new version of its meta-search engine. The meta-search engine is developed by a team of 8 developers led by the founder of Ixquick, David Bodnick. This team is based in New York, the United States. Here in the Netherlands the team consist of Robert Beens, managing director, and Alex van Eesteren, vice president marketing & sales. Yahoo! Search Marketing supplies the sponsored links that are displayed on the results list of Ixquick. Ixquick is dependent on other search engines for its search results. There is a cooperation between Ixquick and other search engines like Yahoo, Ask Jeeves and LookSmart. Ixquick does not have to pay any kind of fee to these search engines for making use of their searching services.

What are the sources of revenue?
The goal of Ixquick is to deliver the user a friendly search experience together with high quality search results. To achieve this user friendliness, they think that advertisements should be kept at a minimum. The only form of advertisements that can be found at Ixquick are sponsored links delivered by Yahoo! Search Marketing. These

sponsored links can be found at the top of the page and are limited to two. The sponsored links are also the only source of revenue of Ixquick.

What are the plans for the future?
After the introduction of the renewed meta-search engine, Ixquick's goal is to let the usage of the meta-search engine grow in both the U.S. and internationally. Their focus is growth of search queries, where other search engines focus on commercialization and maximizing profit. At the moment they don't compete with large search engines like Google and Yahoo. They are currently a niche player in the search engine business, but they tend to expand in the future. Ixquick hopes that users will use Ixquick next to a popular search engine like Google. Quality for the users is their main objective.

Is there a promotion campaign for Ixquick?
The budget of Ixquick is not large enough to finance a promotion campaign. Ixquick relies for promotion on press releases and word-of-mouth promotion among Internet users.

What are the differences between Ixquick and other search engines?
According to Alex van Eesteren Ixquick is a qualitative good product. They focus on the user experience instead of on commercialization and maximizing profit where other search engines focus on. They have the opinion that they have advantages over a single search engine. Using multiple search engines for search results makes them search in a much bigger part of the Internet than only with one search engine. Another advantage is that Ixquick can overcome irrelevant results by search engine optimization companies that trick search engine spiders into thinking irrelevant websites are relevant. The foundation for this advantage is that optimization companies can trick one search engine, but it is difficult to trick several search engines at the same time. Ixquick's focus is to return websites that are found in the top ten results list of the search engines. These websites are placed on top of the search results list of Ixquick.

Is Ixquick profitable?
Ixquick is profitable. They have been profitable for the last two and half years. Ixquick generates enough revenue to operate the meta-search engine. The profits enable Ixquick to invest in software development for improvement of the meta-search engine.

A.5 Expedia.com

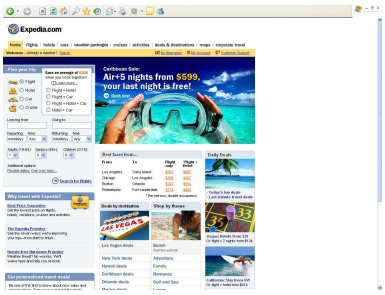
Expedia.com is an Internet portal that provides services to plan a complete traveling trip. Examples of these services are search engines for flights, hotels, cars and cruises. Expedia.com is a U.S. company with activities in the United States, United Kingdom, Canada, Germany, France, Italy, the Netherlands and Australia.

A.5.1 Expedia.com Website

The website of Expedia.com (www.expedia.com) contains at the top of the page 10 links to services like flights, hotels and cars. At the bottom of the website are links to pages that contain information on Expedia, such as advertising on Expedia, affiliate

program, jobs and investor information. From the page Aboute Expedia.com we have found the following description of Expedia.com:

> Expedia delivers consumers everything they need for researching, planning, and purchasing a whole trip. The company provides direct access to one of the broadest selections of travel products and services through its North American Web site, localized versions throughout Europe, and extensive partnerships in Asia. Serving many different consumer segments from families booking a summer vacation to individuals arranging a quick weekend getaway, Expedia provides travelers with the ability to research, plan, and book their comprehensive travel needs. Expedia-branded Web sites feature airline tickets, hotel reservations, car rental, cruises, and many other in-destination services from a broad selection of partners.



Advertising is a source of revenue of Expedia.com. There is a distinction between travel advertisements and non-travel advertisements. Advertisers can list their travel products on Expedia.com, which also offers for more exposure of the listed products an advertising campaign. Examples are banner ads that are displayed on the webpages of Expedia.com. Non-travel advertisers can also advertise on Expedia.com.

Websites can become an affiliate of Expedia.com by selling travel products and services. These affiliate websites have to link their users to Expedia.com. These website earn a commission when users of their website buy travel services on Expedia.com

A.5.2 IAC Annual Report 2004

Expedia.com is part of IAC Travel which is a segment of IAC/InterActiveCorp. IAC is active in deversified businesses. They are operating a portfolio of specialized and global brands in the travel, retailing, ticketing, personals, media, finacial services, real estate and teleservices industries. IAC enables billions of dollars of consumer-direct transactions for products and services via the Internet and telephone. On December 21, 2004, IAC announced its plan to separate IAC into two independent public companies. The new company will hold the travel and travel related businesses of IAC as ”New Expedia”. The other businesses of IAC will continue as IAC/InterActiveCorp. Expedia.com belongs to New Expedia, this company sells stocks on the Nasdaq stock market.

The following description of the business of Expedia.com comes from the annual report of IAC [37]:

> Expedia.com makes a large variety of travel products and services available directly to consumers through its U.S.-based website, www.expedia.com, as well as through localized versions of its website in Canada, France, Germany, Italy, the Netherlands, and the United Kingdom, many of which are leading online travel services companies in their respective coutry. The Expedia-branded websites also serve as the travel channel on

> MSN.com, Microsoft's online services network in the United States, was well as certain international MSN sites. Expedia-branded websites target many different types of consumers, from families booking a summer vacation to individual travelers arranging a quick weekend getaway, in order to provide the vast majority of travelers with the ability to research, plan and obtain their travel needs. Consumers can search for, compare information (including pricing and availability) and book travel products and services on Expedia-branded websites, including airline tickets, lodging, car rentals, cruises and many destination services, such as attractions and tours, from a large number of suppliers, on a stand-alone and package basis.

IAC Travel generates revenue through two business models: merchant business model and agency business model. From the annual report of IAC [37]:

> Under the merchant model, IAC Travel facilitates the booking of hotel rooms, airline seats, car rentals and destination services from its travel suppliers and is, for such bookings, the merchant of record.
>
> The merchant model provides travel suppliers a cost-efficient way to increase marketing and promotion of their brands. Merchant revenue are derived from the difference between amounts paid to the travel suppliers and the amounts paid by the consumer.
>
> Under the agency model, IAC Travel acts as an agent in the transaction, passing reservations booked by its customers to the relevant airline, hotel, car rental company or cruise line. IAC Travel receives a commission or ticketing fee from the travel supplier for its services under the agency model. In the case of agency airline transactions, IAC Travel also receives fees from global distribution systems, GDSs, which control the computer systems through which air travel reservations are booked, in addition to any commissions or ticketing fees paid by travel suppliers.
>
> Agency revenues are derived primarily from commissions and ticketing fees from travel suppliers, revenues from GDSs and fees leisure and corporate customers and are recognized at the time the reservation is booked.

Travel suppliers are important partners of Expedia.com. IAC Travel offers the following benefits to their travel suppliers [37]:

> IAC Travel strives to deliver value to its travel suppliers through a wide range of innovative, targeted merchandising and promotional strategies designed to increase their revenues, while simultaneously reducing their marketing transaction and customer service costs. IAC Travel maintains a supplier relations team, which consists of a staff of account executives and market managers who work directly with travel suppliers to increase the marketing of their travel products through IAC Travel's brands and businesses.

A.5.3 Press Releases

On August 9, 2005, IAC announced the completion of the separation of the traveling activities and the other activities of IAC into two separate public companies. The

traveling activities go under the name of New Expedia, the other activities belong to IAC. Expedia will begin trading stocks as an independent publcly traded company on the Nasdaq National Market under the symbol ”EXPE”.

A.6 FindArticles.com

FindArticles.com is a search engine for publications. They claim that they have 10 million articles from magazines and journals. FindArticles.com is a service provided by LookSmart, Ltd.

A.6.1 FindArticles.com Website

The website (www.findarticles.com) of FindArticles.com contains several links to magazine categories. The search box is placed central on the webpage. The search box has the option to look in all articles, free articles only, and in magazine categories. The categories are: Arts & Entertainment, Automotive, Business & Finance, Computers & Technology, Health & Fitness, Home & Garden, News & Society, Reference & Education, and Sports. A banner is always displayed at the top of the webpage. At the right side of the website banner ads and sponsored links by Google are interchangeable displayed. At the bottom of the webpage are links to the company information and information how to advertise on FindArticles.com

From the website of FindArticles.com comes the following description of their business:

> FindArticles is focused on delivering the best and most essential search results. By working with the best sources, we have assembled all the essential publications covering a wide range of subjects - and are continually adding to our collection.
>
> Our publications and subjects are organized by major categories: Arts & Entertainment, Automotive, Business & Finance, Computers & Technology, Health & Fitness, Home & Garden, News & Society, Reference & Education, and Sports.
>
> Discover exactly what you need at FindArticles, using either browsing or searching techniques. Select a specific publication up front, or start with a general search and then include or exclude publications. Insert new search terms as needed to pinpoint the most relevant results.
>
> FindArticles has articles from thousands of resources, with archives dating back to 1984. That means you get to search for exactly what you need, from millions of articles not found on any other search engine.

Advertisers can contact with LookSmart to advertise on their search sites. FindArticles is a search site that belongs to the family of search sites of LookSmart. LookSmart offers several advertising options such as pay-per-click (PPC), banners and search results listings.

A.6.2 LookSmart Annual Report 2005

FindArticles is a search service provided by LookSmart. LookSmart is a provider of products for advertisers who wish to pay to be included in relevant web search results. The listings of these advertisers are distributed on the network of portals of LookSmart. Some other portals than FindArticles are LookSmart Auto, LookSmart Money and LookSmart Food.

From the annual report we have found the following information [47]:

> Our FindArticles service allows consumers to search a large database of high-quality content from over millions of articles and thousands of publications. This service is accessible to consumers from www.findarticles.com . With the launch of the vertical search sites, we have been able to select appropriate resources by categories and deliver a more relevant search result.
>
> Online advertising revenue primarily decreased in 2005 as a result of the elimination of low converting traffic as we continued our efforts to improve traffic quality and optimize traffic flow from our distribution network partners. Further, revenues also decreased due to the winding down of our agreement with Microsoft, which began in January of 2004. This decrease was partially offset by a 94% increase in revenue from the sale of our consumer products, driven by our FindArticles product. In 2005, we improved our content, which subsequently increased page views and FindArticles revenue. We also began generating revenue from our AdCenter agreements during the second half of 2005, which further offset the overall decrease in revenue.

A.7 Yahooligans!

Yahooligans is a web guide for kids aged 7-12 years. It is a service provided by Yahoo. The websites that are included in the directory are checked by educators to ensure that the content and links are appropriate for the target group. The user can participate into the expansion of the directory by suggesting web sites to Yahooligans.

A.7.1 Yahooligans! Website

Central displayed on the website of Yahooligans (www.yahooligans.com) is a search box. Under this search box is a directory listing. This listing is divided into 6 major sections: Around the World, Arts & Entertainment, Computers & Games, School Bell, Science & Nature, and Sports & Recreation. Guides for parents and for teachers are also available. At the left side there is a menu with links to Yahooligans services like Yahooligans! Music and Yahooligans! Games. At the top of the home page there is a banner. At this moment (December 19^{th}, 2005) there is a banner of Cartoon Network. After surfing through the website, other banner ads are displayed. Every ad that we have seen so far are banner ads for Yahooligans

services like Yahooligans! Music and Yahooligans! Games. The only external ad that we have seen so far is the Cartoon Network ad.

From the website we have found the following description of Yahooligans:

> Yahooligans! is a browsable, searchable directory of Internet sites for kids. Each site has been carefully checked by an experienced educator to ensure the content and links are appropriate for kids aged 7-12. In addition to our directory, we also have a number of fun features, including Yahooligans! Games, Sports and News pages, Movies, Animals, Science and Dinosaurs, a Reference section, Parents' and Teachers' Guides, a Cool page, and lots more.

A.7.2 Yahoo! 2005 Annual Report

Yahooligans is service provided by Yahoo. In their annual report we have found a description of Yahooligans [86]:

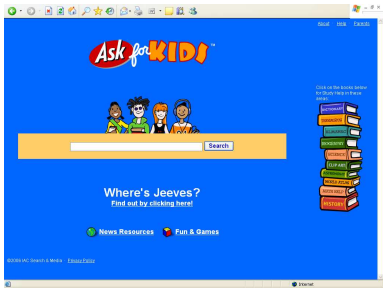
> Yahooligans! is a free entertainment and educational Internet guide designed for children ages seven to twelve. Yahooligans! offerings include games, reference materials and movie information.

Yahooligans is not further mentioned in their annual report. The revenue sources of other services of Yahoo are discussed in the annual report, but there is no information on how Yahooligans generate revenues.

A.8 Ask Jeeves Kids

Ask Jeeves Kids is a special search engine for children. It is different from other search engines. Children can enter a question into the search box. This question is seen as the search query. The difference with other search engines is that children can ask questions to the search engine just like as they would ask their parents or teachers. Ask Jeeves Kids tries to guide the user to the right websites where the user can find an answer to its question. Ask Jeeves Kids is a service of Ask Jeeves Inc.

A.8.1 Ask Jeeves Kids Website



The layout of the website (www.ajkids.com) is kept simple. Centrally displayed on the website is a search box where the user can input a question. Under this search box there is a window where the user can see what other kids are asking at the moment. At the right side of the home page Ask Jeeves Kids offers the user study tools. These tools in the following area's: dictionary, thesaurus, almanac, biography, science, clip art, astronomy, world atlas, math help and history. When a user clicks on one of these topics, the user is actually filling that topic as a search query. For instance typing in *math help* into the search box gives the same results as with clicking on the study tool *math help*.

At the top of the home page there are links to information on Ask Jeeves Kids. One of those links guides the user to the About page of Ask Jeeves Kids. The following information about Ask Jeeves Kids comes from the About page:

> Ask Jeeves for Kids is a fast, easy and kid-friendly way for kids to find answers to their questions online. Designed to be a fun destination site focused on learning and "edutainment," Ask Jeeves for Kids uses natural-language technology that allows kids to ask questions, such as "Why is the sky blue?" or "What's it like to live in space?" in the same way they would ask a parent, friend or teacher. The service combines human editorial judgment with filtering technology to enable kids to find both relevant and appropriate answers on the Web.
>
> Kids have great imaginations and are curious by nature. That's why Ask Jeeves for Kids has created a fast, easy, and safe way for kids to find information and answers online. Ask Jeeves for Kids is primarily an educational Web site that kids can use for homework help and research for school projects. But we've also included fun games and activities.
>
> Each Web site included in Ask Jeeves for Kids is carefully selected by an editor. We include only "G-rated" pages and those written specifically for children. We select sites for the quality and depth of their content, and for safety. This criteria extends to the content of the pages presented on the Ask Jeeves for Kids search results page, as well as the content of pages directly linked to on those results.

The advertising policy is that Ask Jeeves Kids accepts advertising that are appropriate for kids under the age of 13. We have tested the search engine, but we have never found any advertisement.

A.8.2 Ask Jeeves Annual Report 2004

From the annual report [2] we have found the following information on how Ask Jeeves generate revenues:

> We earn revenue primarily as users click on advertisements we display as they navigate the Internet. We refer to users Internet activity as Web traffic and, in general, the more Web traffic we can attract, the more advertising revenue we will generate.

Ask Jeeves generates web traffic in two ways: Proprietary Traffic and Network Traffic. Proprietary traffic is attracting users to the web sites of Ask Jeeves. The more web traffic, the more revenues generated from advertisements. Network traffic is about generating traffic on third parties websites. Ask Jeeves generates revenue by placing advertisements on these third parties websites. Ask Jeeves will collect a fee from the advertiser and gives a portion of it to the website publisher.

Ask Jeeves Kids belongs to the class of destination search sites of Ask Jeeves Inc. These destination search sites contribute to the proprietary revenues of Ask Jeeves Inc.

Ask Jeeves gives the following definition of Destination Search [2]:

> Destination searches occur when an Internet user navigates to a search site, such as Ask.com, in order to submit his or her query on that particular site. To attract destination searchers to our branded sites we employ advertising campaigns and seek to improve the user experience, among other strategies.