Image-based information retrieval using mobile devices

Authors: Ted Möller and Björn Nilsved

Departement of Informatics, Lund University

Master thesis presented January 2006

Supervisor: Konrad Tollmar

Abstract

Image-based information retrieval goes beyond only matching images, as information in other modalities also can be extracted from data collections using image search. Our purpose has been to explore users experience of general mobile photo-based Web searches to find what issues and phenomena it contains. This was achieved in a three-part study by creating and letting respondents test prototypes of mobile photo-based search systems and collecting data using interviews, observations, video-observations, questionnaires, and time-meassurements. Among the issues and questions found, the most important was that a specific image can be submitted to a search engine with many different search questions in mind. Users need ways to specify their information need apart from using the image itself. Suggestions for designs to solve issues like this one have been explored as a part of this study.

Keywords: Filtering, Image retrieval, Image search, Mobile Web search, Photo search

Table of contents

Table of figures

Table of tables

Introduction

Searching the Web, which is full of information of all kinds and subjects, is not a simple task. Users generally find information using search engines - websites that match the users input with existing Web pages using a search algorithm. The input to these search engines often consist of keywords or other written text like a question. But the Web contains not only text, but also information in other modalities such as images. They are however not yet being used as input for general Web searches.

A shift towards ubiquitous computing is evident in the increased sophistication and extended utility of mobile devices, such as mobile phones, PDAs and Pocket PCs (Sweeney & Crestani, 2003). The major difference from accessing the Web using a desktop computer is that handheld devices are mobile and can be used in a larger number of contexts. Further, the input and output facilities are somewhat limited, such as the size of the display and the number of and placement of keys. But compared to a desktop computer the mobile device doesn't only have interaction-wise limitations. Its design and context of use also opens up new possibilities for sound, image and video input.

A lot of effort is being spent on finding compelling mobile Web services such as position-based travel information, online banking or pizza ordering. However, there is little explicit consideration of user interaction requirements and abilities (Jones, Buchanan, Marsden & Thimbleby, 2000).

Imagine submitting an image to a search engine and retrieving hyperlinks to Web pages containing similar images. Then imagine taking a photo with your camera-equipped mobile phone and getting the same functionality anywhere you are (Figure 1). The obvious benefits of such a search system in a mobile context is that you don't need to do the tedious work of entering a query string using the limited text input facilities of most mobile devices. In addition, you may be able to find information about objects and places that are hard or even impossible to describe using only words and you would be able to carry your information search tool with you at all times.

Figure 1. Illustration of a general mobile photo-based Web search.

The research field of image retrieval is about finding images that match in different ways. In this study, we're interested in finding information related to an image. It is not the matching image itself that is sought (as is the main purpose of the image retrieval field) it is the related information found in its vicinity, its context, which is the goal. This is called image-based information retrieval. The information retrieved can be in any format or media, but it's found using an image.

Reasons for using a photo search instead of, or as a complement to, a text-based search is that users' search needs may be vague, and that text as input may be hard to define (McDonald, Lai & Tait, 2001).

The concept of a general mobile photo-based Web search brings many questions to life. How would users perceive such a search system and what would they use it for? Would its availability change users Web searching behavior? How will users interact with such a system? What will it look like?

Creating a general photo-based Web search engine and opening it to the general public on a larger scale is not yet feasible, but this does not make the subject less interesting as it probably will be in the future. Increasing the number of indexed images would make a search engine less acurate, as it is more likely that images will be similar without containg the same object. At this point, using a database of a hundred thousand images is possible, but compared to Google's (http://www.google.com) eight billion indexed Web pages that is not even close (Biever, 2006).

While other researchers continue the work of refining and optimizing image-matching algorithms we explore the interaction sides of mobile photo-based Web searches. We believe this can be studied before the infrastructure is really there and when it is, we will already know about some of the issues that will need to be solved.

Tollmar, Yeh and Darrell (2004) present a project called IDeixis that looks into mobile photo-based Web searches for location related information, which will be discussed later in this thesis. They propose more research into some areas in this field such as new designs for browsing large search results containing thousands of hits, multi-modal interfaces, and image-based object searches. Other researchers like Jones et al. (2000) suggest further research into alternative (and combined) modalities for user interfaces for mobile Web devices in general.

Purpose

The purpose of this study is to explore the area of general mobile photo-based Web searches from a user perspective to find what issues and phenomena it contains. General in this case meaning any search that a user can come up with, as opposed to a specific class of information needs.

We focus on photo-based Web searches in digital camera-equipped mobile devices such as mobile phones. Our purpose is not to design a finished system or define a definitive way of conducting a search. Instead, we're looking for issues that could be interesting for future studies and general suggestions for designs that might solve those issues. To accomplish this we aim to design prototypes and test and evaluate them together with respondents.

This study is not about the technical aspects of how to retrieve matching images from the Web as most things written on image retrieval. Instead, this study is about how to provide better information that helps the user evaluate the results of a search, as stated by Woodruff, Faulring, Rosenholtz, Morrison and Pirolli (2001 pp-198):

> *"The search engine can increase user efficiency by either (1) returning higher-quality document lists (e.g., through better index coverage and ranking algorithms) or (2) providing information that allows the user to evaluate the results more quickly."*

The intention has never been to replace the text-based Web search but to explore an alternative tool that can be used in new situations not entirely supported by traditional search methods, and also for some overlapping searching needs.

Definitions

We make a difference between image search and photo search, where we define image searches as searches using digital images that already exist (e.g. in a database, on the Internet or on other digital media) as a query. We define photo searches more specifically as searches using photos recently taken with digital cameras embedded in devices.

When mentioning a mobile phone, we refer to a mobile device with an Internet connection, Web browsing capabilities, and an embedded camera that people tend to carry with them most of their time. The definition includes some Personal Data Assistants (PDA) as they seem to converge with high-end mobile phones, both in use patterns and in capabilities. In cases where we believe the difference between the two is of greater relevance we will be more precise.

Throughout the study we use the terms: hit, result set, result list and result page. A hit is an item (often a Web page) that a search engine finds matching the input query but it does not necessarily match the users' information need. A result set is the set of all hits that that a search engine finds matching the input query. A result list is a common way of presenting the result set in sequential order. A result list can be divided into many result pages, depending on the design of the user interface.

When referring to services available on the Web we call them by the name commonly used by its users. In this study, this name refers to all functionality provided by that Web site. When referring to a specific service, or a subset of the services, we will also be more specific. We provide a URL in parentheses the first time a service is encountered in the text, so that readers may try them out themselves.

Overview

After this introduction there is a chapter of general background on fields related to the subject. It consists of short reports on different phenomenon and research areas, which are good to be familiar with when following this study.

The study itself consists of three parts that are presented in three chapters. They follow each other in chronological order and present our exploration of the area. Each part is built on the preceding ones so readers are advised to read them in sequence. To conclude we collect important findings from all parts of the study and present them in the last chapter (Figure 2).

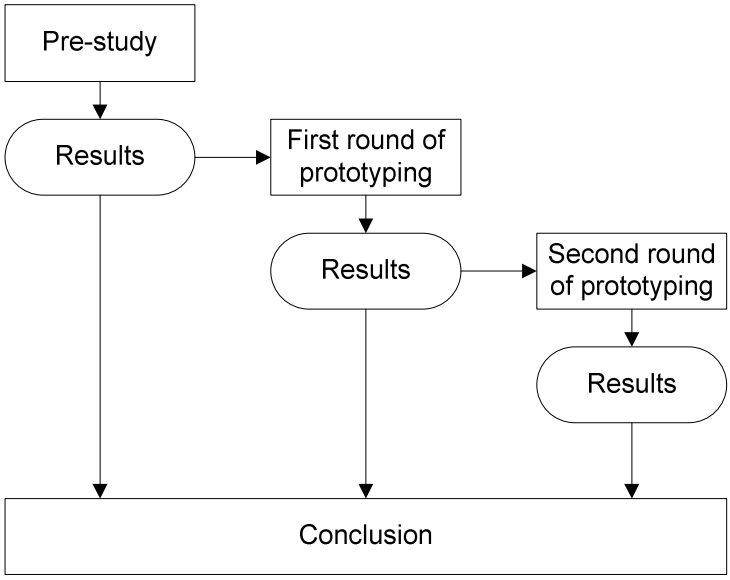


Figure 2. Model of how the outcome of parts of the study has affected the following parts.

Some chapters have a short introduction that describes contents and reasons for its inclusion. The most important results found in chapters are also presented in the introduction to provide the reader the option of understanding the report without taking part of the full content.

Method overview

The empirical data of this study has been collected in three sequential parts, an initial phase and two rounds of prototyping. In all three parts of the study we identify interesting issues and phenomena that could call for further research. These issues, the main result of our study, are collected in the final chapter of the thesis.

To get some input of ideas to build a prototype we started with what we call the Initial phase. It is a first quick exploration of how people react to the phenomenon we're interested in. It's there to give us a starting point, but result-wise it is as important as the other parts. It explored what information people could be interested in if given the chance to use a mobile photo-based search system, and to some extent how they would like their search results presented to them. From the diverse information needs recorded in this part we found that one important issue is that a basic image search does not give a user any means of specifying what kind of information she is looking for using the submitted image.

To be able to study users' reactions and perceptions of using a system that does not yet exist we designed prototypes (Figure 3) of an interface implementing some ideas on how to further specify the information need when doing an image search. A group of respondents helped us evaluate the prototypes and we were able to identify more issues concerning photo-based searches in general and the designs suggested in the prototypes. We refer to this second part of the study as round one of prototyping or the First round of prototyping and it was conducted using unstructured interviews.

Figure 3. The First round of prototyping was conducted using a Nokia mobile phone and the Second round of prototyping was conducted using an iPaq from HP.

The results from the First round of prototyping gave some hints on how to further explore the ideas on how to specify the information need using a second set of prototypes. New, improved prototypes were created (Figure 3) using the feedback from the respondents from the first round, and yet another group of respondents was put together to help evaluating the experience of the second round of prototypes. This time we had a more experimental approach as we let two groups of respondents try different versions of the prototype to compare their experience. Also, in this second round, our interviews were more structured; we did video observations of the test sessions and collected quantitative data using time measurements and a questionnaire.

At first glance, it might look like the method used for this thesis follows a typical and popular system design method - prototyping. We do design prototypes, evaluate them, make changes and then create new prototypes, but it is not made to design a product. The design process is there because no other system available making it impossible to evaluate the experience of using such a system. As a result this thesis contains reporting and discussions about the design of prototypes that in a way are outside of the specified scope, but still necessary to make the study possible.

Apart from the three empirical parts this report also includes a chapter containing background on the subject for those that are not already familiar with theories on Web search, image retrieval and other related areas.

The study, including all data collection and analysis has been made in our native tongue - Swedish, but important parts of the results and analysis have been translated for the appendices.

General method

Our way of conducting research is inspired by grounded theory. Although we did not start without any prior knowledge about the subject at hand, we didn't know much and we tried to keep an open mind and see where the research took us. There is a general criticism against grounded theory concerning this blank start, that no researcher can be completely neutral and no researcher can have no prior knowledge, preconceptions, thoughts or ideas about a subject before starting his or her research (Bryman, 1995). We agree with this but do not see it as a major problem. Of course the researcher has preconceptions, but as long as everyone is aware of that and don't deny it, there shouldn't be any conflict. The main difference between our study and descriptions of grounded theory is that theory and analysis has emerged in our process in parallel with and inspired by the data retrieval and not after the fieldwork is done.

This intervened data retrieval and inspiration from theory throughout the study definitely disqualify it from a phenomenological point of view. We have tended to use terminology and language inspired by readings instead of participants and this is not the way of phenomenology (Bryman, 1995).

To make semi-structured interviews possible, the relation between the researcher and the respondents needs to be pretty close (Bryman, 1995). In this study, the groups of respondents all consisted of people known to the research team before the research started and this has been an advantage in the sense or researcher-respondent relations. It has meant that meaningful conversations could be started rather quickly as both the researcher and the respondent knew quite well the other persons ways of thinking, context and way of speaking.

As the study and its method allow the researchers to follow any lead, an exact replication of the study is very difficult to create. Other researchers starting at the same place would probably end up quite differently. Therefore the appendices where we present research data are extensive. This is to make sure that others have the possibility of analyzing the material on their own to see if their interpretation differs from ours.

General background on related fields

Areas that are presented in this chapter are: How to retrieve information from the Web from a general point of view – this gives an idea of techniques that are used today and how search results can be presented. Image retrieval is presented because it is the technical foundation behind the interaction we're interested in studying. We also look into how a mobile device differs from a desktop computer to support our design discussions. Users' perceptions such as search tasks and usability, and the semantic gap are also covered to provide a better understanding of users' experience of a search process. Presentation of tools to increase usability such as User interface agents, Wizards and Guides are included because the prototypes implement parts of them.

Searching the Web

Searching for information on the Web basically comes down to locating an appropriate Web page and to retrieve relevant information from that site. There are mainly three ways of searching and retrieving information from the Web: By typing URL (Uniform Resource Locator) in the address field of a browser, browsing, and using search engines (Lazonder, Biemans & Wopereis, 2000). These three ways of searching works independently in different situations but a combination of them in different constellations is the common way of searching the Web for information (Glover, Lawrence, Gordon, Birmingham & Giles, 2001). Web browsers and search engines are basically the only tools used to search for information on the Web today. Browsers let the user type URLs directly and browse the Web using hyperlinks. Search engines mostly provide an interface where users can search the Web by typing a text query.

Typing a URL

Searching for information by typing a URL, or as Notess (2005) call it "guessing" a URL, is the straightforward way of retrieving a Web page. The user might try out different Web addresses, to find a desired Web page. This is normally conducted when a user frequently visits a site, or for example when the name of the organization sought is known and there is a probability that the organization's name and the Web address corresponds in some way. However, the top domain address or exact formatting of the address can be hard to guess. If the information the user sought after is located further down in a Web sites hierarchy, successfully guessing the URL is unlikely. One possible strategy is to first enter the domain name, reach the organizations index page and from there browse to the sought information (Notess, 2005).

Browsing and using subject directories

Browsing or "surfing" is the basic way of navigating the Web: the use of hypertext with hyperlinks connecting documents and Web pages to each other. A user can browse from page to page, following hyperlinks and ending up at Web pages that would never have been found using the other two search methods. Further, there is the possibility of browsing a subject directory to find Web pages of interest. Directories select and classify resources into subject categories and sub-categories. The links are presented either as categories or direct links to a Web page (Notess, 2005). Browsing is a well-established retrieval strategy, which users seem to like because of the degree of control it offers them. It is particularly useful when users have ill defined goals or basically want to find out what is available as a goal in itself. It is also useful as a source of inspiration for future query development (McDonald & Tait, 2003). A positive aspect of the browsing strategy is that the user doesn't need to remember or construct queries or URLs. As a consequence, the memory-load is reduced (Notess, 2005). The backside of browsing is that users may not be able to formulate queries and therefore put too much reliance upon browsing. The browsing strategy implies that it is not adequate by itself in occasions where there is a need for supporting it with queries (McDonald & Tait, 2003). Lazonder et al. (2000) found that novice users favored less cognitively demanding browsing strategies, whereas more experienced users preferred an analytical approach. Further, they suggest that there is little need to teach novice users the basics of navigating hypertext – the linked document structure used on the Web. Even though novice users apply less advanced strategies, their browsing performance measures up to the expert users (Lazonder et al., 2000).

Using a search engine

When searching the Web, a user can be overwhelmed by thousands of results retrieved by a search engine, of which few are valuable. The problem for search engines is not only to find topic-relevant results, but results consistent with the user's information need. It is a user's information need that determines which documents are valuable. Typical search engines limit the user to type a text query, even though users consider more than the topic of the query when making relevance judgments of the search results (Mizzaro, 1997). Despite the efforts to develop sophisticated search tools, they still leave a lot of work to the user such as making up good queries (Glover et al., 2001). Search engines are developed for users to optimize their search strategy. The most common way is to let users type a search query that contains at least one keyword.

To exceed just using one keyword, one can refine the search using many keywords combined with Boolean logic. The common Boolean operators used in search engines are: “AND”, “OR” and “NOT”. For example if typing the query: “London AND Paris”, the search engine will present links to Web pages that contain both London and Paris. Using this syntax the user is in a way filtering the results manually. Typing “OR” instead, the search engine will present links to pages that contain either London or Paris, or both. This syntax widens the search result. Consequently, when typing “NOT”, the search engine will present links to Web pages that contain the keyword London, and not Paris. Some pages are excluded by this syntax and the user is again filtering the results manually (Cohen, 2005).

Search engines complement their search with diverse syntaxes. For example a syntax that can be used in the Google search engine is “site:”. This syntax allows the user to limit a search to one domain. As an example: “"Lund university" site:se” results in a list of sites that contain “Lund university” in the “.se” top domain (Delisle, 2004). As “lu.se” is the domain of Lund University, the result of “informatics site:lu.se” would contain hits from the universitys’ Web sites.

How search engines work

The most powerful way of searching for information on the Web is the search engine. Early in the Web’s development, critics pointed out that the Web could never be a well-organized library where one would be sure of finding everything, this because of the lack of a central database and tree structure. They were right, the significant power of the system made vast amounts of information available scattered over many servers around the world. The search engines that seemed quite impractical a decade ago now produce complete indices of a lot of the material on the Web (Berners-Lee, Hendler & Lassila, 2001). The most used search engine Google, executes more than 300 million searches a day (Bosswell, n.d.).

The term “Search engine” has two different meanings (Dictionary.com, n.d.):

- A software program that searches a database and gathers and reports information that contains or is related to specified terms.
- A website whose primary function is providing a search engine for gathering and reporting information available on the Internet or a portion of the Internet.

The two meanings above explain different boundaries, but not how they gather information. There are mainly two distinct categories of how search engines gather information: human-

powered directories and crawler-based search engines. Even if they present the search results similarly they work differently. Human-powered directories are actually not engines as they depend on human editors for its listings. These editors put together information and then decide which sites to list. Therefore, changing data on a Web page has no effect on its listings. Examples of Human-powered directories are Yahoo (http://www.yahoo.com) and AOL (http://www.aol.com) (Web pros group, n.d.). Crawler-based search engines on the other hand are not controlled by humans, instead, a program called a crawler or spider does the search automatically. The crawler automatically downloads Web pages, follows hyperlinks between pages and sites, and collects information that is stored for later use (Sullivan, 2002). The information collected is put into an index, which is a large database of all the sites the crawler visited and read. Here every page is ranked based on the contents, including the title of the page, meta-tags, text and images. In contradiction to directories, changing data on a Web page can automatically change its listings in the crawler-based index; the search engine crawler will eventually find the changes. However, this will only happen after the search engine has seen and indexed the change. Examples of crawler-based search engines are Google and Excite (http://www.excite.com) (Web pros group, n.d.). Further, there exist search engines that searches by using other search engines. In the depth of search engines there are also really specialized once, for example Amazon (http://www.amazon.com) has developed a search engine that allows searching for words in books (Knowthis, 2005).

Internet users spend a great amount of time examining search engine results (Bosswell, n.d.). The user must seek through the search list of Web pages and evaluate each of them for possible relevance to a particular information need. Improving the efficiency of this process directly benefits the end-user. As told earlier under delimitation, there are two ways a search engine can increase user efficiency. Either by returning higher-quality search lists, like better index coverage and ranking algorithms, or providing information element that allows the user to evaluate the results more quickly in a search results page (Woodruff et al., 2001).

There are two key interaction issues to search engines. First, to get a feel of effectiveness of a query and the sorts of information available, users should be able to scan search results quickly. Second, interfaces must also facilitate the flexible and dynamic way users search the Web. Users don't use search engines to retrieve documents until a perfect match is found based on their initial information need. It is not unusual that the user's objective or information need change as the

search proceeds. The information in a result list can change user's perception, change the original goal and trigger off tangential searches (Jones, Buchanan & Thimbleby, 2002).

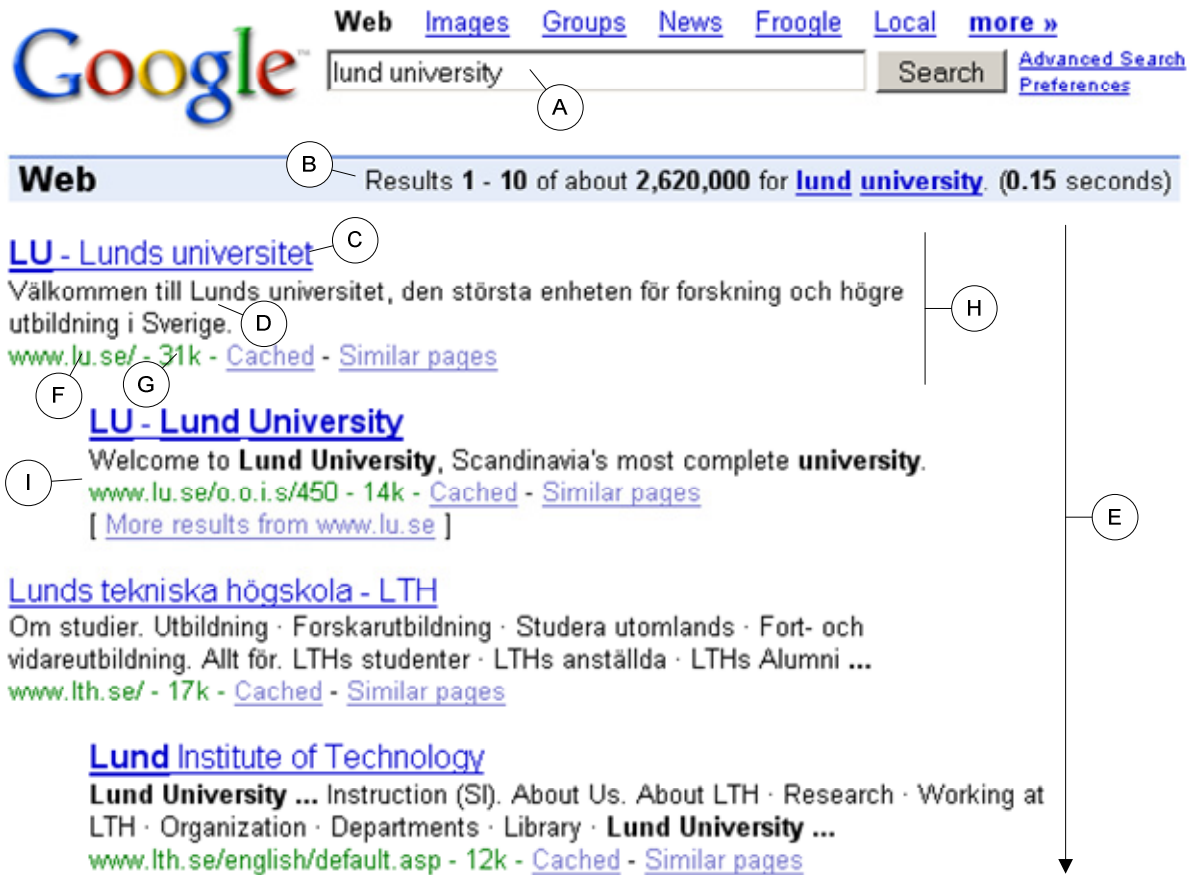
Page rank

Page rank is an important part of the ranking function of some search engines like Google. The page rank of a page grows with the number of and the importance of the pages that link to it. A page rank can be seen as a "vote" by all the other pages on the Web, the more votes, the more important the page is. A link to a page counts as a vote. If there is no link there is no vote, consequently the absence of a link doesn't mean that a Web page is voting against the page (Rogers, n.d.). If a Web page is linked from a high ranked site, this contributes to the Web page's rank much more than a link from a low rated site. Image ranking is based on the same algorithm as page rank; the ranking of an image will be increase with the importance of the pages that contain it (Lempel & Soffer, 2001).

Search result page

The search result page is the page in where a user, after typing a text query in a search field of a search engine is presented a search list of Web pages that matches with the typed query. Here the user can evaluate which Web page to enter by evaluating the information given (Google.com, 2005). Search engines search result pages commonly provide a range of information elements like URL, title and size for each hit, as well as a few sentences that either summarize the document or contain some of the search keywords - the people at Google call this last part a snippet. Because they often consist of text only, they can be quickly downloaded, and often contain a great deal of valuable information about each document. The information elements on the search result page do however not provide much information about the page layout or any images contained in the page (Figure 4).

Many users have a natural tendency to select the first entry in any list. Some search engines, like Google, therefore provide an alternative to the search button, the "I'm Feeling Lucky" button. Instead of presenting the results it directly directs the user to the highest ranked page. Empirical studies show that it is demanding for the user to go through lists of search results, and the average search engine user is unwilling to read through more than a two pages of such listings (Woodruff et al., 2001).



A. *Search field - This is the space where to enter a textual query.*
B. *Statistics bar - This indicates the total number of results, as well as how long the search took to complete.*
C. *Title - This is the title of the Web page found, it is also the hyperlink to the Web page.*
D. *Snippet - This is a text excerpt from the returned result page that has all query terms in bold, the context in which search terms appear on a Web page*
E. *Result list - The list of hits that matched the query.*
F. *URL (Uniform Resource Locator) - This is the Web address of the Web page.*
G. *Size - This is the size of the text portion of the Web page.*
H. *Presentation of a hit – This is all data presented for a hit.*
I. *Indented result - When Google finds multiple results from the same website, the most relevant result is listed first, with other relevant pages from that site indented below it and represented by only one hit.*

Figure 4. Terminology of a typical search result page. (screenshot from http://www.google.com).

Even if Google is the most used search engine, they don't use all possible result information elements one can present, and the elements they use may be excluded in other search engines. For example, Ask Jeeves (http://www.ask.com) exclude elements that Google uses, like size, time it took to conduct the search and indented results. On the other hand, Ask Jeeves sometimes include

a small image of a pair of binoculars. When moving the pointer over it, Ask Jeeves shows a preview image of the Web page (Figure 5).

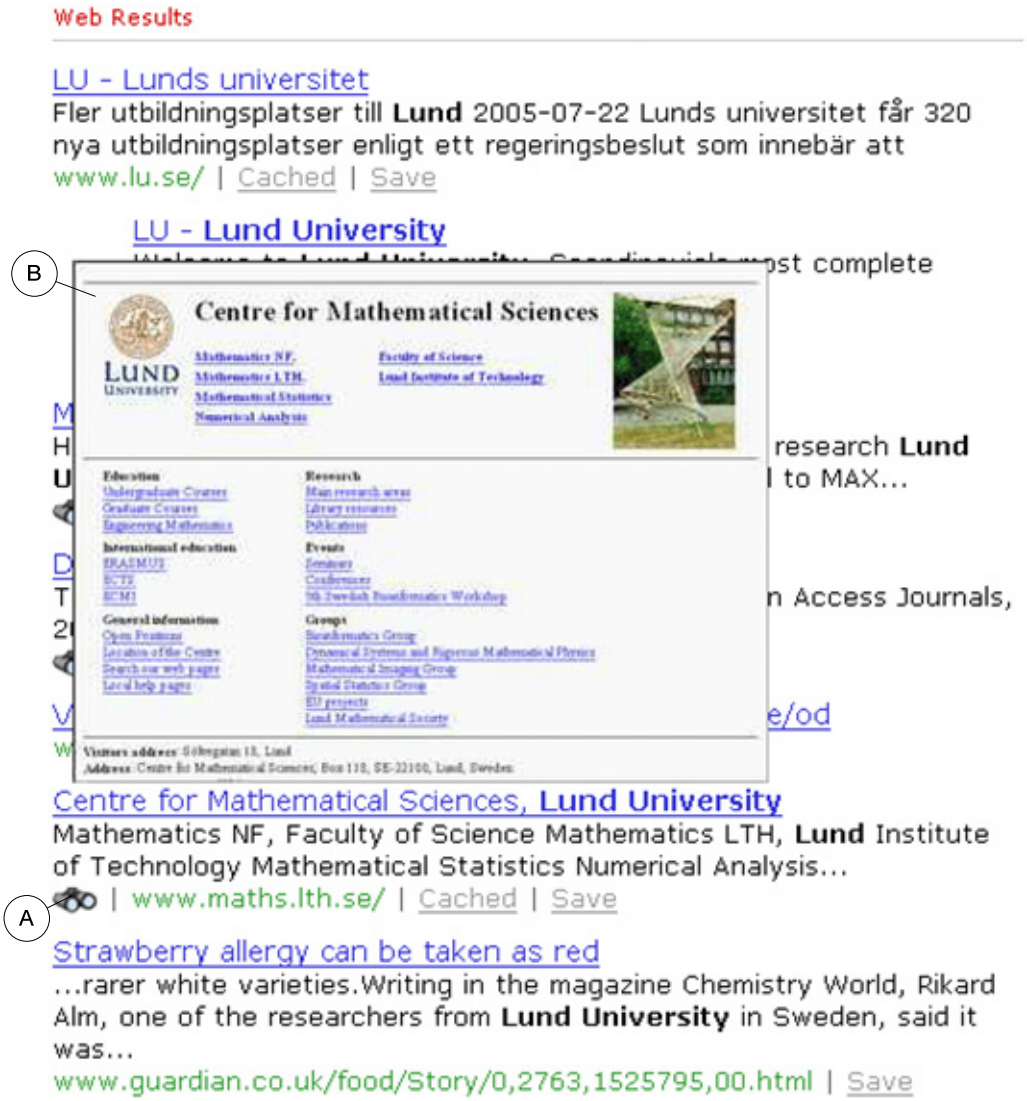


Figure 5. Screenshot from http://www.ask.com where the pointer is moved over the binoculars (A) and a preview image of the website is shown (B) on top of the result page.

In the IDeixis project by Tollmar et al. (2004) one prototype presented thumbnails of the Web pages, a small image of the page, in the result list. Thumbnails of Web pages are typically larger in data size than text summaries and therefore slower to download. Textual content in simple thumbnails is less accessible, as it is difficult to read and is not conveniently summarized. Graphical summaries provide information about the Web pages layout and genre of the page. The visual representation may assist a user that has previously seen the page or a page similar to it by recognizing or classifying it. The human visual system can process images quicker than text and thumbnails can speed up tasks immensely. Results in studies that cover this issue concludes

that users subjectively rank the thumbnail images as the most helpful feature for information retrieval (Woodruff et al., 2001).

A search result page is not always shown as a search list. Kartoo (http://www.kartoo.com) for example presents its results as a map of topics and Web sites. When moving the pointer over a topic, Kartoo highlights Web pages that are related to that topic (Figure 6).

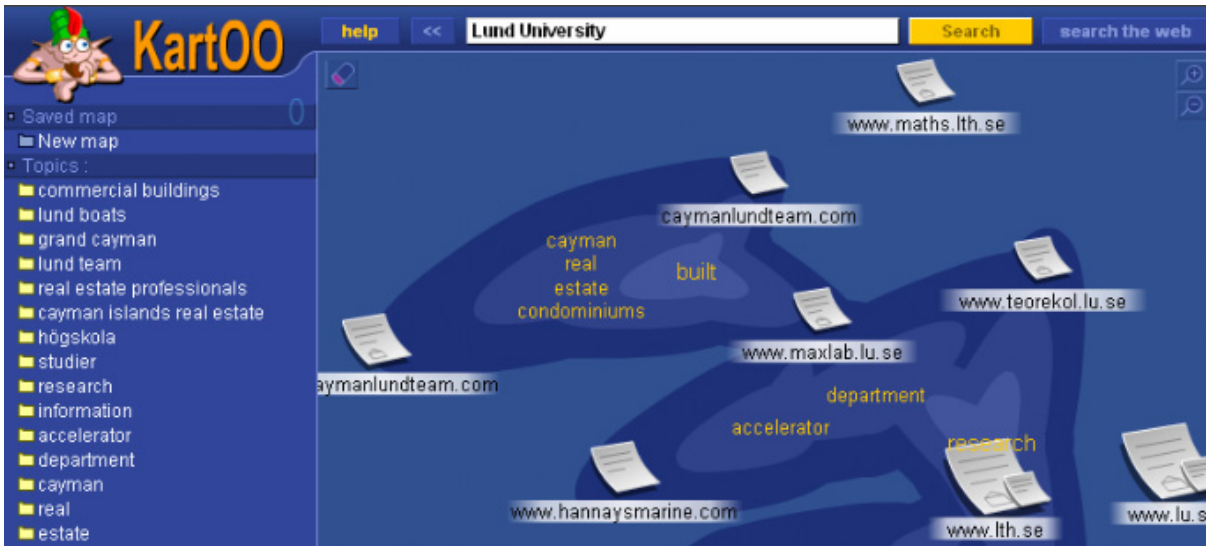


Figure 6. When moving the pointer over a topic, Kartoo highlights related Web pages. (Screenshot from http://www.kartoo.com).

The IDeixis project

One research project that has inspired this study is called IDeixis. It is a study about searching the Web for images using a photo, taken with a mobile phone to retrieve location-based information. Current image Web search engines use keywords as input to find relevant images by analyzing neighboring textual information such as URL and title. In some situations this can be limiting. One example is when a person visits a place where she has never been before the obvious keyword or name for a Web search might be unknown to her and cannot be used as a query. With the current technology however, one can snap a photo using the mobile phone and send it to a friend together with a question to be answered. A negative aspect with this is that the user needs to communicate with another person, and that person may not know what building or place is depicted in the received image (Tollmar et al., 2004).

The basic idea behind IDeixis is that the user sends a photo to a server running an image-matching algorithm that finds matching images in a multipurpose database, such as the Web. Hopefully the matching images depict the same location and the Web pages that contain them

also contain other information that the user could be interested in. In the end this helps the user find location-based textual information of buildings in her surroundings (Tollmar et al. 2004).

IDeixis consists of two major components: a client-side application running on a mobile device which is responsible for acquiring query images and displaying search results, and a server side search engine, equipped with a content-based image retrieval module to match images from the mobile device to pages on the Web. IDeixis does not only focus on pure image search but also includes a prototype of a system that searches the Web using an image as input, extracts common keywords from pages containing similar images and lets the user search the Web using these keywords as a text query. Users that tried this combination of image- and text-based search tend to find answers to simple questions only by looking at the keywords returned, not searching with them (Tollmar et al. 2004).

Mobile devices

On mobile phones users tend to go through the same amount of search result pages as for desktop computers. This page-to-page navigation is expensive in terms of user interactions and time. Often, long lists of search results are divided into separate pages that contain a reduced number of results, due to both transmission requirements and as a means of aiding presentation. The Google search engine only shows 5 hits on each result page on mobile interface and 10 hits for desktop computers. As users normally chooses a hit of a Web page from the first two pages, the user normally makes her choice out of 10 hits for mobile devices and 20 when using a desktop computer (Jones et al., 2002). Studies have shown that if users are required to scroll horizontally in the result pages, users become disorientated (Sweeney & Crestani, 2003).

Dissimilarity between mobile and desktop applications

One of the issues with mobile applications is the absence of a proper keyboard. Typing text-based search queries in a mobile device is more problematical than on a desktop (Kaikkonen & Roto, 2003).

Compared to a desktop computer, mobile devices such as mobile phones and PDAs with ability to connect to the Internet differ quite considerably. First of all the size of the display is smaller, even if a PDA has larger display than a mobile phone it is still far from the size and resolution of an average desktop computer. Second, text input is slower than with a full desktop keyboard. Third, data transfer between the client and the server is slower. Fourth, the amount of storage space is limited, for example the amount of cookie data that can be stored in a mobile device is very

limited. Fifth, the context of use is harder to predict than with a desktop computer, the user could be anywhere and involved in almost any situation (Kaikkonen & Roto, 2003).

Because of the small display size, there is not always space for a navigation bar or other navigational aids on the page. Current Wireless Application Protocol (WAP) applications use a simple tree hierarchy for sites to ensure streamlined navigation (Kaikkonen & Roto, 2003). Less than 1% of the Web's content is marked up specifically for WAP devices using the markup language WML. Standard HTML pages cannot be displayed in WAP-only devices but many mobile devices handle both WAP and HTML (Jones et al., 2002).

Most offices and homes have a fair amount of regularity and predictability. The user can devote relatively consistent attention to performing tasks on the computer. When using mobile applications, there can be a significant number of people, objects, and activities competing for a user's attention aside from the application. Furthermore, since devices are mobile, this outside environment can change swiftly. A mobile device may not be the focal point of the user's current activities, as the user may be trying to balance interaction with a mobile device with other elements in the environment. Users of mobile devices are typically involved in other activities not focused on the device itself (Beale & Lonsdale, 2004). The amount of attention a user can give to a mobile application will vary over time, and a user's priorities can also change dramatically and unpredictably (Campbell & Tarasewich, 2004).

Mobile phones and interaction design

Current mobile devices with limited displays and restricted input methods don't provide optimal user experience for activities such as mobile Web browsing. Information should be presented in a suitable way for the user to categorize and compare data, and to make conclusions and overviews. The user should be provided with both detail and contextual information, which is a very challenging task considering the small size of mobile devices. Further there is a need to manage the presented information in an efficient and easy way. The direct object manipulation user interface and the Window, Icon, Menu and Pointer (WIMP) user interface may be useful with desktops, but the current interaction methods of mobile devices do not support them entirely (Koskela & Vipola, 2004).

As mobile devices tend to have small screens, they are cumbersome to use for retrieving information as traditional interfaces that return lists of matching Web pages in response to keywords. One way to ease the user's burden is the use of a question-answering system. Such a

system shifts a user's work from filtering documents for relevance to describing his or her information need more precisely at the beginning with a question rather than keywords. Question-answering systems on small devices bypass the problem of presenting Web pages and search results on a small screen. This alternative interface for finding information is strong because answers to questions are more likely to fit on a small screen than arbitrary Web pages. A question-answering system demands less of the output-facilities of mobile devices, such as the screen size, and more of their input-facilities as a full sentence is needed (Schofield, 2003).

Image retrieval

Image retrieval is an extension to information retrieval. Approaches to image retrieval derive from conventional information retrieval, and are designed to manage millions of images of almost every imaginable topic. To retrieve these images, researchers have tried to construct effective methods in recent years. There are basically three main approaches for Web image search and retrieval: text-based retrieval, manually annotated image collections, and content-based image retrieval (Lempel & Soffer, 2001).

Text-based image retrieval

This approach annotates images with text derived from the documents that contain them, and then applies text-based retrieval algorithms to the annotated collection of images. The derived text can include the caption of the image, text surrounding the image, the entire text of the containing page, the filename of the containing document and the filename of the image itself (Lempel & Soffer, 2001).

There are two kinds of texts on a Web page adhering to images. The tag text is the pieces of text enclosed in the HTML tag that present the image. It contains information such as the name or the URL of the image, and an alternative text that appears in place of the image on a text-only Web browser. There is also the full text or the body text of the document. Two ways of finding images on the Web using text-based searches are (Aslandogan & Yu, 2000):

- Name or URL match - substrings of the image name or the image URL can be used for matching.
- Shared HTML tags - words that are enclosed in the same HTML tags as an image are likely to be associated with that image and can be used for searching.

Other evidence combination mechanisms include bayesian combination, neural networks and fuzzy sets. These mechanisms and others have been proven to be effective in various contexts including image retrieval. In text retrieval, combining different query expressions and results of different retrieval algorithms have also been shown to be effective (Aslandogan & Yu, 2000).

Google Images (http://images.google.com) is a good example of a search engine where users can search for images on the Web using a text query.

Manually annotated image collections

There are several organizations that specialize in providing visual content to a diverse range of image consumers. These organizations maintain archives where images are indexed and retrieved by keywords, which are manually assigned to each image.

Content-based image retrieval

Content-based image retrieval (CBIR) applies image analysis techniques in order to extract visual features such as color, texture, orientation and shape from the images. The features are extracted in a pre-processing stage, and stored in the retrieval system's database.

One difference between the retrieval approaches is their support of different types of queries. Text-based retrieval systems, as well as the commercial image providers, support natural, topic-descriptive queries. CBIR supports either queries that are formulated in terms of the extracted visual features, or similarity queries, in which a sample image is presented and the system is required to retrieve images with similar visual features (Lempel & Soffer, 2001).

Aslandogan and Yu (2000) made a study of face detection and recognition system on the Web called Diogenes and compared it with three commercial systems: Webseek (http://www.ctr.columbia.edu/webseek), AltaVista (http://www.altavista.com), Lycos (http://www.lycos.com) and Ditto (http://www.ditto.com). They found that the Web offers a rare opportunity for indexing multimedia. Significant amounts of multimedia content such as images are accompanied by textual context. Systems that both searches on textual and visual clues in indexing content should therefore have an advantage. But unfortunately, the textual descriptions are far from structured and the quality of the image may be low. The need for evidence combination mechanisms that can detect if text and images are of high quality is crucial. The study of Diogenes that had implemented face detection as screening and combining text, face detection and face recognition via evidence mechanisms, suggests that it has great potential of finding and retrieve images.

CBIR can be done by making use of the low-level features of perceptual image aspects, such as color, texture, shape, and spatial constraints and indexing these in a database. It is fairly easy to extract these features, apply similarity methods and then presenting these features. High-level concepts are however not extracted directly from visual contents. Instead they represent meanings of objects and scenes in the images that are perceived by human beings. These conceptual aspects are more closely related to users' preferences than the lower-level features. Here, subtle changes in the semantics may lead to conceptual differences (Zhao & Grosky, 2001). The ideal CBIR system from a user perspective would involve what is referred to as semantic retrieval, where the user makes a request like "find pictures of dogs". This type of open-ended task is very difficult for computers to perform – as an example: pictures of different dogs look very different (Sagarmay, 2004). There are two different approaches to CBIR. One that is based on attributes, extracted manually and managed within the framework of conventional database management systems. These attributes are specified using queries which entail a high-level of image abstraction. The other approach depends on an integrated feature-extraction or object-recognition subsystem, this to overcome the limitations of attribute-based retrieval. It automates the feature extraction and object-recognition task that occurs when the image is inserted into the database. There are two types of features, one that is primitive, which is concerned with extracting boundaries from an image, the other type is logical, which defines the image at lower levels of details like texture, color, and shape. Recent CBIR systems combine both approaches. This has raised the efficiency in image representation, data models, query processing algorithms and intelligent query interfaces (Gudivada & Raghavan, 1995). However, it is a very challenging task to extract and manage meaningful semantics and to make use of them to achieve a retrieval that is intelligent and user friendly (Zhao & Grosky, 2001).

CBIR search tools

The low-level visual features for image indexing means that visual features must be used as the basis of image searches. Further the users must use visual queries that represent what the user wants to retrieve. Most search tools within CBIR rely on an integration of a direct visual query tool with thumbnail browsing, most of them focusing on browsing (McDonald et al., 2001). The primary means of submitting direct visual queries in CBIR tools are the use of either "query by sketch" or "query by paint". By submitting a basic sketch or painting of the required image, the system presents a result list that has been ranked in terms of similarity to the image query (McDonald et al., 2001). Another example of using CBIR is the IDeixis project (Tollmar et al., 2004), which uses a photo as the visual query. But IDeixis takes the search one

step further as it is searching for information about the location depicted and not only matching images. McDonald et al. (2001) compared three different visual search tools: a sketch tool, a shape building tool and a browsing tool. They found that while the tools were rated very positively in terms of usability, the respondents believed that successful use of the sketch tool depended upon the users artistic ability. Overall, the respondents preferred a system that didn't depend on any artistic ability, but instead just allowed them to browse the collection of images.

Search tasks and usability

It has been acknowledged that there are two factors that seem to have a particular bearing on the process of image retrieval: the user's task and the type of image they are looking to retrieve. The user's task is of critical importance when determining the usability of a system. Users may have a very specific information need in mind, they may be looking for an image they have seen before, or an image of a particular object or entity (McDonald et al., 2001). Like in the IDeixis project (Tollmar et al., 2004), more information in the context of the photo can be interesting, not the image in itself. That is for instance, taking a photo of a building and wanting to know who built it and not where it is located. However, such information may be located close to the image in the document.

The semantic gap

The semantic gap is the conflicts surrounding when concepts at a high level of abstraction need to be translated into lower, more concrete data. The semantic gap occurs between low-level features and the user's high-level semantic expectations. As an example when searching for an image of a large sea mammal, a system could return a picture of an aircraft as it has a similar shape. The returned image would be relevant in terms of the low-level visual content, but in terms of the semantic closeness of the retrieved image to the user's query there is a significant gap. As a consequence the user has to either browse the search result pages of irrelevant images or engage in numerous query refinements (McDonald et al., 2001).

Filtering

The difference between a search engine and a filter engine is that users can search for documents, using the search engines, which are then processed by the filter engine. The filter engine potentially modifies the documents for example by adding information about link and visit frequencies to hits (Paepcke, Garcia-Molina, Rodriquez & Cho, 2000). There are two main categories of filtering that directly benefit searches: content based value filtering and action based

filtering. The most frequent use is the blockage of irrelevant information, and the ranking of documents to help users process large result sets. Further, value filters can improve browsing activities through guided tours. Using crawlers, Web pages that are of high-value can be explored in more depth than sites that appears to be less important (Paepcke et al., 2000). Figure 7 shows a classification of filters.

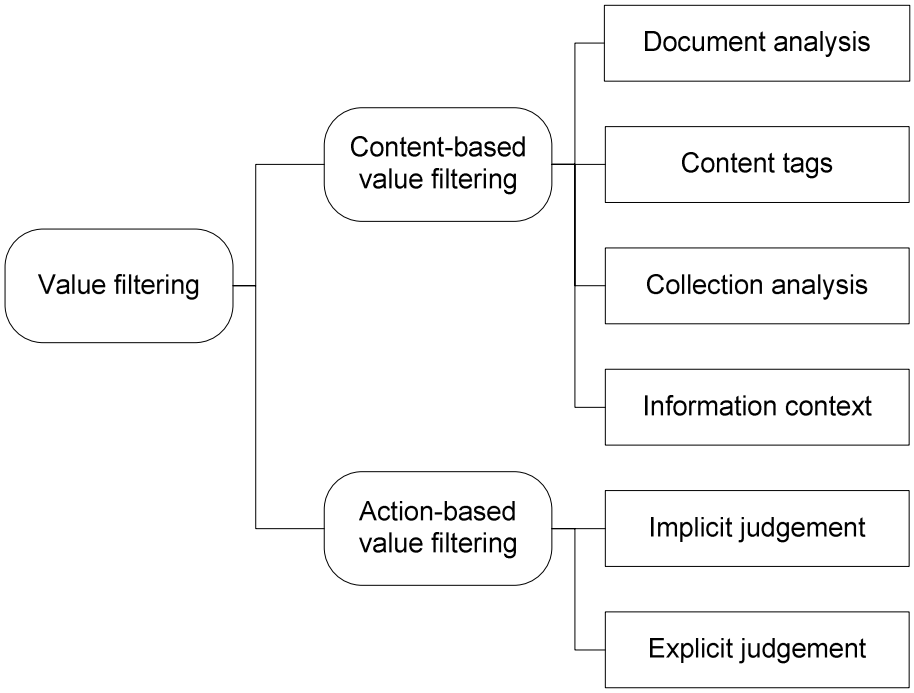


Figure 7. Value filtering and its sub-categories as described by Paepcke et al. (2000).

Content based value filtering

Content-based techniques are when a filtering technique relies on information that is contained entirely within the source collection (Paepcke et al., 2000).

The procedure of content-based value filtering is as follows: a search engine extracts potentially relevant information, and passes it to a filter engine. Then the filter engine removes some documents, ranks the remaining ones, or generates enhanced browser views that are passed back to the user. The filter engine not only removes records from the information stream, they may also annotate, organize or enhance the information that is presented to the user. Content-based techniques fall into four sub-categories (Figure 7) (Paepcke et al., 2000).

Document analysis

Document analysis works on individual documents in a collection. An example is TileBars (Loeb, 1992) which is a technique that relies on document analysis, but is intended to support

users in manually filtering result sets returned in response to queries. TileBar systems maintain a collection of metadata about the location of words within a document. The system's filter engine then uses this metadata to provide visual clues about where in a document, and how often query terms occur. Users viewing these enhanced displays can extract value information from the keyword positions by using knowledge about traditional document structures. For example, keywords occurring in the centre of an article are expected to be part of the document's central theme (Paepcke et al., 2000).

Collection analysis

This technique attempts to analyze the structure of collections instead of the content or the structure of individual documents. As an example, Google crawls the Web and counts the number of links found that point to a given document. With more links pointing to a document the rating of that document increases (Paepcke et al., 2000). This is described earlier in this chapter in the section How search engines work.

Information context

One set of content-based filtering techniques attempts to extract information about the context in which images or documents are situated and uses this information for filtering. Examples of context are the publisher of the document, or the time the document was published. As an example, there have been several wars in Iraq during the last decades and articles written about those wars have different relevance depending on when they were written (Paepcke et al., 2000).

Document-internal content tags

This technique of content-based value filtering involves explicit human initiative and maintenance. It assigns most of the analysis work to the authors or publishers of the information. These approaches rely on human beings physically changing the documents inside the collections by adding mark-up tags that make statements about the document contents. Filtering engines then examine the markings and draws conclusions about the value of documents (Paepcke et al., 2000).

This was in part developed in response to sexually explicit material on the Internet. The idea is to have publishers add tags to documents. The tags would then indicate what kind of contents the documents contained, like nudity and violence. These tags allow filter engines to prevent the browser from viewing this sort of content (Paepcke et al., 2000).

Action-based filtering

Action-based filtering techniques are based on observing human actions associated with a collection of documents. Some techniques observe whether documents in a collection are accessed and manipulated and how. The assumption behind these techniques is that human beings are most reliable to judge information. One can divide the judgment in two sub-categories: explicit and implicit judgment (Figure 7) (Paepcke et al., 2000).

Explicit judgment

This concept was originally used for query refinement, rather than the use of a filter engine module. Feedback takes place when a user retrieves documents that match a query, and then explicitly indicate which documents are relevant. The retrieval system then modifies the original query, and finds more documents that are hopefully equally desirable. Relevance feedback can be obtained from one or more users. The latter is used in an email system called Tapestry where it lets readers comment on a document and others use the collective judgment for filtering purposes (Paepcke et al., 2000). Another example is Pricerunner (http://www.pricerunner.com) where users can filter products like consumer electronics by other users' ratings.

Another filtering technique based on explicit judgment is data-triggered filters that are sometimes just like standing queries. The judgment filter consists of manually added filter expressions. The most known are mail filters which automatically discard messages from particular sources, or with particular contents based on the users judgment (Paepcke et al., 2000).

Synthesized filters, in contrast to data-triggered filters, use an abstraction between the filtering machinery and the user. Instead of specifying precise instructions to the filter, users describe their task or current context. Based on these high level descriptions, the filter engine then chooses among different filter technologies. A commercial system using synthesized filters is LyricTime (Loeb, 1992), which helps the user selecting songs to listen to. The user of LyricTime can chose among different moods to control the filter engine's operation by choosing among "sad", "calm" and "romantic" (Paepcke et al., 2000).

Another value filtering method that relies on explicit user participation for generating the value metadata is asking users to evaluate the information they retrieve. The search system then considers these opinions during future searches (Paepcke, Chang, Garcia-Molina & Winograd, 1998).

Filters built on explicit judgment involve user participation to install triggers, to construct profiles and to declare context. The drawback is that these require time and attention from the user and therefore have a risk of low participation. To overcome this obstacle the filtering can be enhanced so that users don't mind the extra work. Another drawback is that most explicit judgment filters improve slowly over time therefore having a delay in gratification for users. When it comes to collaborative techniques using explicit judgment, users also may fear that they are supporting freeloaders (Paepcke et al., 2000).

Implicit judgment

To overcome the obstacles of explicit judgment, recent systems tries not to involve users focus on an evaluation task but instead letting the system observe user activities and speculate about users opinions. Two classes of approach to collect implicit judgments are identified: conjecture from observing collective user behavior, and conjecture from observing individual user behavior. Within the first class one approach is to analyze access logs and monitor how visitors browse a collection, the more frequent a document is visited the more valuable it is. When recording collective behavior data it is only effective when there is large enough traffic to generate statistical significance. This filter can only work if other users previously have visited the document to be evaluated. When it comes to conjecture from observing individual user behavior, any document can be evaluated, even if the user, or any other user, has previously entered the document (Paepcke et al., 2000).

An obstacle to overcome is the collection of judgment metadata for each individual user. Two primary sources for this metadata are observations of which documents a user accesses, and observations of what the user does with the documents she views. Another way of judgment meta-data involves identifying user behaviors that are effective predictors of user interest. One predictor is time: if a user is studying a document for a long time, the document is in some way interesting to him or her. Time not spent on a document therefore shows disinterest of the document (Paepcke et al., 2000).

Profiles

A method of filtering data is the use of an initial profile that the user fills out with data, later used as metadata for searches. Depending on the goal of the system and the user, this data can contain for example age, gender, interests, and occupation of the user (Paepcke et al., 1998).

Classifying Web sites often requires a great deal of data. Users may be unwilling to manually rate many sites before accurate predictions are made. Pazzani and Billsus (1997) created an agent to learn profiles that allowed the user to provide an initial profile that is revised when the user rates visited sites. This resulted in a more accurate classification and this user profile can be used to determine what other pages might interest the user. Further, the result showed that revising an initial user profile can increase the accuracy of the classifier and the major benefit of the user profile is in selecting features that are relevant to the classification task.

Instead of using an algorithm that checks the amount of links to a Web site, the frequency of access might be assumed more valuable and consequently rated higher. Mixing access frequency and user profiles will get even better accuracy. When deciding between two Web sites, systems that use these techniques will favor the Web site that most closely matches the current user's profile. Combining this mix with counting links would probably give even better accuracy (Paepcke et al., 2000).

Usability tools

User interface agents

A user interface agent (UI agent) guides and helps the user in different ways. Incorporation of UI agents in applications has increased in the forms of search agents, desktop support assistants and collaborative filtering like shopping recommenders. Some agents' purpose is to assist the user by decreasing task complexity. Other purposes are to act as an expert to the user (in the form of expert critiquing, task completion, coordination), or to provide a more natural environment with which to interact (Maybury & Wahlster, 1998). Many UI agents observe the activities of the user and suggest better ways for carrying out tasks. They can also automate a series of tasks based on actions of the user (Raisamo, 1999). UI agents are most useful when assisting a user with a task that is difficult and important. During easy tasks, they provide less help. The more important a task is the more important it is that the user succeeds. Further, the frequency of how often a user will perform a task is of some importance. Agents can help users to learn about tasks in a system that are in frequent use, which ends up in a "no need" of the agent. In this case, the agent acts as tutor. If tasks are infrequent, it is less important that the user gets educated from them. The importance is that users are able to be successful at these tasks. The latter is mostly concerning guides (Dryer, 1997).

> *"Technology design must concern itself with ways to remain connected with the world so that the interactions with the technology take place within the context of on-going interactions between the user and the world."* (Höök, 2000, pp.3)

The metaphors glass boxes and black boxes are two distinct extremes where the first lets the user control and see (transparency) through it. The other is a closed system where the user doesn't know what's going on under the shell of the user interface. For Intelligent user interfaces, the big issue becomes what to place in the glass box and what to place in the black box. Three different levels of transparencies have been identified for the context of tutoring and learning systems, however these principles can be applied to any computer based system (Höök, 2000):

- **Domain Transparency:** Allows the user to see through the system and see the domain behind it.
- **Internal Transparency:** is concerned with the tool itself and how the user can see through the tool's interface into its internal workings.
- **Embedded Transparency:** refers to the whole environment in which the tool is going to be used.

Wizards

The most common UI agents are called wizards. Wizards are often used to automate a task and are frequently used for tasks that are complex or performed infrequently (Burton et al., 1999). The characteristics of wizards are that they assist the user by breaking down a task in a linear series of steps and present them to a user one at a time. Wizards work best with a linear series of steps and they are most successful when the task has algorithmic solutions. Generally they don't use artificial intelligence however one can perceive them as intelligent because they are task specific. A suitable situation for wizards is to replace a multi-purpose task interface with a task specific interface. This is to guide the user through a path to success while the wizard completes autonomously the steps that do not require the user's attention (Dryer, 1997). An example of a wizard is the tool for setting up a new mail account in Microsoft Outlook.

Guides

Unlike wizards, guides are intelligent UI agents. They typically provide the user with task assistance by monitoring the interaction between the user and the information system so that appropriate information is provided. A well-designed guide will direct a person to the next step in

a task. For example a guide could annotate a graphical user interface in a way that communicates how to perform the next step in a task so that a person's attention is drawn to the location of that step. A guide annotates an interface whenever it is most likely to be useful, so it is in a sense intelligent. Guides are more flexible and annotate more complex interfaces than wizards. It is often easier for a user to make mistakes using a guide than a wizard. Instead of changing the interface, guides annotate it. Guides can assist either algorithmic tasks or heuristic tasks, for example tasks for which different procedures will lead to success with some uncertainty, and they therefore adhere to the intelligent user interface family (Dryer, 1997).

Intelligent user interfaces (IUI) can help users tackle information overflow or filtering problems. They can provide help on how to use complex systems or take over tasks from users so they can handle cognitive overload in stressful situations. IUIs are also being proposed as a means to make systems individualized or personalized, thereby increasing the systems flexibility and appeal (Höök, 2000).

A problem is that IUI interfere with usability principles designed for direct-manipulation systems. Those principles include giving the user control over the system, making the system predictable so that it always provides the same response given the same input, and making the system transparent so that the user can understand something of its inner workings (Höök, 2000).

Initial phase

The Initial phase is the first part of three in this study. It's an exploration of what information people could be interested in if given the chance to use a mobile photo-based search system, and to some extent how they would like their search results presented to them. This studys was made because we couldn't be sure of the kind of information people might be looking for if given the chance to use a photo-based search tool. We only had vague ideas and we needed to make sure that we did not start off in the wrong direction. To find the issues and to get some input of ideas to build a prototype according to our purpose, we needed to know where to look for them. One important issue found in this Initial phase is that a basic image search does not give a user any means of specifying what kind of information she is looking for. This led us to the idea of Search modes, a way to further specify what kind of information the user is trying to find related to the submitted image using a wizard.

At the start of this study, we didn't really know what kind of information people in the real world could think of retrieving when introduced to a photo search system. Other studies, like the IDeixis project (Tollmar et al., 2004), start out with a context of use already in mind. We were curious about what sorts of searches people would come up with when they begin understanding the possibilities of a photo-based mobile Web search. Is it basic information about objects or places, or is there more or less advanced information that one wants to retrieve? The Initial phase was done to give us a first glimpse into how mobile photo-based Web searches are perceived by potential users.

Schofield (2003) investigated what questions were asked when searching the Web, and then claim that these questions may be asked on a mobile device as well. We think that a mobile photo search offers another way of searching and therefore the questions asked may not be the same as the questions a user have when accessing the Web using a desktop computer. Not only because of the differences in context, but also because of the differences in search methods. A person may not even imagine the variation of information she may want to retrieve, if not given the tool. We believe you can explore and expand your information needs when using a new search system. A person's information need is not static; it changes over time because of changes in context, but also because of a growing awareness of what information is possible to find (Jones et al., 2002).

The purpose of the Initial phase is to find examples of what kind of information people may look for in mobile situations and what kind of information a person can imagine to find, using a photo matching system on the Web. We were also interested in how and what sort of information elements users would like to have presented to them in a result page from such searches on a

mobile device. We believed that different kinds of search questions would benefit from different ways of presenting search results.

We wanted respondents to ask themselves, what questions they want answers to with a freshly taken photo and give suggestions what kind of information elements they might need in the result list to be able to evaluate links in the result list. This follows the lead of improving the presentation of the search result as introduced by Woodruff et al. (2001) and discussed in the background chapter about search engines.

Method

Respondents

We selected three volunteers to participate in the Initial phase. The prerequisite for participation was that the respondents should all be regular users of camera-equipped mobile phones. Respondents owning and using a camera-equipped mobile phone was of course important when carrying out the task of taking photos but we see other benefits too. We hoped that people owning camera mobiles already had started reflecting over the new technology and its possibilities or at least gotten used to it. Using people who already own camera-equipped mobile phones at least saved us the trouble of introducing the basic concept of using a phone to take photos.

This part of the study was supposed to provide some ideas of what users would like to do with a mobile photo search. We did not strive for generalization so if just one single person would bring up a subject it would still be interesting to us. We could have used just one respondent but realized that a higher number would give us more data to choose from and perhaps the possibility to see some indications of possible patterns.

Tasks

The respondents were informed about the future possibility of searching the Web with photos using a mobile phone. They were then given instructions to use their own mobile phone to take five photos of objects or places that they in some way were interested in finding information about, save them and write down any thoughts of what information they hoped to retrieve using the photo. This should be carried out during a period of 24 hours.

E-mail was used for these instructions to lessen the researchers influence over the respondents' reflections and ideas. If instructions had been given verbally, there would have been a significant risk that a discussion between the researchers and respondents would have started right away.

This could lead the respondents to think in the paths of the researchers, something we wanted to avoid.

Data collection

The day after the respondents carried out their task a meeting with each one of them was set for an interview. During this interview, we talked about each photo and what information the respondent would have liked to retrieve from the Web with that photo, using the simple interview guide found in Appendix 1.

We also presented a list of information elements that can be presented in a result list and asked the respondents to rate them for each photo they had taken and the search question they related to that photo.

After asking these questions a little test was set up to see if the respondents could find the sought information for each photo on the Web using a standard search engine and a desktop computer connected to the Internet. We basically asked them to try to find the information they asked for with each photo using Google, as this is a very commonly used search engine. We also asked respondents which search engines they generally use to make sure they were all familiar with Google. If not we would have had to let them search using another search engine.

Analysis

The 15 photos gave us 15 corresponding search questions that are valuable on their own as they are all true examples of possible searches that users could initiate. We also grouped some of them to find patterns through clustering (Miles & Huberman, 1994). These patterns in search questions cannot be generalized to any larger population but they still give us better insight into what search questions could arise when people get to use a mobile system based on the idea of photo search.

Results

The data collected during the Initial phase can be found in Appendix 2.

Examples of photos respondent 1 took were (Figure 8):

- A bottle of ketchup: The question respondent asked for with this photo was: “What is the price in other nearby stores?”. The information the respondent wanted to know were price, name of the stores and distance to the store.

- Different food products: Here the respondent asked for some recipes that contained the food products in the photo taken.


Figure 8. Reconstructions of images from the Initial phase.

Respondent 1 believed that for all searches the most valuable information elements in the result page would be to categorize the hits, and statistics about the total amount of hits. Other useful information elements were: the URL, image of an image on a Web page, and an image of the webpage. For the bottle of ketchup, the information element of when the web page was last updated is essential because prices of such products may change from day to day.

Examples of photos respondent 2 took were:

- A trashcan containing a bag from a burger restaurant: The respondent asked where the closest burger restaurant was located. The information the respondent wanted were the distance to the restaurant and a location like an address. But most of all, the respondent would like to have a map that showed the distance.

- An Ashtray: Here the respondent asked: Who was the designer? When did the person design it? The respondent wanted a name and a year.

Respondent 2 had a more scattered perception on what was more valuable as information in a result page. URL, the respondent believed was more valuable on three of the five photos taken. The respondent also thought that summary (snippet) and an image of the Web page was valuable.

Examples of photos respondent 3 took were:

- A car: The respondent wanted to know the brand of the car, the name of the color and where to buy such a car. The information the respondent wanted were a name of the car

and the color. Further, the respondent wanted a Web address where one could buy used cars, for example at Blocket (http://www.blocket.se).

- A movie theater: Here the respondent asked what films the theater is showing. The information the respondent wanted was titles, abstracts and show times. The respondent also wanted to be able to book a ticket.

Respondent 3 believes that URL, summary and category are important information to decide which hyperlink to enter. To some extent the respondent believes that information of when the Web page was last updated is important.

In 9 out of 15 cases, the respondents thought that being presented some kind of category for each hyperlink in a search result would help them evaluate if the links could answer their questions. An example would be that shopping-related pages in the search result could be tagged in some way.

By doing a text-based search (using Google) on a desktop computer trying to find the information the respondents wanted to retrieve by taking the photos, 4 out of 10 search questions could not be fully answered for respondents 1 and 2. During the interview with respondent 3, no connection to the Web was available but through discussion we agreed that the sought information about all but one photo probably could be retrieved through a text-based search. In total this would leave 5 out of 15 search questions unanswered if only a Google text-based search was available.

Some of the questions could partly be answered using a Google text-based search. Such as the question related to the ashtray where the respondent wanted to know who designed it and when it was designed. The name of the designer was found but not the date when it was designed.

Analysis and discussion

We found that some search questions were so specific they would probably need a very specialized service or agent to answer the question. E.g. after taking a photo of a bunch of food products, respondent 1 asked for recipes for possible dishes to cook. To retrieve an answer to such a search questions, a system not only need to match images, it must also be able to add up the different food products and use a recipe database, suggesting what ingredients need to be added or what can be done with the products in the picture. This seems like a two-step search method based on image matching, but also, there is a need of an agent that can offer recipes. One can also think of a next step in such a scenario, that is combining this with respondent 1's other search

question of price information (Appendix 2). If a recipe to be completed needed an ingredient that was absent from the picture, the next question would perhaps be where it can be bought at the lowest price nearby. One respondent took a picture of the license plate of a car and wanted to retrieve a phone number to the owner of the car in order to move it. Sequential and very specialized search questions like these are probably not what general photo matching systems will be able to handle initially, but they are still interesting. A system could use techniques derived from semantic Web, Ontologies and Knowledge representation to handle such questions, but they are however not yet available at a larger scale (Berners-Lee et al., 2001).

It seems like information about unmarked objects, that is objects without a text, and that are not known to the respondent are hard to find on the Web using text-based searches. For example the ashtray, the keywords the respondent used were "kulan design". Without the sticker at the bottom of the ashtray that was marked with the text Kulan, it would be much harder to find the needed information. An image search could help out in those situations if similar images exist on the Web. Then there would be no need to examine objects for text or other markings.

Search result information elements that the respondents' thought would be helpful or necessary to evaluate the hits in a result list was not far from what Google currently presents, such as number of total hits of the search, the URL and snippet for each hit. All respondents thought that image of an image on a Web page, and an image of the webpage would be useful; respondent 1 even thought it would be very useful. It might be the same as when searching using text queries. The keywords that a user searches with become bold in the snippet for each hit. The respondents' answers indicate that the respondents think a category indication of hyperlinks and a representation of the matched image would be useful to see if the system matches correctly.

All three respondents gave examples of search questions related to local shopping, like the ketchup bottle, the trashcan containing a bag of Burger King and the cinema where the respondent wanted to buy a ticket. This indicates that a at least a part of users' information need adhere to local shopping.

We initially thought that different search questions and different objects would call for different information elements to be presented in the result list, but we hardly found any support for this hypothesis but for the photo of the ketchup bottle, where the user wanted an up-to-date price information element. We found no pattern in the result of what information elements different search questions would benefit to present in the result page. But of course, our sample of three

respondents is not big enough to rule this out. We suggest this as an area for future studies. The information elements that are presented to the users of small mobile screens might be beneficial to vary. This can easily be implemented when browsing categories, where hits from different categories can be presented using different information elements.

From the widespread types of search questions found in this Initial phase and the fact that the respondents thought that displaying categories in the search result list could help them evaluate hyperlinks better, we were inspired to the idea of Search modes which is presented in the next section.

The idea of Search modes

The results of this Initial phase led us to the problem of different kinds of search questions being asked with the same photo. A general photo-based search system has difficulty knowing what kind of information need the users want to satisfy with each photo. An image of a hat could be connected to so many different questions. A user may desire different information like what kind of hat it is, where she can find one, or maybe what famous person was seen wearing such a hat. The amount of questions that is possible to ask with a single photo is probably so high that browsing a result list to find answers to every one of them is impossible. This issue could be solved by adding a text input field to an image search for the user to further specify the information need, but it would require typing text on the limited keyboard of the mobile device.

The idea of Search modes is based on combining the photo search with some kind of categories to help a user filter a search result. It does not use any text input to save users the hassle of entering text on the limited keyboard of most mobile devices. A schematic model of how Search modes could work can be found in Figure 9.

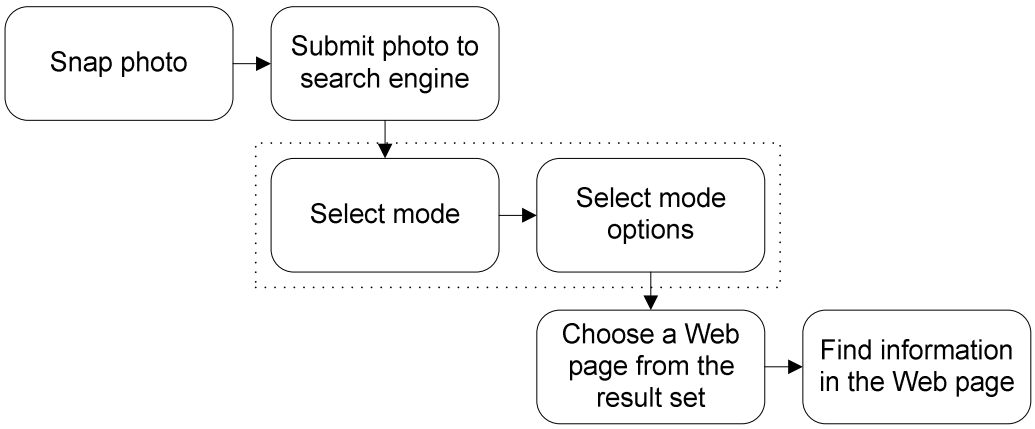


Figure 9. Interaction sequence of a photo-based search with Search modes. The two steps marked in the middle of the sequence are what differ it from a straight photo-based search.

Search modes have been inspired by wizards as they are described in the theory chapter. It is a way for a user to enter parameters for a search in a linear series of steps. Basically a search using Search modes starts with the user taking a photo and submitting it to the search engine. A selection of Search modes is presented, the user selects a mode and then additional options for that mode can be presented. At the completion of the wizard a search result list is shown, but it is not the complete result retrieved from the search photo, but a filtered result depending on the selected mode and option. The result list could also look different and contain different kinds of information elements in different Search modes.

Aslandogan and Yu (2000) propose combining different query expressions and results of different retrieval algorithms for text retrieval. This goes well with the idea of Search modes even though Search modes are using images instead of text.

Special Search modes can be created to support many different types of photo searches. How each mode is technically constructed behind the scenes is not important to us. They can either be based on algorithms filtering on image features, on textual information, on manually created directories, or a combination of these. Any of the filtering techniques mentioned in the theory chapter, or combinations of them, could be used to create Search modes with different purposes. An example would be a Shopping mode that filters on prices and places to shop. This would be appropriate for the ketchup search that one respondent wanted to do. In Figure 10 some examples of possibly interesting Search modes are presented. Some are suggested or inspired by participants in the study and others come from discussions within the research team. One of them, the Location mode, is inspired by the IDeixis project by Tollmar et al. (2004). We do not provide any detail of

what results these Search modes would give or how they would work. We have discussed these issues to some extent during our work, but that is not a part of this study. The important thing is that it is possible to come up with many suggestions for Search modes. We leave defining them in detail to future research.

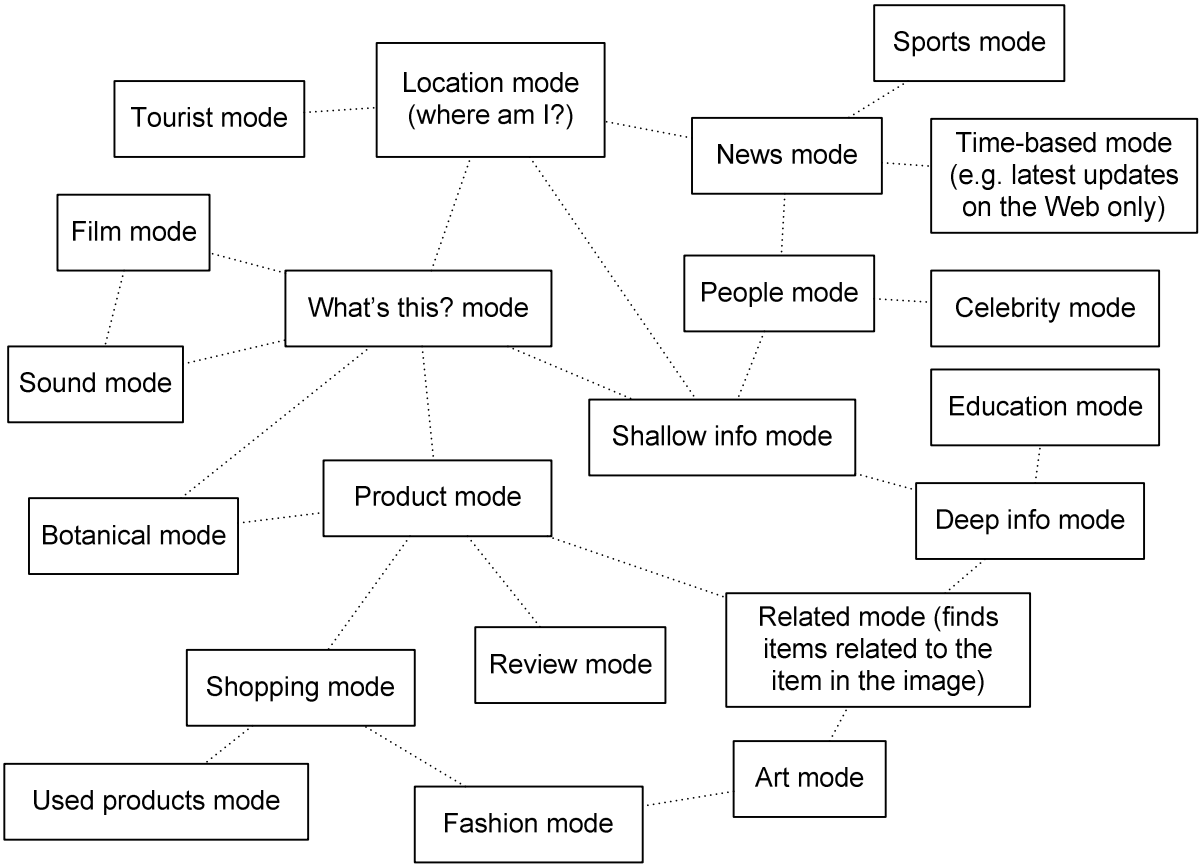


Figure 10. Suggestions for Search modes. The edges in the graph give an idea of which Search modes were inspired by each other.

Conclusion

The most important issue found in this Initial phase is that users of a mobile photo-based search engine will want to conduct very innovative searches but a basic image search does not give a user any means of specifying what kind of information she is looking for. Apart from this the results also indicated that the information elements users wanted presented to them in a search list was not far from what Google currently presents, such as number of total hits of the search, the URL and snippet for each hit. A thumbnail of the matching image or of the complete hit page seems to be valuable as well.

First round of prototyping

This part of the study is the second of the three parts. It explores in general how people react on searching the Web using photos. It also explores users perceptions of a straightforward general photo search and a search system using the idea of Search modes mentioned in the previous chapter. The results indicate that photo-based search is quite easy to understand and to use even though it was novel to all respondents. Some kind of search refinement like modes was considered useful, as the image itself doesn't tell what information is sought. At the end of this part of the study we still felt we could get more and better data if we continued working with respondents and prototypes but with some changes in research methods.

In the Initial phase we found some general categories of search types that inspired us to the idea of Search modes and of course we were interested in users' reactions of using search tools that support Search modes. The main purpose in this part of the study was to get a first glimpse into how users may experience a photo-based search system enhanced with Search modes but we also wanted to further explore the general usability of photo-based search on mobile devices. To do this, prototypes had to be built.

In the Initial phase we can see that situations for photo-based search could be about price for an object. Furthermore location information, such as where to buy an object, was also ranked highly with a special interest in local stores. Based on these observations we decided to study the user experience of a Shopping Search mode, as well as a General Search mode. The General search mode will use the image-matching algorithm without filtering and therefore be a sort of reference. The Shopping mode on the other hand uses filtering techniques to limit the result set to those hits concerning shopping and the options for the Shopping mode handle geographical location.

As in the Initial phase our purpose was not striving for hard scientific evidence, but rather exploring the area, finding indications and inspiration to guide further studies. We also wanted to see if using observations and interviews could give us the data we were looking for. This part of the study quickly evaluates the idea of Search modes and at the same time evaluates methods to study the phenomenon more closely.

Method

For this round of prototyping, user-centered evaluation techniques are used to evaluate the users' perceptions. Researchers have praised these techniques when evaluating systems based on image and information retrieval (McDonald et al., 2001). User-centered design allows comparing system

performance with different users and for a variety of tasks. One main advantage of the user-centered approach is that it allows us to observe how people actually use a system as opposed to just expecting how they would use it.

Respondents

For this round we used four volunteers. These four respondents, one female and three male, were all students studying at the master level at the department of Informatics at Lund University. The effect of gender on this study is probably insignificant compared to the impact of them all being quite used to new ideas in technology and interaction design and also being used to evaluating different kinds of information systems. This was also the premise for selecting them for this part of the study.

The respondents' ages ranged between 24 and 26. They had all owned and used between three and five models of mobile phones since they got their first models four to eight years ago. Three of them had cameras in their latest phone but none of them had used it very much. When using it, it was mostly for fun and as photo reminders to themselves. They tend to think that the image quality is to low and that they rather use a dedicated digital camera for photographic needs. The same three people also had some sort of Internet connection in their mobile phones but none of them used it frequently. When they did, it was to try downloads and some surfing. However, none of them had done a Web search using a phone.

All respondents reported using Google as their primary search tool on the Web. Some also mentioned that apart from Google they sometimes used other search engines like AltaVista and Google Images.

The test environment

Some, like McDonald et al. (2001), argue that a real world system should be evaluated in the real world. Kjeldskov, Skov, Als and Høegh (2004) instead believe that little value is added to a usability study when taking it out into the field. With the subject of mobility and photo search, field studies will probably gain more importance as live image matching systems become available. At this stage in the research, issues can be found in a lab environment. And because the mobile phone is mobile, the real world, as McDonald et al. (2001) mentions, can be wherever one can bring the mobile phone. Because there are many contexts a user can be part of, the context is a critical element that affect the interaction (Maybury, 1999). To get a focus on one possible context and be able to study users equally, a decision was made to carry out the study in a

laboratory environment, more specifically the interaction lab at the Department of Informatics, Lund University. Because a mobile phone is a mobile device some critisism about this choice of context is justified but one single context can't easily be generalized into other contexts and therefore the lab setting would be as satisfying as conducting tests in one live context. We believe that carrying out the study in a lab environment is acceptable as long as it generates new interesting data. If we would reach a point where it didn't, we should take the study to other contexts.

Objects

From the Initial phase we got indications that it is harder to find information of unmarked objects than marked objects by using text-based queries. Out of this we chose two objects for each prototype to search for, the ashtray Kulan designed by Gunnar Larson and manufactured by Gustavsberg, and the chair Myran designed by Arne Jacobsen (Figure 11). One respondent in the Initial phase took a photo of an ashtray and this inspired us because the respondent had some difficulties to find information about it using a text-based query search. The chair was chosen because it has the same attributes of information as the ashtray. That is, both objects have unique forms, and both can be found on the Web. Further, after removing the sticker at the bottom from Kulan both objects were unmarked, as they had no text on them saying what they are or where they come from. This might make them a little harder to find using text-based queries. However, the respondents might already be familiar with them as they are both classic examples of Nordic design. We predicted that it was likely that the respondents would recognize one or both of the objects.

Figure 11. The ashtray Kulan by Gustavsberg and the chair Myran by Arne Jacobsen.

Respondents' tasks and search questions

The respondents were introduced to the idea of searching the Web by using a camera mobile phone. Then, they were given two similar search tasks, one for each object. The information they were trying to find was the name of the object, the name of the designer, where it can be bought,

and a price suggestion. The first search was a straight search showing all hits directly without letting the user select Search modes and options. The second search was a mode-based search with the variety of using General and Shopping mode. The respondents were told that there was a possibility that not all of the information requested could be found.

General mode - Search for information based on visual similarity and without any assistance. The search result is presented as hyperlinks and text summaries of Web pages.

Shopping mode - Search for information based on visual similarity and filters the search result based on price information and location found within each document. In this mode the search is filtered by a couple of question where each step produce a smaller and a more detailed search result.

There is a weakness in this way of exploring a prototype. When the users are given tasks by the researcher they do not have real internal motivation to complete the task and this makes the situation different from a real usage situation. But to conduct tests in real situations with truly motivated respondents we would need at least a functional prototype and this was not available.

Observation

Respondents were asked to constantly talk about what they experienced and were thinking about during the test using the "think aloud protocol" (Preece et al., 2002). This puts some additional strain on respondents as they need to concentrate on the task at hand and at the same time communicate what they do and why. We were interested to see if this way of conducting a test would be feasible for this kind of prototypes and setting. If respondents would seem to have trouble concentrating or the data collected wouldn't lead us anywhere, we would have to try other methods.

During the test, notes of respondent's comments and observations were taken. We expected that just observing the respondents carrying out the tasks would give us a good feeling for what's important and interesting. Very little of these influences can actually be written down when taking notes, as the researcher may not be aware of them at that point, but we need to recognize that they exist anyway. Another issue with observations is of course that respondents are not unaffected by the researchers presence and will adjust their behavior to the situation.

The respondents had no time limit when carrying out the test. Instead we chose to terminate the tests when our observations no longer added any more valuable information, or if the test person lost interest in the task.

Interview

After the respondents were done doing the first search task, we asked the first question from the interview guide: "How does searching by taking a photo and using it as input to the search feel?". This was to inspire the respondents to get into a reflective state and get the conversation going. After the second search task we conducted a semi-structured interview based on the interview guide in Appendix 3 to be able to evaluate the respondents experience and collect their views on the subject. The questions were designed to capture respondents' opinions on mobile photo-based searches in genereal, Search modes, and information elements presented in result lists. But as we used semi-structured interviews the questions should be seen more like a way of keeping the discussion going and touching the topics we're interested in. We continued taking notes during the interviews as we had done during the actual tests.

All notes from the observations and interviews were collected in a matrix, sorted by question. This data can be found in Appendix 4.

One of the respondents seemed particularly interested in discussing the subject and the conversation went on beyond the questions in the interview guide. Notes of this discussion are also collected in Appendix 4.

Analysis

To analyze the results we sorted our notes of answers to the questions into a matrix having respondent on one axis and question on the other. This way we could compare respondents' answers to each question and identify patterns and differences. We also identified some themes among the notes from our observations (Figure 12).

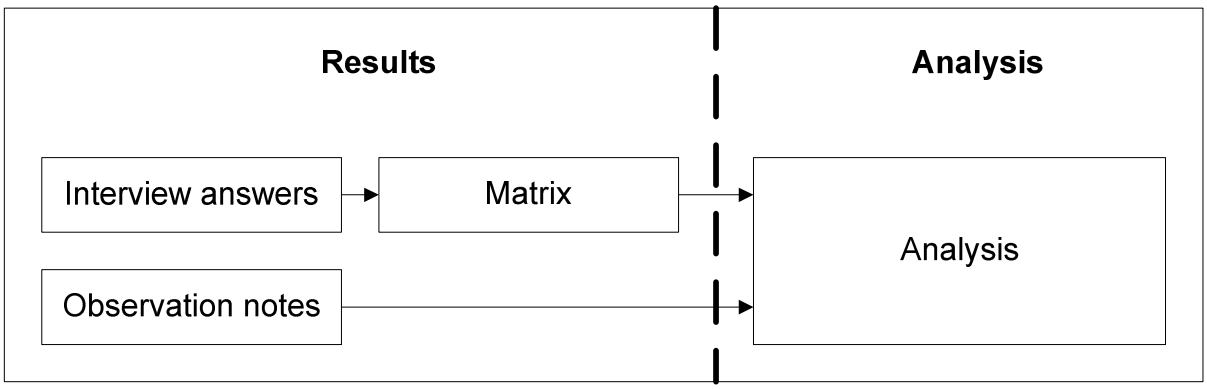


Figure 12. Model of the analysis of data in the first round of prototypes.

The prototypes

To conduct the test we used a Nokia 3660. This is a quite old mobile phone for a study like this, but it had all the embedded technology needed to do and test the prototypes, a camera, a Web browser and a memory card to store content to browse.

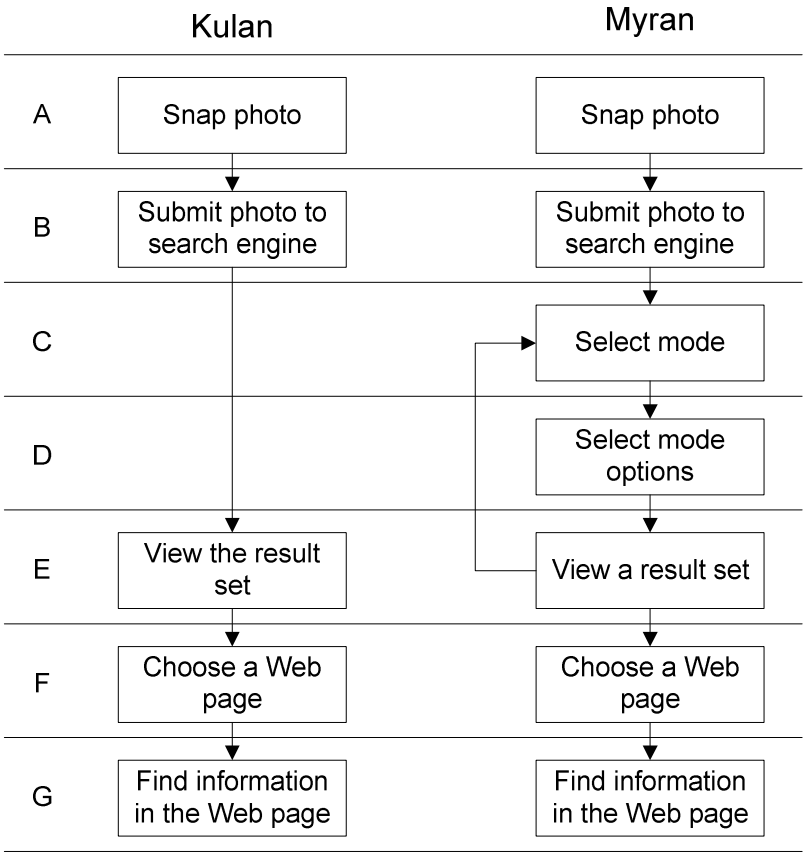


Figure 13. How to navigate the two prototypes, Kulan and Myran.

The navigation of the two different versions of the first prototypes can be seen in Figure 13. The Myran prototype is designed to implement modes and the Kulan prototype implements a straight search. For the prototype Kulan, a straight navigation would be: The user first snaps a photo of an object (step A). Then the user submits the photo to a search engine (step B). When the result is presented on the display, the user can make decisions by browsing the result pages (step E). After finding a potentially interesting Web page (step F) the user may find the sought information in the chosen Web page (step G). It might also be possible to find the sought information for example in the snippet. If this is the case, the user skips step F and G. For the prototype Myran, the navigation differs some from Kulan. Step A and B is the same, but instead of directly browsing the result pages, the user choose between General and Shopping mode (step C). Then, the user chooses a location in the mode settings (this only applies to Shopping mode) (step D). The user can now browse a filtered result set (step E). Step F and G is the same as for Kulan. If the user is not satisfied with the result set, the user can go back to select another mode (step C).

Simulating the search and filtering

As we didn't have access to any real image retrieval database with indexed Web pages, the basic idea has been to simulate the content-based image retrieval by using text-based retrieval on the Web and then selecting a matching image from each search result manually.

The prototype itself consisted of static html files that are stored on a memory card in the phone. The phone has no active Internet connection, so all Web pages had to be downloaded and stored during creation of the prototype. Pages that present the search result list as well as all external pages that are linked from there are stored on the memory card. Also all images, stylesheets, etc referred to by all downloaded Web pages, needed to be stored locally. A negative aspect is that we neither encounter the time it takes for the algorithm to find images, or the download speed.

Obtaining search results using the Google API

This section describes the technical aspects of how the result sets and the prototypes were generated. It's presented here for those especially interested and the general reader can skip or skim through this part.

We created a prototype generator to lessen the manual work of collecting hits and presenting them. It was written as PHP scripts, run on a standard Apache Web server and has a simple HTML form-based interface for input. The generator connects to Google via their API, (Application Programming Interface) using the SOAP protocol. The API can be found and downloaded for free at the Google website (http://www.google.com). This API returns search

results as data objects instead of HTML files as ordinary searches on Google do. Each link returned contains the data items Google usually displays in a search result list, like a title, a URL, a size, and a snippet. The snippet is generated out of the text on the originating Web page, with items from the search string in bold. Using this information, the prototype generator can build HTML files presenting the results in a design we choose.

To initiate the prototype generator a query is needed. This query corresponds to the General-mode. If other Search modes are included, extra search words are appended automatically to the original query for those Search modes, and each query defines its own search result list. This means that the search results are not subsets of each other. The pages in Shopping Lund don't necessarily appear in Shopping Sweden or Shopping Europe.

The design of the query strings is based on trial and error. We have tried different strings and modified them until we found search results that were acceptable and contained as many pages with matching images as possible.

For the first object, "Kulan", we created a straight search prototype with no options or choices to make after the photo is taken. The Kulan prototype takes the user to the search result list directly after taking the photo. The Myran prototype on the other hand uses the General and Shopping modes, where Shopping mode in turn have three options of geographic location. The Search modes and options and their corresponding queries can be found in Figure 14.

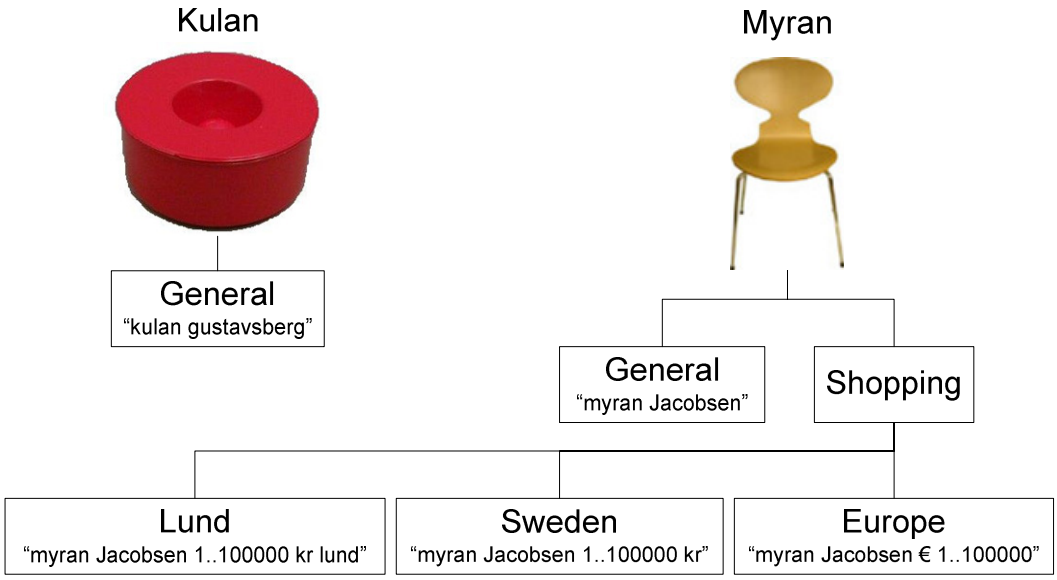


Figure 14. Queries used for the Search modes and mode options in the two prototypes.

Lund is the city in Sweden in which all respondents were resident and also where the tests were carried out. The choices of Lund, Sweden and Europe together gave the user a possibility to specify how geographically narrow the shopping search should be. The idea was that this would create a feeling that the system is aware of the user's current location.

In Google-queries the double dots specify a range search. "1..100000 kr" will find pages containing a number between 1 and 100000 close to the string "kr", which is the abbreviation for the Swedish currency "krona". Our idea was that this would find pages containing price information, which is essential for shopping related pages. We decided not to search for objects worth more than € 10000 to make a search for numbers possible as a range search by definition needs a floor and a ceiling. 100000 kr roughly corresponds to € 10000, according to exchange rates for 2004, and this accounts for the difference in numbers between the search options.

The abbreviation for the Swedish currency, kr, is also used for the Danish and Norwegian currencies. This accounts for some erroneous Danish websites in the Shopping Sweden search result.

Designing Search modes using different queries like this would fall under the filtering sub-category explicit judgment as it basically is about creating a data-triggered filter. The triggers in this case are the currency and location information. The act of using this mode however would fall under the category called document analysis as the filtering is done automatically and its results are dependent of the contents of the document itself. An ideal Shopping mode should perhaps be filtered on a manually created directory of shopping sites and local stores and in that case it would fall under the category called implicit judgment.

Using the queries, the generator downloaded search results from Google and created static html files showing the search results. Also all linked files in those results were downloaded to a separate folder as well as all images, stylesheets, and java scripts connected to those pages. Links to PDF files and HTML files containing frames were skipped as the browser in the mobile phone we used couldn't present such pages properly anyway. Because we didn't have the mobile phone connected to the Internet during the test, the prototypes contained some broken links, missing style sheets and misinterpreted scripts. The results list was limited to the first 10 hits for every mode, and all hits were shown on the same page.

When the result pages had been created, the matching images had to be manually added to each item in the search result lists. All images from each hit had already been downloaded to separate

folders, one for each hit. We manually selected the best matching image from each folder, copied and renamed it to a standard name. This made the selected image appear in the search result list. The image is resized to 50 x 50 pixels by using the size property of the html image-tag. This makes some images appear stretched.

The prototype does not support using the built in camera of the device. Instead we manually selected a photo of the object and put it in the snap-page of the prototype. This was the photo the respondents used to search with.

Design issues

Below follows a discussion on design decisions made when creating the interface for the first round of prototypes. The decisions with motivations can also be found collected in Table 1.

For the prototypes to be made, some design elements from the Initial phase were considered. For example, respondents thought some information elements in the search list were more critical than others. They were URL, snippet, amount of hits and category. Respondents also thought that it would be useful to have the matched image in the search list so that they could see if the system matched images correctly. This was implemented to see if it actually could help users to find the sought information.

A problem is that the mobile phone display is small and so the information elements in the search result list shouldn't take up too much space. Also, the user may be on the move and so the search result information must be accurate for the purpose of the user because attention and time has to be shared between the search task and other activities.

The first page of the prototype is similar to the one used in the IDeixis study (Tollmar et al., 2004). It contains a static image of the object and at the bottom a hyperlink saying Snap photo (Figure 15). In the "Myran" prototype, this hyperlink takes the user to the mode selection page and in the "Kulan" prototype, directly to the result list (Figure 15).

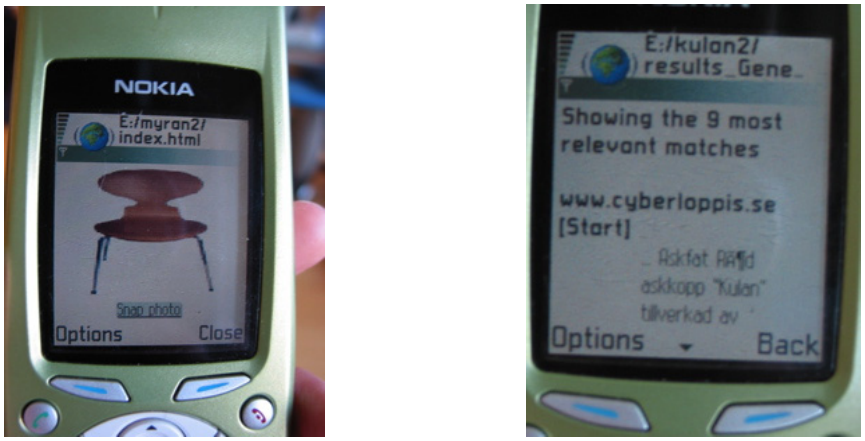


Figure 15. To the left a static image and a hyperlink saying "Snap photo" which in the "Kulan" prototype takes the user directly to the result page shown to the right.

The mode selection page just gives the user a choice between the Search modes General and Shopping using hyperlinks. The General hyperlink takes the user directly to the General search result list and the Shopping hyperlink directs the user to the options page for Shopping mode (Figure 16).

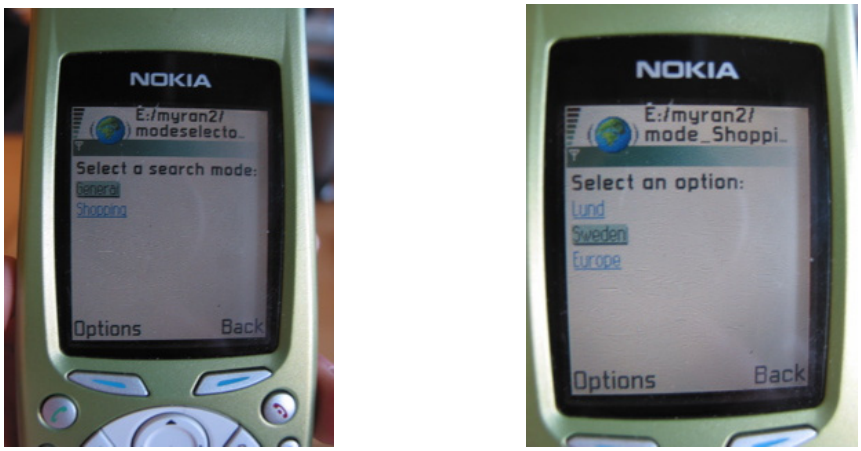


Figure 16. The mode selection page and the mode options of the Shopping mode.

The options page for Shopping mode simply contains hyperlinks to the three choices of geographic areas Lund, Sweden and Europe (Figure 16).

The results page presents the text "Showing the n most relevant matches for mode option" if using Search modes and otherwise "Showing the n most relevant matches" in bold. This is to inform the users how many hits are available and also to remind them what mode and options they have selected (Figure 17). This all falls under the design principle of feedback as discussed by Preece et al. (2002).

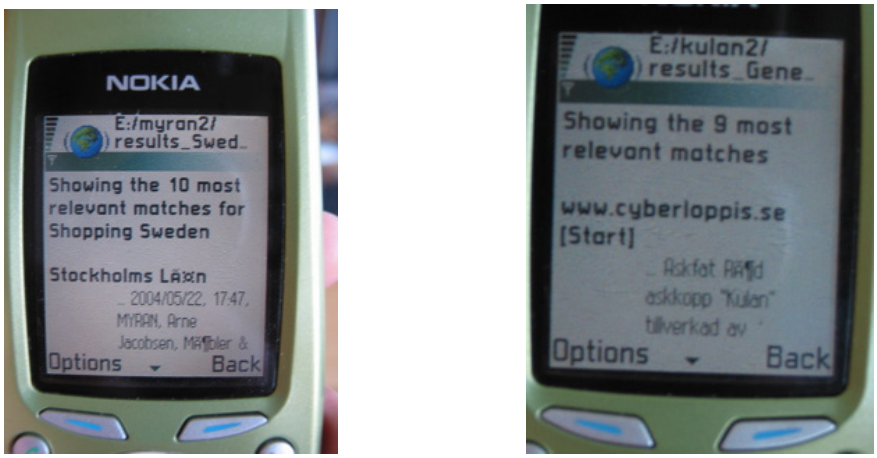


Figure 17. The result pages of Shopping mode to the left and General to the right.

At the end of the search result page is a hyperlink that says "Change search mode". This is to remind the users that they can go back and choose another mode if they are not satisfied with the hits in the current mode (Figure 18).

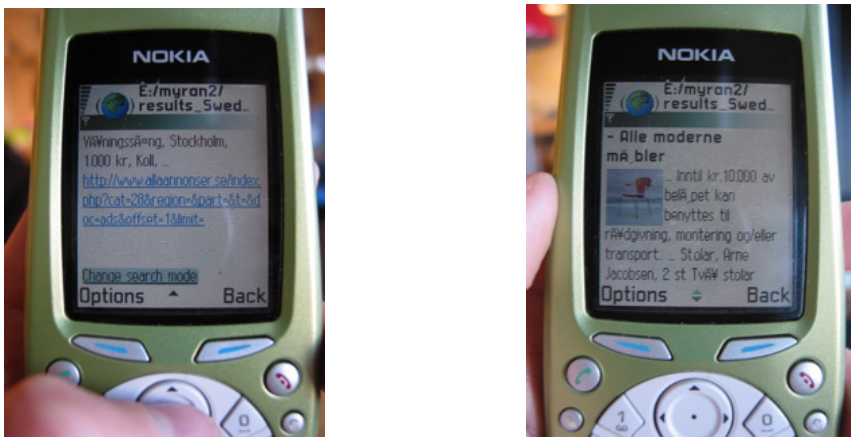


Figure 18. The option "Change search mode" and a hit with a thumbnail of a matched image.

In the Initial phase the respondents suggested that the matching image, snippet and URL are the most useful information element in the result list. We did not ask respondents in the Initial phase about the title of a hit but it is such a common element in most search engines that we decided to use it anyway. Presenting the matching image for each result also has support from Woodruff et al. (2001) who claims that searching for an image of a particular object among other images is faster than searching for the name of that object among other words.

We used no special color-coding on the search result pages. The title and snippet are both in black and the hyperlinks appear underlined in blue. The currently selected hyperlink is

highlighted in grayish blue. This is all the standard interpretation of the devices Web browser. The only formatting applied is to the titles, which are placed inside bold-tags.

The snippet Google returns for every result in the result set contains the HTML bold-tag (<b>) around all words from the search string. We stripped all HTML tags out of the snippet to avoid highlighting words from the search string and reveal the underlying search criteria to the user.

In the mode prototype there are four steps before the result page is presented. The usual navigation aids showing which step one is currently at and the amount of steps there are in total are excluded. This, because of the small screen size and that the purpose of the system is that the user will use it frequently, and as a result this information is of less importance when the user has learned the steps. This is against what Kaikonnen & Roto (2003) are proposing as they find it important to give the user a navigation control and the possibility to go back. For the latter we have implemented a hyperlink to be able to switch to another mode (Figure 18). The back button of the browser can still be used if the user wants to go back one or several steps.

Table 1. Design decisions and their motivations for the first round of prototyping

Mode and mode options selector	Results from Initial phase indicate that users may want some kind of categories to specify their information need
Not presenting what total number of steps and current step when going through wizard	Users will use the system frequently and once they learn the flow of interaction this will not be important. The number of steps may also differ between different Search modes as they have different number of levels of options
Total amount of hits presented in the search result page	Results from Initial phase and comparison with other Web search engines suggest this
Selected mode presented in the search result page	Design principle of feedback
No special color-coding on search result pages, just using the standard interpretation of the device's Web browser	We decided not to make a decision to avoid drawing attention to certain elements as this could affect users reactions
URL and snippet presented for each hit	Results from Initial phase and comparison with other Web search engines suggest this
Matching image presented for each hit	Results from Initial phase indicate this and Woodruff et al. (2001) claim that searching for an image of a particular object among other images is faster than searching for the name of that object among other words
Title presented for each hit	Other Web search engines have this and we keep it to avoid confusing users
Not presenting any other information elements than the ones in other decisions for each hit	The screen of the device is small and space is valuable to avoid scrolling
Removing the bold formatting of keywords used in the search from the snippet	To avoid presenting the underlying search criteria to the user and revealing the workings of the non-functional prototype
No back functionality	This is already implemented and available in the browser functionality
Change search mode link at bottom of search result page	Reminder to users that they can modify their search criteria

Results and analysis

Interviews

The data collected during interviews was put into a matrix that can be found in Appendix 4 and this is the analysis of that matrix. All respondents' answers to each question have been analyzed to find trends and issues that we follow in our exploration.

How does searching by taking a photo and using it as input to the search feel, compaed to text-based search?

Result: Three out of four were positive or very positive to photo searches. They thought it was good way of searching if one doesn't know the name of the object or can't come up with key words for a text-based search. Two of them thought that they rather would use text-based search if it is faster. One respondent reported preferring the single click one do when conducting a photo search over typing a textual query using many buttons. Another respondent used the same search strategy as when using a text-based search that is looking in the snippet for information and confirmation. The same respondent was worried about finding copies of the original object instead of the very object in the photo.

Analysis: Some of the respondents recognized the pros of a photo matching system brought up in earlier chapter, especially the issue of eliminate the work load of coming up with relevant key words as one do for text-based searches. The issue about copies of object is interesting. As for now it is up to the user to determine if the matching image is correct and if it is a copy or not. It can be the opposite as well, that one is taking a photo of a copy and receives hits of originals. The problem would arise when there are many copies presented on the Web, and the Web pages containing the sought information exist in the original only. The few hits to Web pages containing the sought information could drown in a search result of thousands of hits.

What was your thought on being presented a choice of Search modes?

Result: One respondent thought it was hard to choose between Shopping and General mode, but started with Shopping mode, and if that wouldn't work the respondent planned to go to General mode instead. One respondent believed that if you just want general information you enter General, and if you want price information you enter the Shopping mode. One respondent hoped to find prices under "Shopping".

Analysis: It seems that the respondents doesn't identify Shopping as a subset of General but as two different result sets. Other than that they seemed to predict that Shopping was related to pricing information.

What did you think of having to go through several steps to reach a search result list?

Result: Generally the users were positive to go through several steps to reach the result list if it resulted in fewer but more useful hits. One respondent thought however that there should be no more steps than the prototypes currently have. One respondent thought that four to five steps would be a maximum amount of steps, otherwise one loose interest. One respondent thought that automatic sensing of location would be better than the ability of letting the user choose. One respondent was disappointed that the location didn't work as the respondent wanted a Swedish site and ended up at a Norwegian.

Analysis: Too many steps using Search modes for filtering might cause frustration. A location sensor could be appropriate, like using GPS. However, it is not always the case that one wants to buy the object where one physically is located when conducting the search. In the case with the respondent who got hits from Norway when looking for hits from Sweden is a fault in the prototypes, the search was made on the currency of the Swedish "krona", the same name as the Norwegian currency.

Did you feel that the result became more precise when you used the Search modes?

Result: One respondent thought that the result was not as specific as the respondent wanted. Another respondent thought that it was hard to tell if the Web page found actually was located in Lund when entering Lund as the location.

Analysis: It seems like the respondents wants specific hits but not too many steps to go through. An ability that lets the user be able to determine if the site is in the location chosen is of great importance. As it looks today it is up to the Web sites to present such information.

Can you think of other types of Search modes that could be handy with mobile searches?

Result: One respondent suggested tourist info as a mode and the same respondent thought that Shopping mode might not fit all products, because not all products come with the same shopping behavior. Another respondent suggested Education mode and Used products mode. One respondent suggested a Geographic mode with a filter that would find the Hilton hotel in Paris and not the celebrity Paris Hilton. The same respondent thought that there should be an auto sense of geographic location and an option to type in keywords. A respondent suggested a review mode, a mode where one can read others opinions of different things like ratings of restaurants.

This respondent suggested that Search modes should not be too specific as the respondent wanted to see the results close to the filtered area.

Analysis: Not all products fit into the use of Shopping mode, such as things that are not present at the time of searching and therefore not able to take a picture of. When searching for products using the Web, one is using a certain shopping behavior that involves the Web. One idea is that after taking a photo, the system recognizes the object and puts the user in a special designed mode for that product. But if this is necessary, we don't know. The principal of narrowing the result might have a downside if the filtering is to fine. A problem is that users may want to explore more than what the chosen mode presents. When it comes to the Hilton issue, the use of a photo search probably already solves it, as an image of a building is quite easy to distinguish from an image of a person.

How did you make the decision of which hyperlink to follow in the result list?

Result: All four respondents thought that the snippet is the most valuable information element in the result list. All four also thought that the thumbnail images are important in the way that they confirm or disconfirm a hit as a correct hit. One opinion a respondent had was, if the system works very well there is no need for images. The respondents preferred domain name before URL. Two of the respondents thought that the title was unnecessary and that it is untrustworthy. One respondent didn't like to scroll whilst inside a Web page. Instead the respondent just wants to enter a page, and go back and follow another hyperlink if the sought information didn't turn up at first sight.

Analysis: Snippets as other search engines use them seems to be one of the most important information elements in a result list. The respondents are however not satisfied with the URL, instead they believe the domain name is sufficient. The reason might be that URLs may take a great amount of the small screens valuable space. The title seems not to be the prime information element that the respondents are using when scanning the result set. It's also set by the author of the document and therefore can't be trusted.

If a search engine would identify perfect hits, it wouldn't need to show a result page, and instead redirect the user directly to the Web page, like Google's "I'm Feeling Lucky" button. There is however no guarantee that a future system always returns 100% correct hits and the results indicate that images seems to be a good way of confirming the hits.

It seems difficult for the users to scroll inside a Web page. A suggestion is to let users arrive where the image is located on the Web page instead of at the top left corner. The sought information should hopefully be close to the matching image on the page.

Are there any information elements in the result list that you find unnecessary?

Result: One of the respondent thought that title gave a hint of what kind of information the Web page contains but said that it can't be trusted. Another respondent reported not using the titles to evaluate hits.

Analysis: The author of the Web page sets the title and that doesn't necessarily conform to what the user wants. As a response to question 6 above, two respondents said they didn't trust the title but it still seemes to be used to some extent. So the title can be of importance when one wants a hint of what kind of Web page is behind the hyperlink but to save space on the small screen of the mobile device the title could perhaps be left out. It would be interesting to see how users respond to a result list without titles for the hits.

Is there any information element that you think is missing in the result list that would allow you to evaluate the links better?

Result: One respondent suggested that the date of last update is a good element to have in the result list. Another respondent talked about the importance of the size of Web pages in bytes, this because if the connection is slow, one wants to know how large a Web page is before downloading it.

Analysis: Date of last update is important in some types of searches. One thought is to provide a default setting of certain information elements for every mode. If the user is not satisfied with the information elements, she should have the option to include or exclude information elements in the result list.

Observations

Searching the Web using images was found to be an unfamiliar concept but easy to understand and use. It was notable as the respondents got used to the idea, that photo-based search was evaluated to be most useful in mobile situations. A few respondents also commented more directly that typing on a mobile device is inflexible, and searching using an image by the click of a button is a clear favorite.

Search modes were considered to be very useful, as one respondent expressed it: "I really like the possibility of filtering the search". However, it was also suggested that it should be possible to

filter the search by both pre-defined questions as well as adding keywords. Another subject commented that his ordinary search strategy, that is to open many hyperlinks at the same time, failed in the mobile situation.

A couple of respondents expressed a concern that some Search modes could lead to a lot of advertisements and would like an additional filter to avoid this kind of information.

The respondents spent a lot of time navigating inside the Web pages. It seemed hard for them to find the sought information within the pages, and it seemed hard for them to rule the hit out if the page didn't contain the information.

One of the respondents was unsure about how General and Shopping mode relates to each other, and what the term General stands for.

No respondents used the hyperlink saying "Change search mode" at the bottom of the result pages.

Evaluation of prototype

Some respondents seemed to be unsure of how General mode and Shopping mode related to each other. This might be because the hits in General and Shopping mode were created as separate sets. One way to get around the confusion can be to make Search modes and their options as subsets of each other and General mode and let the interface communicate this in some way.

The prototype didn't reflect the mental picture of how users perceived the different relations between the different options General and Shopping mode.

The reason why respondents didn't use the hyperlink "Change search mode" at the bottom of the result pages might be due to that the respondents found the way of searching as new, and that they thought that the algorithm should present correct pages. Another reason might be that they were not comfortable with what "Change search mode" meant or that they just didn't notice it at all.

The prototypes were not as good as we would hope for. The screen size and resolution of the mobile phone was not adequate when navigating inside and between Web pages. This made our respondents put more focus on the issues of the browser and the device instead of the photo search and the Search modes. The result lists weren't good enough as they contained a lot of missing images in the result lists. This was because all the hits in the results were taken straight

from the first hits returned from a Google query. Not many of these hits had any appropriate images that would look like they simulated an image search and some pages didn't have images at all and therefore images in the result lists were sparse. This in turn made the prototypes quite unsuitable to study respondents' reactions to images in the result list and how images could make their search strategy different. There were some broken links among the results and this irritated some of the respondents. It didn't give the respondents a feeling of a fully working system.

Evaluation of method

The techniques used were quite satisfying as were the amount of data, especially from one of the respondents. The observations gave us quite satisfying results. This might have been because we were two researchers with divided responsibilities with one interviewer and one researcher taking notes. Of course we did not catch all possible data but we got enough to find some interesting issues concerning the user interaction. It would however be interesting to compare the time respondents' used for a search task with the qualitative data from interviews to see if there are any interesting patterns.

Even though we can't know if we would have gotten different results if we conducted the test in another context, we got more data than we asked for. Of course there will come a time when the lab environment doesn't give any more interesting data but that is a bit further into the future. The prototypes and research methods can still be refined and more data collected in a lab environment before extending the study to other contexts.

The group of respondents was well suited for this part of the study as they were used to this way of working and showed interest to the subject. They were not many but we got a good deal of data from them that would have needed many more respondents to find if they were randomly picked out of a larger population. However, the amount of respondents we think was not sufficient. Even if we had no intention to generalize over a whole population, the data can only just be used as inspiration to some new hypotheses. It seems respondents have no problems understanding the concept of image search but this say nothing about how people from outside a systems design environment will react. The same goes for the method of using prototypes and interviews. It worked for this group of respondents but that doesn't necessarily mean it will for another group. We think the test might need to be a bit stricter, both the respondents' tasks and the interviews, to efficiently collect data from a wider population.

The way of interviewing in a semi structured way worked well with this kind of respondents. The respondents seemed to consider the questions and subjects we wanted to discuss thoroughly. Some of the interview questions didn't really fit into the flow of the interview so they were skipped. But this was not any greater loss, as there is an overall feeling that the questions were a bit too many anyway. Some questions gave more useful data than others and these were questions 1, 2 and 3. Perhaps better results would have been obtained if we had fewer questions and a few more respondents.

The objects used for searching in this round were satisfying, however with both objects being unmarked, we felt we wanted to compare one unmarked object with a marked to see if there is a difference in respondents' search behaviors.

The answers to all the four search questions could be found within one single Web page making the search questions a bit redundant. The respondents also seemed to have some trouble to remember all of the questions during the test.

We found that it was not too stressful to have four respondents doing the test during one day. The interview with respondent three went on a bit longer than the other interviews, and after that the fourth respondent gave us as much data as the first two.

Issues found

A drawback that can arise is that an image matching system can have difficulties recognizing and excluding copies of the objects in the photo from the result set.

Search modes seem to be helpful in as a way of filtering the results. The amount of steps that a Search mode presents to a user shouldn't be too many as too many steps may loose user's interest. But the more steps, the more the user can filter the results and the higher the chance is to be presented accurate hits. A geographic location sensor like an embedded GPS, connected to the photo search, could decrease the amount of steps. However, it looks like in some cases users doesn't want perfect hits all time, this might be true when users have vague ideas of the information sought. People may not like to be exposed to advertisement or offensive material so it may be suitable to provide a step where one can exclude such images. An option apart from providing too many steps could be to make the prototypes more transparent, so the user can do changes e.g. while evaluating the result list.

Different products may fall under different shopping behaviors. Modes like Shopping mode may have sub-modes like shopping artifacts or services.

The information elements that are of most importance seem to be the Snippet, the Domain name and the thumbnails of images on Web pages. In the Initial phase, we found equivalent results but in the Initial phase respondents were more positive to URL than Domain. It seems like Title is of less importance and that the size of Web pages in bytes and the information element of last update can be of importance. By looking at the results, we believe that the respondents' taste of information elements in the result list is very personal, this indication we also found in the Initial phase. An ability where the user can choose what information elements should be presented in the result list could be appropriate. However, a larger test group could show patterns. There is still a possibility that the simplicity of not having a choice is better than the flexibility of having too many.

Because Web pages normally aren't designed for mobile phones they are often very large. Navigating these pages may not be an easy task. The user should be able to reach the information inside a Web page quick and easy. One suggestion is that the user arrives where the image is located assuming that the information sought is closely located to the image.

As people aren't used to not include keywords when searching, the requests to do so may decrease as time goes by. Nevertheless, it seems that combining words with a photo may increase the efficiency even if it the workload increases from just pushing a button. Strategies that people use when they search on a desktop computer may not be applicable to the prototypes as they look today, however, it would be interesting to see if strategies like "open many links at the same time" is possible to apply to mobile environments.

Conclusion

In conclusion we found that photo-based search is a novel way of searching, but quite easy to understand and use for information retrieval. The usefulness is especially notable in mobile situation and unfamiliar places. However due to the limited interaction in mobile situations, there is need for more active help in the search process. Search modes are considered useful for this task, but they may need to be configurable to personal needs.

After carrying out this part of the study we felt that there were many more issues to be revealed if we rearranged the method, made new prototypes and used a more modern mobile device. We

therefore decided to make another round of prototyping, the Second round of prototyping, which also focused more on if a system using Search modes can deliver better user experience than a system not using Search modes.

Second round of prototyping

This is the last part of the study. As the First round of prototyping, it explores how people experience searching the Web using photos and a mobile device, but this time with some changes to the method and the prototypes. The results from the First round of prototyping indicated that Search modes could be interesting and therefore we wanted to continue the study. The focus for this round beside the main purpose of finding issues is to see if Search modes can deliver better user experience than a system not using Search modes, and if there are any differences between marked and unmarked objects. The method was changed to include video observation and questionnaires. Further, the interviews in this round are more structured and the number of respondents has grown from four to ten. Instead of having one group of respondents that tested the prototypes, we used two groups of respondents. This was to compare the user experience of two different prototypes. The prototypes were redesigned in many ways, now using a PDA instead of a mobile phone, the result set are now true subsets of each other and no longer separate. There are still two objects to search for, but one of them is marked with text in this round. The results indicated that even if it took longer time to find the sought information using Search modes than non-mode, the perceived value in time efficiency was higher using Search modes. Further, the results indicated that a photo-based system is more useful when the search type is an unknown, unmarked object and less useful when the object one wants to find information of is not in the user's vicinity, or phenomenon that one can't see, like feelings or atoms.

In the First round of prototyping, the respondents spent a lot of time navigating and scrolling inside Web pages, it looked like it was hard given that small screen resolution. Users tend to dislike horizontal scrolling on handheld devices, vertical scrolling seems to be a bit more accepted. To avoid horizontal scrolling as much as possible there are two future possibilities. Either the screen resolutions of devices will grow or, and as more and more people access the Web from such devices, Web pages will be redesigned to fit smaller screens better. More likely, there will probably be a compromise between the two (Etemad & Newth, 2004).

The device and browser used in the First round of prototyping just wasn't good enough. Browsing the Web on such a device isn't practical and there is no use trying out a photo search system on it when there is a feeling that no one will use it for browsing on a large scale anyway. There is reason to believe that mobile phones in the future will have higher resolution displays. It's not feasible to redesign all the Web pages in the search result lists so instead we were looking for an alternative device with a higher screen resolution. We found what we needed in an Ipaq PDA.

With the Ipaq we did not only get a higher screen resolution but also a browser that handles special characters better, scripts and other features of Web pages. The Web browser used in this Second round of prototyping was the version of Internet Explorer that comes with Windows CE. Also the Ipaq has support for wireless access to a local network which means we no longer had to

store all the content used for the prototypes locally, but could access Web pages directly on the Web and therefore also make it possible to follow all hyperlinks on all pages, providing a more realistic browsing experience. In addition, the Ipaq had a built in digital camera and although we never used it physically in the study it does add up to the feeling that photo-based Web searches really can be made with such a device.

Method

As mentioned earlier, the method used in the First round of prototyping has changed. This round has more of an experimental approach than the First round. In experimental studies like this one, McDonald and Tait (2003) points out that it is important to use a variety of techniques within an experimental context and that performance data is important when testing particular hypotheses. Further, they point out that verbal data used in combination with an evolving search tasks can provide more detailed insights into the search behavior that accompanies or leads to performance outcomes. A mix of qualitative and quantitative gives the possibility to triangulate the data (Yin, 2002). These arguments and experience from round one resulted in a method using many different techniques.

Experimental Studies

In our evaluation, we focused as McDonald et al. (2001) did, on the use of controlled experiments, with measures of efficiency and user perceptions of efficiency and effectiveness. There are of course a number of other methods available within this general framework; constructive interaction scenarios, protocol analysis, semantic differential scales, etc. The evaluation of round two of prototyping focus on the navigation of a mobile photo-based search system, not the infrastructure such as algorithms or bandwidth, although we don't deny its influence on the usability.

We divided the respondents into two groups and manipulated one variable between the groups and one variable within each test to compare responses. We have also tried our best to make the experience of testing the prototypes as similar as possible to all respondents to create comparable data. This means the experiments in the Second round of prototyping has been more controlled than in the first round, but it is still not a very strict experiment, and this is not necessary, as we're not trying to generalize our findings to a larger population.

The respondents were divided into two groups, A and B. Group A used the system type All, which conducted a straight search without any modes. Group B used the system type Mode. All

respondents in both groups first did a search with the known object (CD) as the reference, and then they did a search with the unknown object (Iittala). The independent variables were System Type (All vs. Mode) and Search Type (Known object vs. Unknown object).

Respondents

The participants who took part in the experiments were all students at Lund University. As McDonald et al. (2001) points out, the use of students as experiment respndents is justified because the system was intended for use by a wide population of general users which will certainly include students. In the First round of prototyping four respondents were used, this we thought were enough because the intensions were not to generalize the results. However, we felt that it was difficult to make the analysis out of that few respondents and we hoped to get richer data if the test group was bigger. We also wanted two groups of respondents that would test different systems to make a comparison and this also calls for more respondents. Further, we wanted to try the prototypes not on informatics students, but on a more general group of students. Ten volunteers were used in this round of prototyping, none of them studied at the department of informatics. The group consisted of four female and six male respondents aged 23 to 26. We realized that using only 10 students as respondents still makes it hard to generalize the results to a wider population but as mentioned before, this was never our intention. We wanted to further explore the issues of mobile photo-based search and we don't need any generalization to do that.

None of the respondents were involved in any of the previous parts of the study. This so they would all come into our lab with equal knowledge about the subject. This is not necessarily an advantage. If we could follow up the respondents from the Initial phase to the final round of prototyping, the feedback might be greater the further they got into the idea. However, we wanted to compare non-biased persons that had never tested the previous prototypes to be able to make comparisons between the two groups. So the composition of the group of respondents is a reflection of our intention of running the tests more like an experiment.

The ten respondents were divided into two groups. They were placed into groups by the order they arrived, so that the first person who could come was put into group one, the next into group two, the third into group one, and so on. Respondents were not informedof the group system and that we tested two different prototypes with the groups.

The test environment

The environment for the First round of prototyping did well. However, we wanted a more natural setting for this round. One extreme was in a relaxed atmosphere like a café. The advantage of such environments is that it is a public place that might make the respondents more relaxed when not feeling inferior to the researcher. The disadvantage is that such environment can be too noisy for sound recording and for the respondents to be concentrated on their tasks. The other extreme was to carry out the study in a laboratory environment. As mentioned in round one, there is little value to be added taking the usability study out into the field unless experiments in the lab doesn't provide any more valuable data. The advantage with the laboratory is that we can prepare for each test person and that it is equipped with computers and other facilities that public areas are not normally equipped with. Another advantage is that in a café it is hard to know how many guests there will be at the time and as a result hard to predict the level of noise. In a laboratory, you can on the other hand exclude the outside world from the study. We found public areas in Lund where we detected a satisfying wireless network that could be used but these areas were all too noisy and in areas where the noise levels were satisfying the network was not stable enough. However, if we had found a public place that was not too noisy and that had wireless Internet, we would have considered conducing this part of the study there.

The place selected, as in the first round, was the interaction lab at the department of informatics, Lund University, which we thought had a creative and a relaxed atmosphere. Further, the wireless Internet connection was very stable and fast. In the first round, we used respondents who already felt comfortable in the interaction lab, as they had worked there before, for this round the respondents were all from outside the department. A difference from round one was therefore to get the respondents into a relaxed mode. To get them in a relaxed mode a seating area was arranged, placing chairs around a low round table with a screen separating the area from the rest of the laboratory, all to create a less strict environment.

Duration

In earlier studies like IDeixis (Tollmar et al, 2004), they found that 15-20 minutes is a good time limit to keep in mind for each test person. Make the test session longer, and the test persons will be unfocused. But if it is shorter it will be hard to get sufficient data. The estimated duration, based on the First round of prototyping, was set to 45 minutes for each respondent including: introduction, the test, interviews, and a survey sheet. The results later showed that the actual test and the interview took 20-30 minutes and the total for each respondent about was 45 minutes, as

predicted. Other studies have set a time limit for their respondents when doing a search experiment as for example Lazonder et al. (2000) did. However, in the First round of prototyping we didn't state a time limit for the respondents to do the test, this because we had so few respondents and it would be a shame to set a too low time limit just to end up with even less data. However, if the respondent would not be getting anywhere we gave ourselves the right to end the experiment. This gave us the possibility to continue observing a behaviour or discussion that might seem interesting.

In round one we found that it was possible to handle four respondents during one day. The test period for round two therefore covered three days with 3 to 4 tests per day. Our aim was to do all the test in a period of time as short as possible to keep the method intact so there would be no major difference from one test person to another. We are aware of that in all such studies the researchers experience from previous tests affects his or her behavior during the ones after, hopefully in a positive way, for example that the interviews get better and better. To run too many tests during a day can also result in the researchers not focusing as much on the later tests as the earlier.

Objects

Instead of using two unmarked objects as in round one, we wanted to study if there are any differences in search strategies by using one marked and one unmarked object. This was to find out if the respondents would react differently on searching for the two types of objects using photo search. The marked object was a CD, Radiohead's album "OK Computer", and the unmarked object was a glass tray called Iittala designed by Alvar Aalto (Figure 19).

Figure 19. Both objects used for the prototype searches.

Respondents' tasks and search questions

We started each test session by letting the user do an ordinary text-based search on Google using the PDA. This was to introduce them to the device and the browser, but also to give the respondents a feeling for what a text-based search is like when you don't have the normal input facilities of a desktop computer. We let the respondents try any query they could come up with, but if they would have a sudden lack of imagination we asked them to search for their own name.

The next step was to let the respondent do a search for the marked object using the prototype. Here the respondents were told to come up with a search question on their own, and if they couldn't, we had a backup question: "What year was the CD produced?". We wanted respondents to come up with their own questions to produce a stronger feeling that we're testing the system and not the respondent's skill. When the respondent seemed happy with the information found we moved on to the last step of the test, searching for information on the unmarked object.

For the unmarked object two search questions were used: "How much does it cost?", and "Where can I buy the object?". In round one we used four search questions and we found this to be a little too many to keep track of for the respondents. And also, all answers to all four questions tended to be located within the same Web page. Here we wanted all respondents to complete the same search task to be able to compare respondents search behavior but mainly to compare the two groups using different prototypes.

Video observation

In the First round of prototyping we took notes of comments and observations during the test. This technique was satisfying in that part of the study but at this stage we wanted to try a new well-established technique to create a better recording of each session, that is video observation. This technique was possible to do because we were two researchers doing the study, where one of us could introduce and conduct the interview, and the other focuse on handling the video camera. The advantage of this technique is that we receive detailed video and audio data that we could play and analyze over and over again. Other options would have been to either use a tape recorder or taking notes. But neither of these would collect any detailed data of how respondents navigated the prototype. With the help of the video recording, it was also possible to measure the time of each search task without putting pressure on the respondent by having a timer-clock on the scene. One of us filmed each session from a distance behind the respondents shoulder to lessen the attention to the video camera. We told respondents the intention was to film the screen and not

their faces. The video recording started just before each experiment, and ended after the interview. During the interviews the camera was used solely as a tape recorder. That is, both researchers sat down at the table, one with the video camera in the knee pointing in another direction but still capturing the sound of the interview. There was really no point in filming the respondents during the interviews because we we're not interested in analyzing the interviews on that level. In the first round we told the respondents to talk loud about what they did and why during the study without the use of recording. In this round we got better data because we got it all properly recorded.

Interview

For each respondent, an interview was set up after the test was conducted. The questions asked regarded the respondent's perception of the prototype. For each respondent six questions were asked about the prototype and the overall system. For the respondents that tested the mode prototype, three more questions were asked concerning the Search modes. Some of the questions are similar to questions in the first round. This was simply because the answers to those questions gave us interesting data in round one, and we believed that more data could arise reusing them in this round. As mentioned about design of questions in the first round of prototyping, we were more interested in encourageing the respondents to talk about their experience and to reflect upon certain issues than to answer the actual questions. The overall trend in the questions in this round of prototyping is that they no longer focus on the information elements in the result list but more on the general experience of using a mobile photo-based search system. The interview guide can be found in Appendix 5.

Questionnaire

For this round we wanted some quantitative data to complete the qualitative using triangulation. The questionnaires purpose was to compare the perception of the respondents of how text-based searches and photo-based searches can help a user to find certain information. Apart from measuring the time to complete a search task, we also gave the respondents a questionnaire to be filled out after the interview. Even if the answers could not be analyzed in a quantitative way, due to the small amount of respondents, we hoped this would give us further insight to the answers of the interviews. Five questions were asked with the range of rating for each question 1-5, where 1 is poor and 5 excellent. The questionnaire can be found Appendix 7). The reason we wanted to do the interview before the questionnaire was to make the respondents start thinking and reflecting about their experience before filling it in. Of course there is a risk that the interview may

have affected their answers in the questionnaire but if they had been given the questionnaire right away after trying out the prototypes they would probably have trouble understanding the questions in the questionnaire.

Analysis

The answers from the interview and issues that we observed on video were clustered (Miles & Huberman, 1994) into categories that emerged during the clustering and every category was given a name. We wanted to do this because analyzing questions one by one like we did in the First round of prototypes gave us insight into just the particular question and not the larger picture.

The diagram showing time to complete the search task, the rating diagrams from the questionnaires and the clusters of observation and interview data were analyzed separately and then together (Figure 20). When looking at the respondents' interview answers, we found as we expected that issues derived from the answers didn't necessarily become clustered under each question. One respondents answer to one question could be discussed by another respondent during another question. This is of course because of the nature of semi-structured interviews. The issues that we found interesting when studying the video observations were clustered into the same set of clusters as the interview answers. The reason of this is that they carried information about the same issues as the interview answers, and because of their nature, they could not be presented in a diagram. The clusters are not isolated from each other. Some of the issues could be placed under many clusters.

The time it took to find the information the respondents needed were presented in a time-diagram. The only comparable times in this study is the time it took for respondent's to find the sought information about the unmarked object as all other searches differed from respondent to respondent. Out of these three (Clustering, Time Diagram and Rating Diagrams) we did a final analysis (Figure 20).

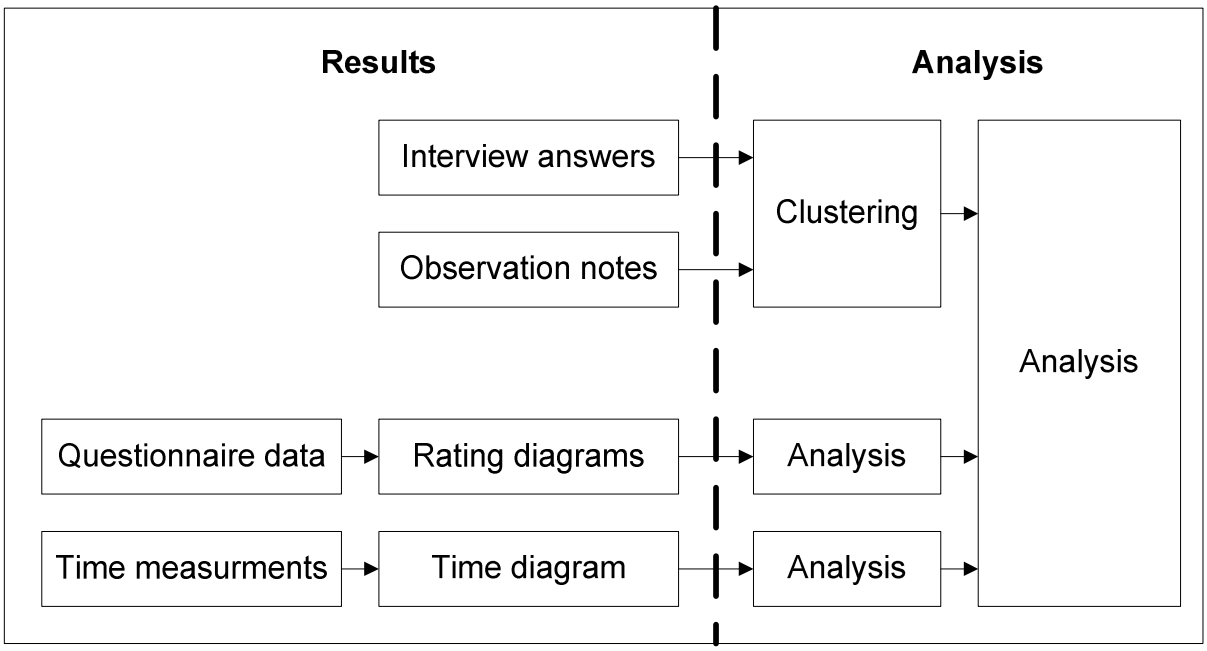


Figure 20. Model illustrating the process of how the results were analyzed in the Second round of prototyping.

Both researchers measured separately from the video recordings the times it took for the respondents to complete their tasks. The times were crosschecked between the researchers to make sure there was consensus about when tasks were started and completed.

Both researchers analyzed the video material by taking notes of the respondents' answers. Everything else that seemed interesting like behavior and comments not related to any of the interview questions were also noted. As two researchers did this separately for the same material we added researcher triangulation to the method triangulation (Yin, 2002).

Data from the questionnaires was presented in diagrams that show all respondents ratings and trends for both groups.

Together both researchers clustered interesting issues from the observations and the interviews using post-it notes. The names of the clusters were defined after clustering the issues. Each post-it note was marked with the originating respondent and which question it was connected to. If it originated from an observation this was also noted.

The prototypes

The prototypes of the second round were created using a modified version of the first prototype generator. The new prototypes handled the results of Search modes and mode options as true

subsets. For example, all hits found in Shopping Lund also appear in Shopping Sweden among the other hits for Sweden. The choice of options in Shopping mode has been changed from [Lund, Sweden, Europe] to [All, Sweden, Lund, Malmö]. The results of Lund and Malmö are true subsets of Sweden that in turn is a true subset of Europe. Europe is a true subset of Shopping All, which is a true subset of All. This is illustrated in Figure 21 and Figure 22. All queries used for different Search modes and options can be found in Table 2 and Table 3. The number of hits and error rates for each mode and mode option are presented in Table 4. The error rate is the percentage of hits in a results subset where the object in the image does not contain the sought object. The target error rate was entered into the prototype generator that suggested hits from the result of the query or the error query to reach this target. However it proved to be hard formulating search queries that gave good hits so our target error rate of 20% gave varying results as can be seen in Table 4.

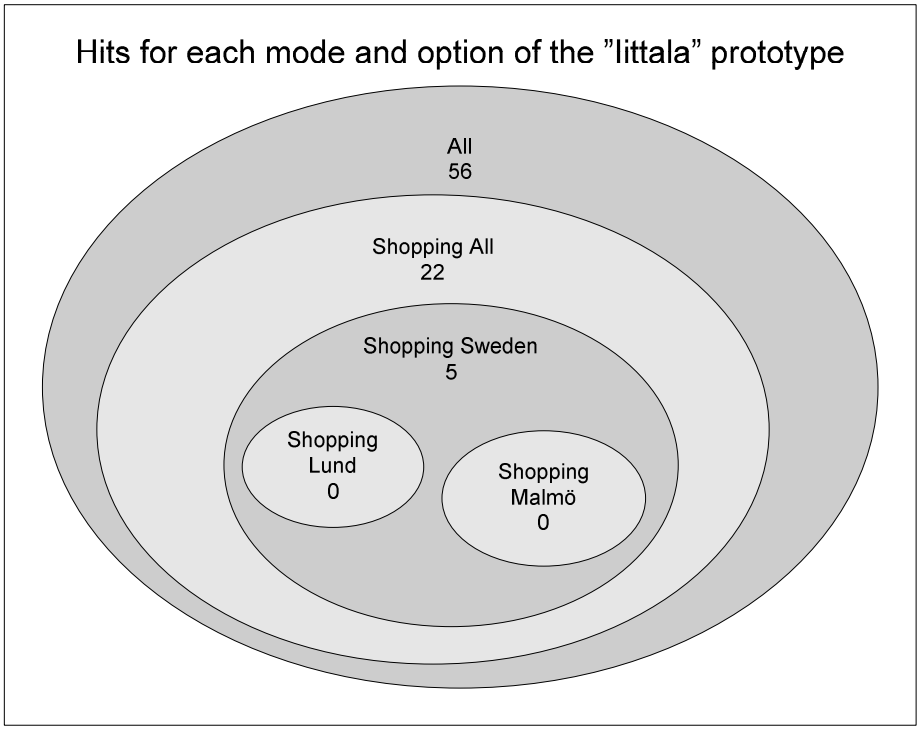


Figure 21. The number of hits in each mode and option for the Iittala prototype.

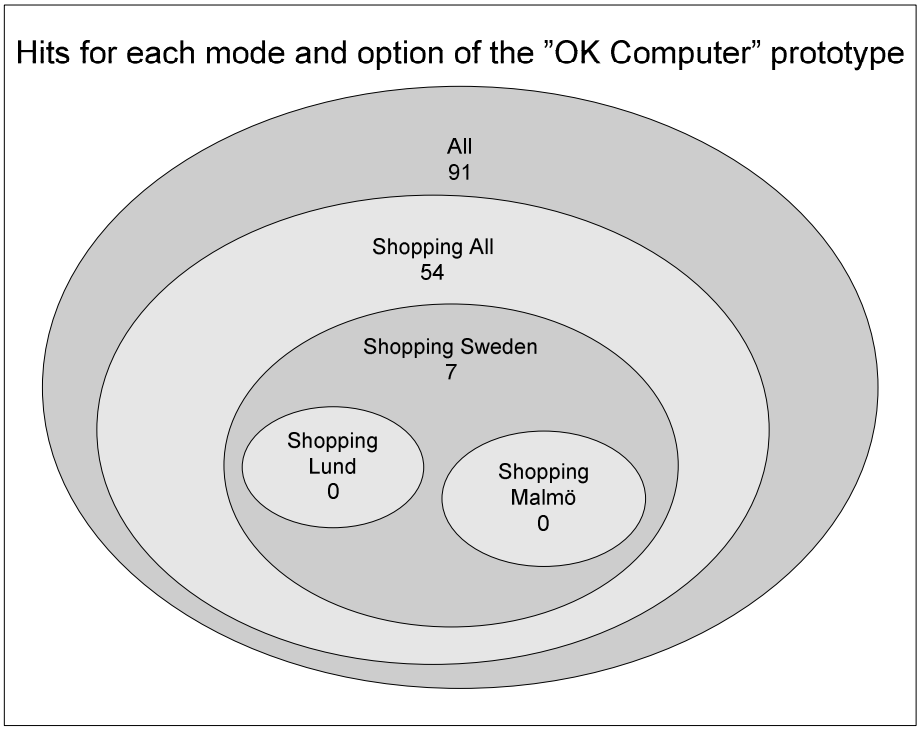


Figure 22 The number of hits in each mode and option for the OK Computer prototype.

Table 2. Queries used in the OK Computer prototype

	Query	**Error Query**
All	"ok computer"	"radiohead cd"
Shopping	"ok computer 1..10000 pris OR price"	"radio head cd 1..10000 pris OR price"
Sweden	"ok computer 1..10000 kr"	"radiohead cd 1..10000 kr"
Lund	"ok computer 1..10000 kr lund"	"radiohead cd 1..10000 kr lund"
Malmö	"ok computer 1..10000 kr malmö"	"radiohead cd 1..10000 kr malmö"

Table 3. Queries used in the Iittala prototype

	Query	**Error Query**
All	"iittala aalto"	"glas tray OR fat OR modern OR nordic OR design"
Shopping	"iittala aalto 1..10000 pris OR price"	"glas tray OR fat OR modern OR nordic OR design 1..10000 pris OR price"
Sweden	"iittala aalto 1..10000 kr"	"glas tray OR fat OR modern OR nordic OR design 1..10000 kr lund"
Lund	"iittala aalto 1..10000 kr lund"	"glas tray OR fat OR modern OR nordic OR design 1..10000 kr"
Malmö	"iittala aalto 1..10000 kr malmö"	"glas tray OR fat OR modern OR nordic OR design 1..10000 kr"

Table 4. The number of hits and error rates for Modes and Mode options in the prototypes

	Total hits	**Hits**	**Hits from subsets**	**Errorrate**
Iittala:				
All	56	34	39%	74%
Shopping All	22	17	23%	53%
Shopping Sweden	5	5	0%	80%
Shopping Lund	0	0		
Shopping Malmö	0	0		
Ok Computer:				
All	91	37	59%	19%
Shopping All	54	47	13%	26%
Shopping Sweden	7	7	0%	57%
Shopping Lund	0	0		
Shopping Malmö	0	0		

The new prototype generator can create an overall better feeling of good hits as it suggest hits for each mode and option and we were able to manually specify which hits to include, which to leave out, and which of the included hits were that good matches and which were not. This made it possible to avoid hits that were completely wrong and to introduce a known number of erroneous

hits, i.e. pages containing images similar to the submitted photo but not containing the very same object.

The generator allows setting of a desired error rate, and asks if each page is a correct hit or not. It uses two queries for each prototype, one query that should for creating hits as good as possible, and the other to create erroneous hits, i.e. pages with images containing similar but not the same object. All hits are sorted by their position in the same order as Google presented them at the time. This ordering is kept constant over all groups in the study. This rating also becomes a part of the simulated filtering (Figure 23).

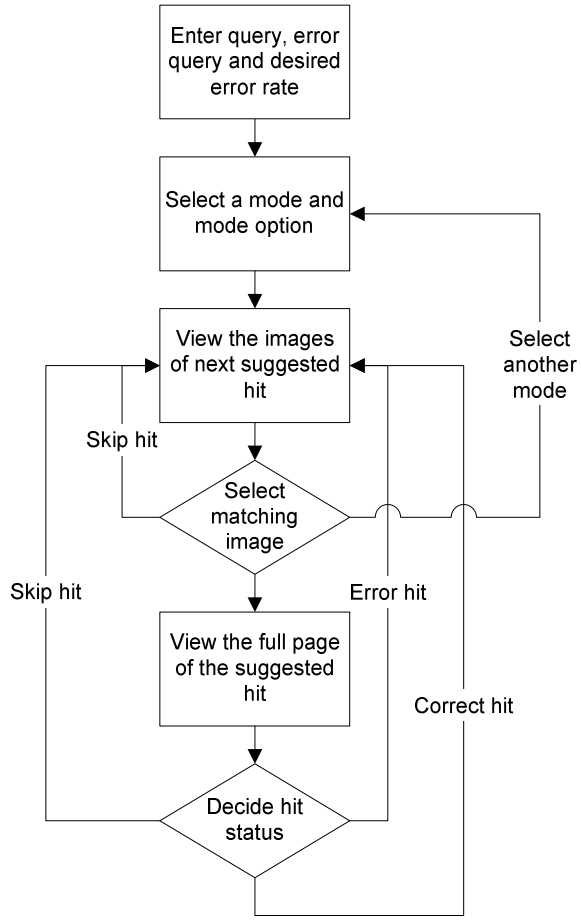


Figure 23. Flowchart describing the process of creating a prototype.

After the two queries have been entered, the search results are built up for each mode option. Then the generator suggests hits and shows them one at a time in the browser. For each hit there is a choice of rejecting it, accepting it as a correct hit, or accepting it as an erroneous hit. If the hit is accepted the next step presents all images of the hit page. One of these is selected to be the matching image for the search result list.

This is a form of manual filtering used to simulate a feeling of an automatic filtering to the user. A misunderstanding of this would imply that this is not a filter engine but a directory, because people and not an algorithm select the hits. But it does create the user experience of a filtering search engine, not a directory.

At the end of the process the generator saves two versions of the prototype to files in two separate folders, one using Search modes and one that doesn't. The one not using Search modes is basically just limited to the All mode. The hits it shows are identical to the ones in the All mode of the Search modes prototype.

The result pages are no longer limited to 10 hits for each mode or mode option. When a prototype is created, the only limit to the number of hits is how many hits the Google query has returned. To avoid having all the hits in one single result page, the hits were presented five at a time with next and previous hyperlinks (Figure 24 and Figure 25).

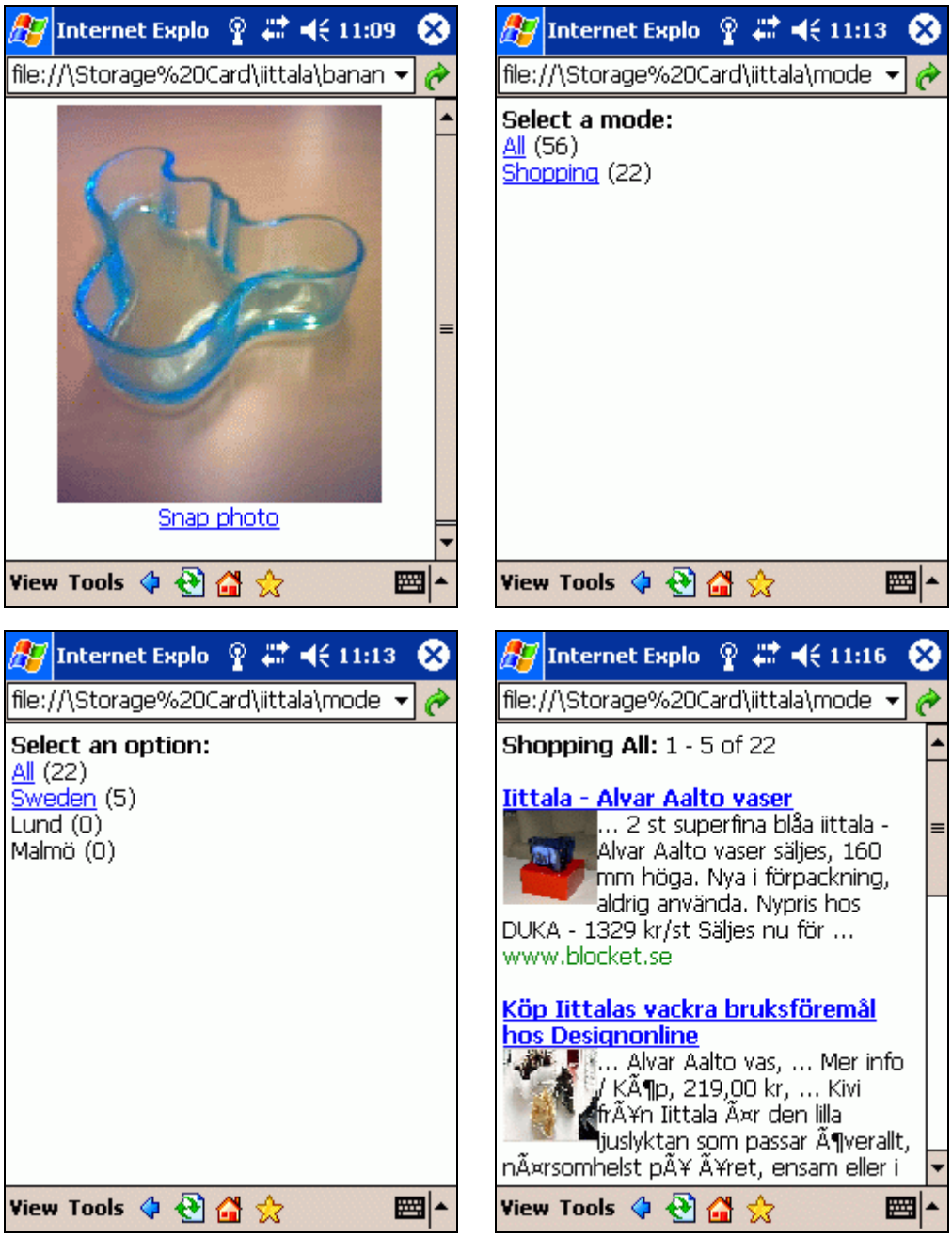


Figure 24. Screenshots from the Iittala prototype implementing modes.
The snap page (upper left), the mode selector page (upper right), the mode options selector page (lower left) and the first page presenting hits of the All mode (lower right).

Pages from a specific domain are returned a maximum of 4 times from each query. This is to avoid too many hits from a single domain. But as the Search modes and options are subsets of each other, pages from a single domain may appear more then 4 times in a search result.

In the first round of prototypes there were many missing images in the result lists. This was because some pages in the result lists didn't have matching images or any images at all. In the second prototype generator the hits for each mode are manually selected from a Google result set

and all pages do contain a matching image. However, the image selected to be the matching image is not stored locally, it is only a reference to the image on the originating Web server. This can lead to broken images if the image is removed from that server after the prototype has been generated. This did happen in the second round of prototypes as well, but there is still a significant improvement in matching images compared to the first round of prototypes (Figure 25). As mentioned in the theory in the chapter about search engines, in a real context this image would come from the index database and therefore always be available.

Figure 25. A broken image at the bottom of the result page.

Design issues

Kaikonnen and Roto, (2003) set up some guidelines regarding design that we were aware of but didn't follow strictly. They point out that some of their guidelines are not necessarily important in the mobile environment because of the nature of mobile browsing. In devices that do not provide a pointing-device, it is not important to make obvious what's clickable. The prototypes we built used a clickable screen (it used a Stylus pen) we had this in mind when designing the prototypes. Other guidelines like "Focus on user goals and needs" are similar to the one of desktop computers. The design decisions made are discussed below but can also be found collected in Table 5.

Five hits on each page might lessen the need for scrolling and decrease the possibility of getting lost in the search result list whilst doing so. As mentioned earlier there have been findings that a user is willing to go through about two result pages using desktop computers before giving up and

this also holds true when using mobile devices (Jones et al., 2002). Kaikonnen and Roto (2003) found that in a mobile device related information should not be split on too many pages and that the optimal length of a page depends on its type (informative vs. interactive). They found that users navigating lists of items preferred a compact list of items on one page. From the list of all items, the user should be able to go to a detailed view of one item. point out that minimizing the number of steps in navigation does not increase users' feeling of insecurity. We decided to maximize the hits on each result page to five even though Kaikonnen and Roto (2003) propose that it's not wrong to show all the hits in one page. The mode and options selector-pages show the number of hits for each selection to aid the user in deciding where to go next and prepare for what to expect of each choice.

Kaikonnen and Roto, (2003) also points to the importance of the Web browsers back functionality, which it is the most important way to navigate backwards. When users need to go back several steps, hyperlinks to home and subsection main pages are useful; this was considered, as it was in the First round of prototyping. The hyperlink saying "Change search mode" at the bottom of result pages in the first prototype has been replaced with "Select another mode" (Figure 26). This was done because no respondents in the First round of prototyping used the hyperlink. It is only a minor change but we thought it might describe the functionality better and make the user more aware of this possibility. The prototypes didn't include back buttons as the browser itself provided this function.

Figure 26. At the bottom of each result page there is a hyperlink to select another mode.

We used the same color scheme as Google to make the users feel comfortable. Changing those colors to another set would probably not give any benefits except than differentiating the prototype from Google. We suspected that all our respondents would be familiar with Google as it is the most popular Web search engine as mentioned in the chapter of Search engines, and that all respondents in round 1 of prototyping were using Google as their primary search engine.

During the development of the system the importance of dealing with a certain percentage of returning correct images was considered. However, user testing in other studies like McDonald et al., (2001) suggested that users do not consider this aspect to be important.

In a fully working system, restarting the search and using a new photo should be possible at any time. This possibility could be given to the respondents, but to test that in our prototypes we would have had to exchange the method for one called Wizard of Oz as described by Preece et al. (2002). In this case the operator (Wizard) could control the quality of the result pages by acting as the search engine and typing queries during the experiment. However, to be able to do this the operator would need to know almost everything about the objects to type the queries and it would also delay the search results a bit. It would therefore not be feasible to do for this study.

For the prototypes of the second round, we changed some information elements for the presentation of the search results. From the First round of prototyping we experienced that the respondents thought that only the domain should be presented and not the entire URL in the search results list. This would save some space without compromising too much with valuable information used for evaluating the quality of the hit. Further, we found that respondents evaluated by looking at the matching image. In the interview some respondents said that they evaluated by reading the snippet. In addition, even if the respondents were of different opinions when it comes to the title, we thought that it was appropriate to keep it for the prototypes of the second round to make the respondents feel comfortable as they would recognize them from other common search engines. In summary, the prototypes of the second round contained the title, the snippet, the matching image, and the domain.

Table 5. Design decisions and their motivations for the second round of prototyping

Five hits on each serch result page	Decreases need for scrolling and getting lost
Presenting number of hits in each Search mode and mode option in mosde selector wizard	Guiding choice by telling user what to expect from each option
Changing wording of "Change search mode" to "Select another mode"	No users in round one used this functionality so we changed it to see if it was because they did not understand what it meant. We believe the later wording is more accurate and clear
Using color scheme from Google	The device screen presents color better than the device in round one. Using only black and white whould make it look dull. Decision to use Google colors as template is because Google is popular and users are familiar whith it's visual presentation of search results
Not giving users choice of snapping a new photo	This would be desirable but would require us to create the prototypes using completely different methods
Showing only domain instead of full URL	To save space and respondents from round one never seemed to use the full URL to decide what hits to select

Issues of the prototypes

The prototypes contained some issues that could be seen as a negative aspect for the study. First of all, the filename shown in the address bar shows the name of the object (as in the first prototype). This because the folders had the same name as the objects. This tips off the user regarding the objects name. Second, some images don't seem to match the original image at all, e.g. other CD covers with the same band. This can be "explained" by telling the respondents that the image-matching algorithm also uses text recognition. Third, the preferred mobile device would be a mobile phone with the same screen size as a PDA, but without a pointing device. The mobile device for this round of prototyping had a Stylus pen for input, not buttons like a common mobile phone.

Results and analysis

Time meassurements

The time it took for the two groups of participants to find the requested information using the prototypes shows a quite significant difference. All respondents found the sought information about the glass tray. In general the respondents in the group using Search modes seemed to take longer time finishing the task than those not using Search modes (Figure 27). The fastest search

time was just above 2 minutes and the slowest just above 8 minute. The average time it took for both groups to finish the task is about five and a half minutes. In the group not using Search modes the average time was about four minutes, and for the group using Search modes the average time was just below seven minutes.

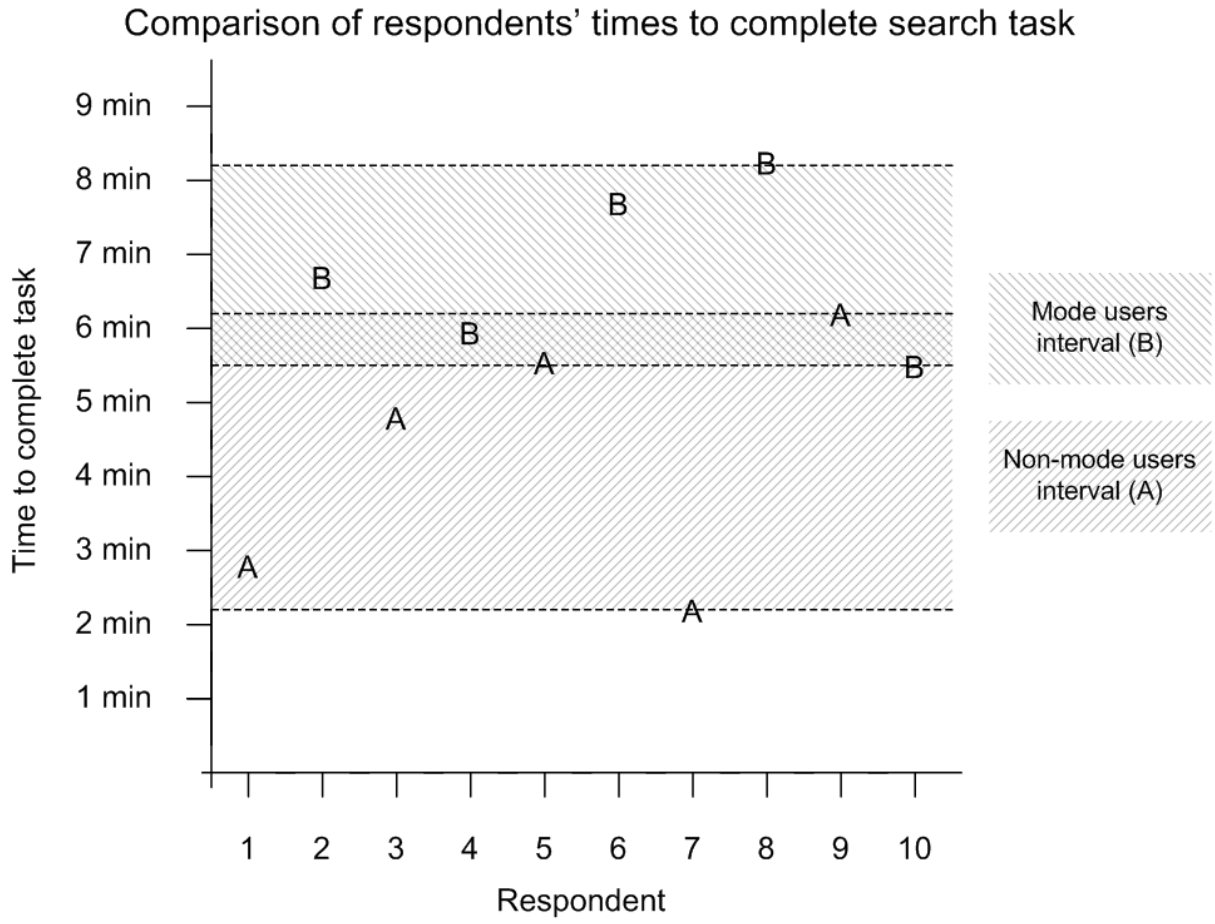


Figure 27. Diagram showing how long time it took the respondents to find the sought information. Group A is the group not testing the mode prototype.

The time difference between the two groups can to some extent be explained by the group using Search modes having the possibility to change mode while the group not using Search modes had their result presented in a linear list. The members of the group using Search modes first looked through the results in one mode and then switched to another mode to start all over again when the group not using modes just had one list to go through. This could account for a part of the difference in time between the two groups.

The Search mode Shopping-Sweden only contains five hits, and none of them contain the sought information. Looking at the videos, we observed three out of five respondents using this mode

seemed to spend a lot of time looking very carefully through all hits before realizing they can change mode and this might also be a probable cause for the time difference.

We expected the group using Search modes to find the sought information faster because the Shopping mode would provide better links earlier in the results list then the group using the non-mode prototype. This is true if entering Shopping-All, but because Shopping-Sweden didn't contain any correct hits, the user needs to go back and enter Shopping-All. This was perhaps an unfortunate flaw in the prototype, that Shopping-Sweden did not contain any correct hits. From the observation we found that the respondents, after choosing a mode, believe that they acted correctly and therefore were reluctant to go back and choose another mode.

Questionnaire

The complete questionnaires can be found in Appendix 7. To get a better overview of the results, the data collected in the questionnaires is arranged as rating diagrams. In these diagrams each arrow represents one respondents answer to a question in the questionnaire. The arrow starts at the respondents rating of text-based search systems and ends at the rating of photo-based systems. This means that an arrow pointing upwards indicates the respondent rated photo-based search higher than text-based for that question and vice versa. The length of the arrow indicates how much the rating differed. In addition the respondents are sorted by group with the results of the group not using Search modes to the left and the group using Search modes to the right. We decided to present the data in this way because the groups are too small to apply proper statistics to.

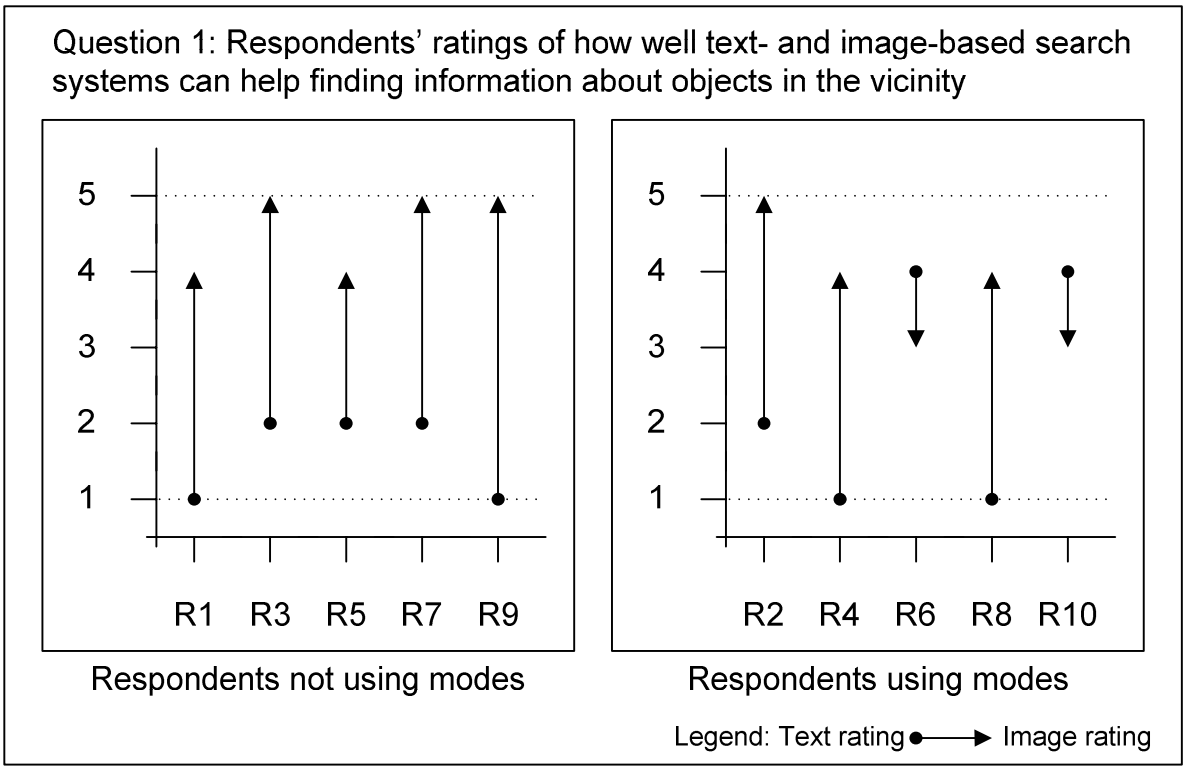


Figure 28. Data from questionnaire question 1 about objects in the vicinity.

Objects in the vicinity

All respondents that didn't use Search modes rated the text-based search low (Figure 28). The same group of respondents rated the photo search high. The other group, the one who used the mode prototype, was not as homogenous. Three of the five were more positive to photo-search than text-based search, and the other two rated photo-search lower than text-based search but their difference in rating is just one step on the scale making the arrows short. Overall, when looking at the length of all arrows, eight out of ten were very positive to photo-based search compared to text-based search when it comes to an object one has in the vicinity. Explanations of why the mode users were not as positive in this issue could be that Search modes made the photo-search a little to complex for users to be comfortable with, or that the prototype not using Search modes better communicates the idea of photo-based search.

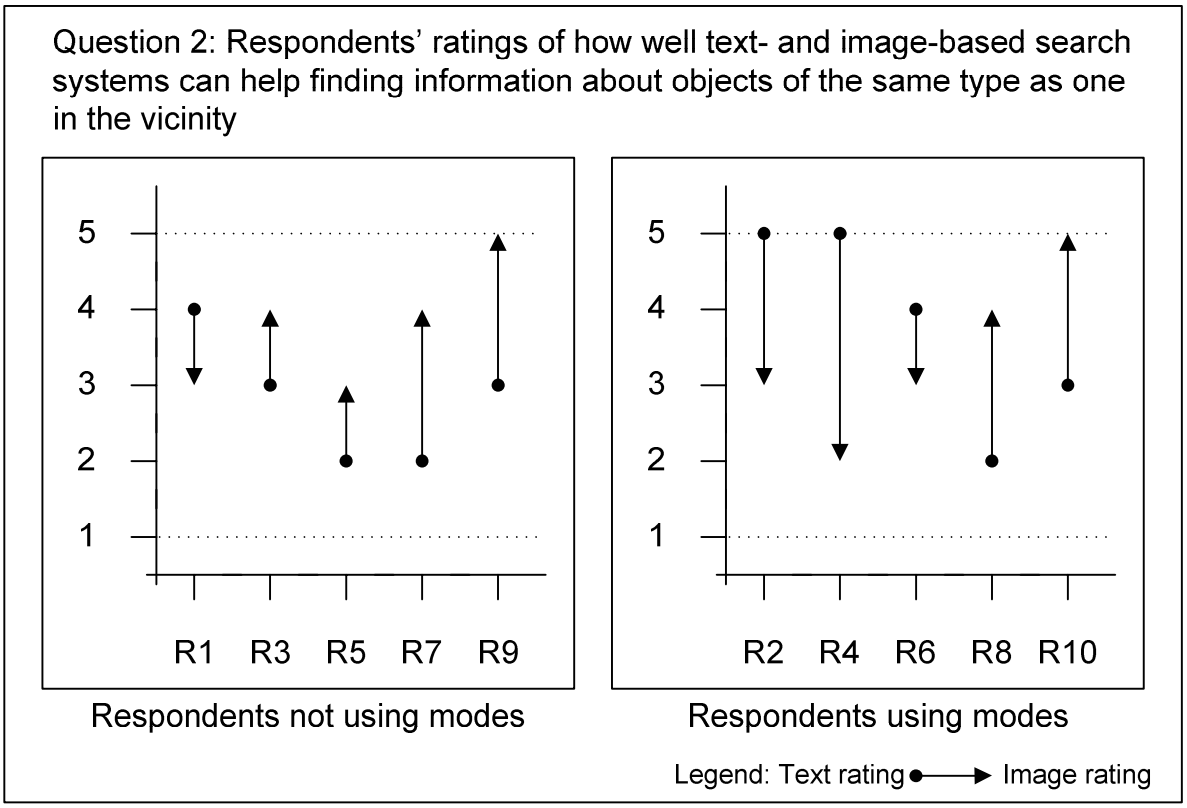


Figure 29. Data from questionnaire question 2 about objects of the same type as one in the vicinity.

Objects of the same type as one in the vicinity

Six out of ten were more positive to using photo search comparing with text-based search when it comes to searching on object of the same type in a user's vicinity (Figure 29). These six are however not very favorable for the photo-based search when it comes to the length of the arrows. The overall trend seems to be that our respondents don't really believe in a photo-based search's ability to find objects of the same type as another object, which is probably true as image matching in general has trouble with this. Remember that dogs can look very different and still they are all dogs. Once again the trend also seems to be that the group using Search modes rate the photo-based search lower and this could be explained by its complexity.

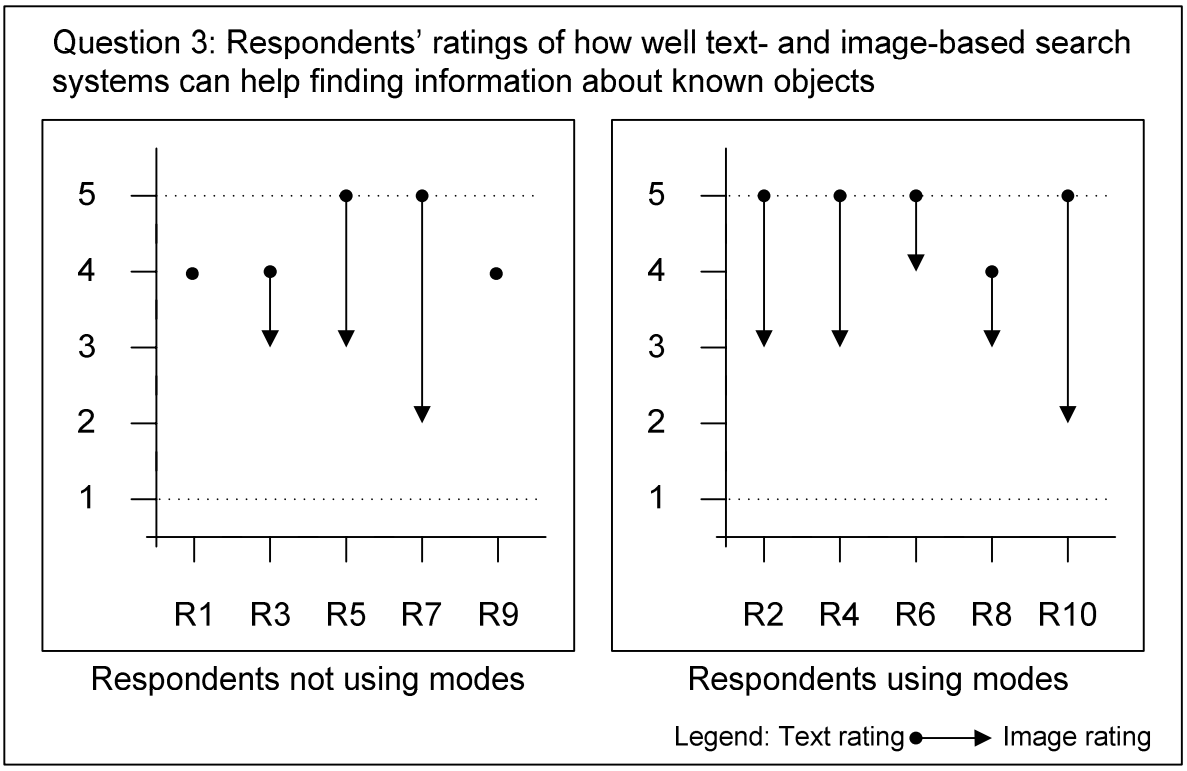


Figure 30. Data from questionnaire question 3 about known objects.

Known objects

None of the respondents believed that the photo search would satisfy their needs better than text-based search when it comes to searching for information about known objects such as objects marked with text (Figure 30). The ratings could be an indication that the error rates for the prototypes were too high, or at least higher than if one knows the name of an object and uses a text-based search. It's easier to match words and therefore it is more likely that the hits will be more accurate. With a system that would return better results the answer might be different. The results here may also be dependent on whether the respondents thought of a mobile situation or not when doing the rating. This is an example of where we would probably have got another result if we evaluated the prototypes in another context. If the respondents were not seated while doing the experiment but instead were walking on the streets of a city at the same time, they might have put more value into the ease of taking a photo instead of typing keywords into the device.

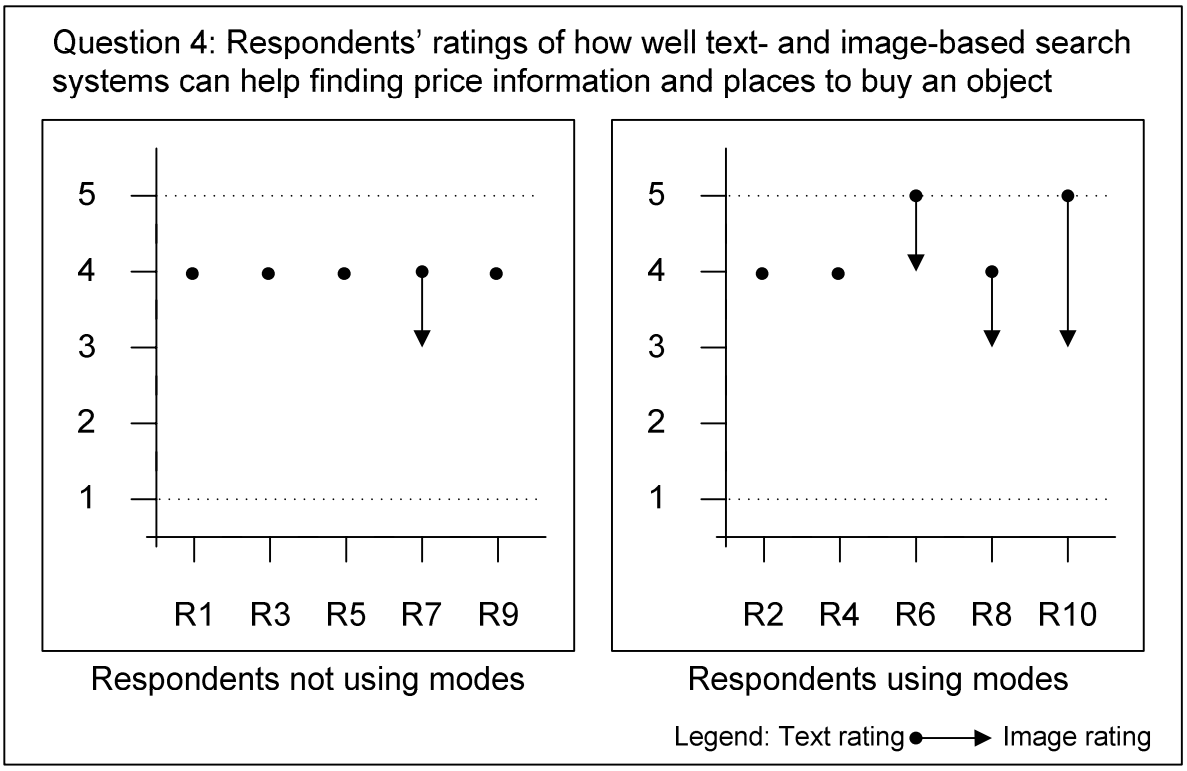


Figure 31. Data from questionnaire question 4 about shopping information.

Shopping information

Six out of ten respondents thought that both systems satisfy their needs equally when it comes to finding prices and places to buy an object (Figure 31). In the group that didn't use the mode prototype the consensus of equally good systems was more significant. All respondents tend to rate both text- and photo-based search systems quite high when it comes to finding shopping related information. This could be because the Web generally contains such information. Three of the respondents in the group using Search modes thought that text-based search was advantageous to photo search. The results indicate that even if the mode prototype was designed to find prices and places to buy objects, they still thought that a text-based search would be more helpful. However, none of the respondents believed that the photo-search prototype using Search modes was unusable as the ratings was 3 and above.

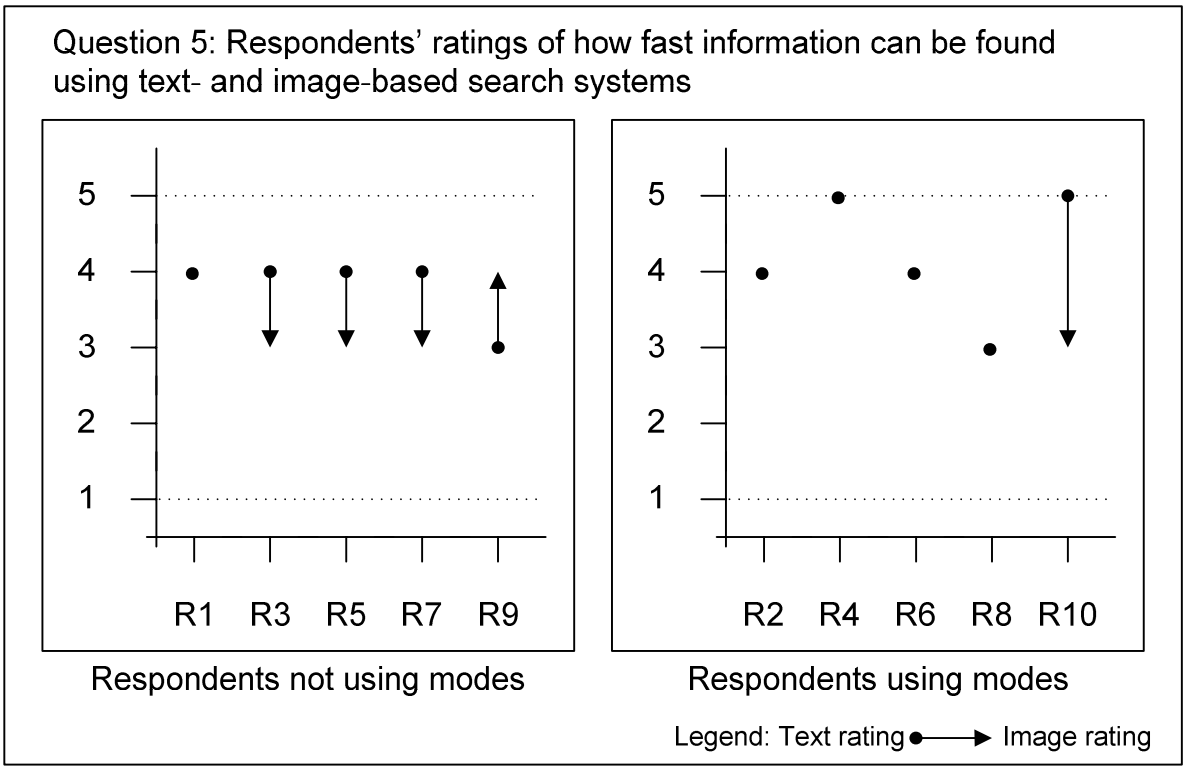


Figure 32. Data from questionnaire question 5 about time.

Time

Five out of ten respondents think that the two search methods would need the same amount of time to find sought information, and the others think it wouldn't be much difference (Figure 32). If a photo-search is more suitable in a mobile situation, because it demands less attention, the tradeoff of a little more time-demanding search process is probably acceptable to the users.

An interesting aspect about these results is when you compare them to the times it took respondents to complete the search task (Figure 27). The group that did not use Search modes tended to rate text-based search as a faster way of searching than photo-search. The group using Search modes took significantly longer time to finish the task and yet they tended to rate photo-search as equal to text-search. In conclusion: the respondents using Search modes spent more time on the task but still felt that photo-search was a little more time efficient than the respondents not using Search modes. Perhaps the Search modes make users feel more time efficient.

Clustering of interview and observation data

During the clustering of video observations and interview data we found 16 clusters of data that are preseted and discussed here. All data in each cluster can be found in Appendix 6. We have set category names for the clusters according to their contents:

- **The mobile device** - Actual and perceived limitations of the hardware.
- **Go to** - Issues of arriving at the wrong place in a Web page.
- **General attitudes and feelings** - How respondents felt about searching using photos.
- **Prototype** - Opinions and responses attributed to the prototypes.
- **Search with text queries and photo queries** - Comments on differences between the two.
- **Algorithms and accuracy** - Responses and expectations.
- **Image, Text, and a combination of these** - How information is commonly presented on Web pages.
- **Search mode** - Attitudes about the Search modes in general.
- **Search mode requests** - Comments on the Search modes and ideas of improving them.
- **Verification** - How respondents verify information to confirm that it matches the query.
- **Image features** - Issues with the taken photo.
- **Modality** - Which modalities respondents use when evaluating.
- **Objects** - About the objects, the tray and the CD.
- **Prerequisite knowledge** - Issues of knowing what kind of information common Web pages contain.
- **Areas of use** - What the system can be used for.
- **What's on the Internet** - Issues of what information and what images are commonly available on Web sites.

The mobile device

Two of the respondents thought they had some problems with the device, they were not used to it or they thought it limited them a bit. Respondent 4 thought that technical features of the device could limit good search result. Three of the respondents believed that the better quality the taken photo had, the better the result would be. Comparing with text-based search, a bad photo is like a misspelled text query or a bad query. Search engines like Google checks queries and suggests another if the suggested query get more matches.

They believe that a photo has to be of good quality to get sufficient matches. In a context of everyday life we believe that a good photo is a subjective judgment. Applying this on the photo matching system, a user might believe that a photo is of bad quality, but it may be good enough for the system to find and match other images. The opposite can also be true, the user thinks the photo is sufficient, but the system can't handle it, or presents bad matches. Issues regarding the device when increasing the photo quality are the camera (lens quality, resolution etc) and the image-matching algorithm. Most interesting is how users perceive what a good photo is. When is a photo good enough? Is there a need for trial and error? It would be interesting to study what the general population believes to be a good photo. This could be used for research for the quality and focus of algorithms. Probably the user expectations will increase over time as the cameras increases in quality.

Technical aspects on how a device should be designed, like a horizontal screen is of great importance. However, it's more of a design issue for the hardware and not about the functions of applications.

Go to

Three respondents found it difficult to find the image and the text information belonging inside the Web page.

Web pages today are not fully designed for small screens. Respondent 4 suggests that instead of redesigning all Web pages one let the system automatically focus on the part of the Web page where the matching image is located. This might lessen the use of the scrollbars. This solution was discussed during the First round of prototyping.

General Attitudes and feelings

Respondent 3 expect that the system, at a start, will be overused, that everyone will take photos and search for everything in their way. However if this is true, people will learn how the system

works and when to use it. Further, they might get a feeling of what a good photo is for searching as discussed under the first cluster. This is probably not an issue, it will just train users and help spread the word about the service.

Generally the respondents are very positive to the overall system, especially using it for unknown objects. Overall respondent 1 was against new technologies, but searching with photos could be useful, especially for unknown objects.

Prototype

Four of the respondents commented the existence of several hyperlinks to the same Web page in the result list. When presenting multiple results to the same site, Google uses "Indented results" to warn users that the hyperlink is to the same source. One could test the user satisfaction and apply the same feature on a new prototype. Maybe there are other ways to solve the user satisfaction, maybe it's not of importance to have several hits, and instead the system could exclude replications by filtering recurrent Web pages. In the prototypes the amount was set to a maximum of four from the same domain.

All participants in the group that used Search mode reported that they comprehend the relationship between the Search modes as a subset relationship, which was our intension after receiving feedback from the respondents in the First round of prototyping.

Search with text queries and photo queries

In general, the respondents seem to think that text-based search is more accurate and of more use than photo search, but that it's faster to snap a photo than it is to enter a text query. It's interesting to compare this to the answers in the questionnaires where the text-based search was rated as a little bit faster. There could be an issue of what we're measuring here, is it the time it takes to enter the query or is it the time it takes to find the sought information?

A text-based search engine let the user add and exclude words for the query; this could however not be done with the prototypes. Under the cluster "General attitudes and feelings" most of the respondents were positive, but comparing it with text-based search engines, they found drawbacks with what the prototypes could achieve.

Respondent 10 changed search strategy to text-based query search when searching for information regarding the CD. In the beginning of the test the respondent changed the browsers location to http://www.google.com and typed an ordinary text query. However, the respondent didn't change strategy to text-based when the searching for information about the unmarked,

unknown object. This goes well with known objects and marked objects, when one knows information about the object to use in a text-query or when one can find out by examining the object.

Most of the respondents think that the possibility to combine text with photos, adding words in the search query to filter the results, would be a better way of searching,. A way is to filter the query instead of using Search modes as a filter. For instance, one can combine two photos, or combine a photo with text, or maybe a photo with sound recordings.

Respondent 4 mentioned that if one doesn't get any hits when doing photo-based searches there is really no way to modify the search query. Another respondent suggested that a possible solution would be that the system should had a slide bar so the user could decide how precise the photo match should be, a precise match would return fewer hits. This issue was also discovered in round one of prototyping.

Search with text queries and photo queries

One respondent was concerned that the matched images will not always return hits, and therefore it is not enough just using image-matching to get a sufficient amount of hits. This concern may lie in the relative small result set the prototypes presented. We didn't know how many hits a large result set contains but obviously this indicates that the prototypes had a relatively small amount. However, as theories spoke of the unnecessity of returning millions of hits when just one is enough as long as it contains the sought information, the users might feel uncertain with small amount of hits. Maybe the user gets a lack of confidence because they are used to Google providing millions of hits. The knowledge that there are more pages to evaluate maybe makes users more confident to find information they need. If this is so, an interesting way of finding out is to compare different prototypes that simulates different amount of hits, and evaluate when the users find them good enough. This may not however be the case at all. Maybe respondents just want more control, being able to choose among a larger amount of Web pages. One suggestion that the respondent come up with was that the user should be able to control and change precision of the matching algorithm. If the search result is not sufficient, the user lowers the precision so it will for example always present at least 300 hits.

Algorithms and accuracy

Respondent 1 commented that the prototypes search didn't contain as many hits as text-based searches usually do. With a real working prototype, this might not be true as it depends on how

accurate the algorithms are. Respondent 4 suggested a slide bar that controls the accuracy fo the algorithm.

Respondent 1 and 9 suggest a ranking based on how well the photo matches with the images. One way is to set accuracy ranking on images. Better matched color, texture, and so on would give the image a ranking number, one example is 893, 8 of 10 for color matching, 9 of 10 in texture, and 3 of 10 in brightness, and then rank the image. This might be interesting for future research in the indexing field. Should the ranking based algorithm about "most hyperlinks to pages" be the dominant algorithm? Suppose that best matched image will give the best results and not the Web pages which are the highest ranked.

Image, Text, and a combination of these

Respondent 4 points out that many Web pages contain information the users need but most often doesn't contain any images and this is a good point One example is Wikipedia (http://www.wikipedia.org), which holds a lot of information users could want to retrieve but not many photos. One problem could be if a user relies on the system to 100% and the result contains only bad hits. This could lead the user to believe that there is no information on the Web at all that satisfies her information need.

Respondent 6 assumes that where there is an image there is always related text, however this is not always right. Image banks don't have much text related to images other than meta information, often just a word.

The respondents were all aware of how to search the Web, and what Web pages that can be found. Some say that there are not always related images to information; others claim that there is always related information to an image. The first is more crucial, if there is information related to the photo but not a matched image on the webpage, should that Web page be excluded? Or should the system find keywords in other Web pages that have a high accuracy, and reveal these hits without images then searching on these keywords as well? This could lessen the ability for users to confirm the information on that Web page because there is no image. But it also returns a higher set, which might result in some cases with the information need that otherwise wouldn't be presented. A suggestion for a future study is to try to determine if the need of an image as an information element is crucial for the user to accept the information that is presented, or if not presenting such information, if the user is aware that there might be more information related but with no images.

Mode

Respondent 8 points out that it is advantageous with Search modes even though it is more decision making when to choose a hyperlink scattered over many steps instead of making all decisions at the result page. The respondent assumes that the use of Search modes results in less workload. Further, it seems that the user made few decisions when using Search modes, maybe because the respondent doesn't evaluate all hits, just the first pages of results. Future research could be done to evaluate how many steps are appropriate, if there is a trade-off between perfect filtering using many steps and the workload of evaluating hits of a large unfiltered set?

Mode requests

Respondent 7 requested a language mode. One can realize the demand for this as a lot of the Web's content is in different languages. Would the systems be localized based on languages, or have a functionality of choosing languages? Google is using a language and a country filter that filters the results based on the language the text is written in, or the country where the computer is connected to the Internet. Would a language mode be a necessary step as a mode, or should the system always provide the possibility to filter more as Google do? This could be done automatically if there are personalized preferences in the system.

Respondent 4 indicates that even a more precise Shopping mode could be appropriate, that just returns hits of prices, but this would need even more steps of mode options to choose from.

A mode request could also be to not use any mode at all, this feature is already working in the prototype in a way. Selecting the All mode takes you to a non-mode result, but still you need to pass through one more step compared to a strict non-mode search.

Verifying

Most of the respondents used the thumbnail of the matched image as a way to verify the hits of the Web pages. One respondent didn't seem to reflect upon the fact that the photo was just taken seconds ago. Instead, the respondent focus seemed to be on the information to find. Did this respondent memorize the photo or was it not important anymore when the results were given? Because the prototypes did not present the taken photo more than just at the photo moment, the respondents had to keep the image structure, color and so on in their minds. Text-based search engines present the query that the results are based on in the result pages (see background chapter: How search engines work). This memory load should not be hard to minimize, the question is more how to do it. As many text-based search engines, one could show the query photo in the result pages. A suggestion for future prototypes is to keep the taken photo in the

background in a semi-transparent fashion. However this might lessen the comparability because characteristics like colors change when changing transparency.

Image features

It seemed that Respondent 3 thought that the scale of objects was a problem. The respondent estimated the size of the physical object (Iittala), and then compared with the surrounding text of the matched image that specified sizes. How far away should the object be when taking a photo? Should a future system analyze distances in a photo from camera lens to the object, and as a result get information of the actual size? Or should the user estimate this in some way? Another issue is about authentic objects. Is it a model of an object or is it the real object? Does the user want information of models of an object, or the real object? The same issue comes when one takes a photo of a painting. Does the user want a match, based on an object in the painting, or the painting as the object? Respondent 4 identifies a problem on how to choose an object of many in a photo. One could study if users consider the centered objects to be the object to search on. Possible solutions are that one could draw a circle around the object in the photo with a pen (if the screen is a touch screen), or use a movable circle with adjustable size. A solution to the painting issue is to provide an Art mode where the result list is based on the painting and not an object in the painting.

Mental Modality

It seemed that Respondent 10 changed mental modality from image to text when the respondent found an indication of the name in the result list of the unknown products. This indicates that snippet can be useful when one doesn't know the object by name initially, when taking a photo. Further, it indicates that if a user's primary information need is the price and not the name, the name can be seen as part of the information users need because it is information that may take the user to the price. In this case the respondent could evaluate by triangulate, using the name (text) and the thumbnails (image), instead of just evaluate the thumbnails.

Respondent 1 switched mental modality totally, from image to text, and didn't care about images in the thumbnails. Can this be due to the respondent being used to no images in a result list, and consequently only cares about the traditional information elements. An interesting question is when does a user change their mental modality during the search process? If the purpose is to retrieve information from text, at some stage they will change mental modality from image to text. It would be interesting to study when people change modality and to what modality. We suggest this area for future research.

One respondent looks at the domain address, and makes decisions based on this. This indicates that domain addresses are helping users to evaluate the hits.

Objects

One disadvantage that respondent 1 and 4 identifies is that to be able to search, one needs to stand in front of the object and take a photo. This is an argument that the system is just a complement of searching, not meant to replace other search strategies like the use of text-based search. Things that one can't see like atoms, air and feelings are hard to photograph. Further, context where one wants information of an object but don't have it in sight, the systems usefulness is decreased.

Prerequisite knowledge

Respondent 8 indicate that a pre-knowledge about sites is advantageous when searching. It seems that respondent 8 looks at the domain address when looking for "OK computer". This indicates that the respondent is looking for a Web page that the respondent is familiar with. Lazonder et al. (2000) mentions that advanced users use a more analytical approach when browsing, but that in the end the results are the same as novice users. Applying that theory on this study, and if respondent 8 is an expert user (the respondent looked at the domain addresses) that should result in no differences between the respondents. However, respondent 8 was the slowest of them all with 8 minutes 12 seconds and the average time was 5 minutes and 31 seconds. The respondents that used less amount of time evaluating the domain address and more time on the images found the information they needed faster. The results are that images in the result list are of great importance. But for users that want to use their prerequisite knowledge about Web pages, the study indicates that showing the domain is helpful in some extent.

Areas of use

Respondent 6 suggested that one doesn't need to search immediately when taking a photo and instead search for information later on. This makes the time span more flexible. If one is stressed at the time taking the photo, one can search later on in a relaxed context. Maybe even transfer the image and search using a desktop computer with its advantages. This strategy also applies that one could take photos with a stand alone digital camera and then search with a desktop computer.

What's on the Internet

What kind of images there are on the Web is nothing one can do much about, neither the amount of images. One respondent (4) suggest another mode that is of persons. Further, the respondent

wants to find hits that are up to date (see background about Filtering). A proposal for a future study is to evaluate what kind of information user wants to know about, a study that is more thoroughly than the Initial phase of this thesis. A hypothesis could be that date control might be appropriate in some cases, but the adhering texts to the images are of more importance, this because that's most often the information the user needs. To meet the respondent's wishes, a way of filtering the dates could be appropriate to get more accurate hits (see theory chapter about Information context).

As the theories about that users in a specific domain finds information easier might also be true with photo search. Two of the respondents mentioned that some knowledge is advantageous such as what kind of information Web pages contains. A webpage that is just designed for selling CDs might also present information of artists. Therefore, it is an advantage if the user is familiar with what Web pages present even if their main business is not providing general information.

Evaluation of prototypes

The prototypes contained some issues that we mentioned in the chapter "Problems for the study". One issue was that the filename was shown in the address bar showing the name of the object. However, the observations imply that none of the respondents indicated that they were aware of this issue.

Evaluation of method

The approach to get data for the second round has been bit different from the methods in round one. Neither of them was bad, but to get more data after the first round we had to change the method. We found that the technique of interviews combinined with video observation gave us the most input. We used the video observation not only to observe the respondents, but also to record the interviews. This leaves us with a lot of different data that we can crosscheck and triangulate and find possibly interesting links between.

Because there is no photo-based search system in the real world, commercial or public, this was the first time the users encountered such a system. The use of visual search tools means that we were asking respondents to search in ways that were quite different from what they normally do. As McDonald et al. (2001) noted, this clearly have some affect on people's search strategies and success, and therefore must be considered.

The new device used where very rewarding for the study. With a higher resolution, larger display and a real Internet connection, the respondents felt that this was for real, and not a mock up that may impose a feeling of science fiction.

Time measurements, as a well-established method of defining usability has to be confronted with the users mentally perceived time. Most users doesn't measure time when deciding if they are satisfied or not, and if they do, what time is satisfying for a task? The time that users perceive mentally might be of more importance. By giving respondents the choice of decision-making, they seemed to estimate the time as shorter. Respondents that didn't use the Search mode prototype found information faster than the other group, but they believed the time was not satisfying when answering the questionnaires. The active role a user has seems to affect this. It appears that the less involvement of a user, the more time consuming the perceived time is.

The mode prototype has some advantages that make a time comparison unfair. The implementation of modes does implement the concept of a directory and this gives users possibilities of comparing similar results to each other that a non-mode prototype can't. Perhaps in some situations the users information need can be satisfied just by comparing hits in the result lists. A similar situation appeared in the IDeixis study (Tollmar et al., 2004). Time measurement does not consider similar variations.

Issues found

This part of the study showed as McDonald and Tait (2003) found, that the performance of the respondents was dependent on the search type. If they were given a known object, the need of using the prototypes as their search strategy tool is not definite. Instead, a search strategy could be a search using a text-based query search as one respondent did. Or it could be a phone call or to ask someone nearby. In this part of the study we first introduced the respondent to the prototypes and then asked them to do a search. As we didn't tell the users that they were free to find the information in any way, and that one respondent used Google text-based search, this is quite interesting. As McDonald et al., (2001) puts it, in some cases it is more effective and faster to use a text-based search engine. However, with the unknown object, the respondents' strategy to find the sought information was to use the prototype. This indicates that photo search is more valuable when the search type is an unknown object. Future stydies could examine in what contexts and with what classes of objects people prefer a certain strategy after introducing different strategies of getting information.

Some of the respondents estimated the tray objects size, and then compared with the printed size information on websites. One issue when taking a photo is scale. Let's say that you are in Paris and take a photo of the Eiffel tower, and then you get a lot of matched images depicting souvenirs modelled from the tower. A suggestion for future work is to examine if and how a system can measure the size of an object when taking the photo? Another issue that was revealed is the necessity to define the color of an object. Let's say that the user wants a tray in another color, but the system retrieves only trays of the same color as the one in the photo.

It has been confirmed in previous studies (McDonald et al., 2001) that users' search needs may be less precise or even vague. Users may be looking for an image to illustrate a general theme, not exactly the object in front of them. They point that some evidence suggests that the nature of the task can affect search tool choice and performance. In this thesis the respondents confirm this theory; they suggested, not knowing exactly how the images was retrieved, that in some cases a less detailed photo would be better to search with. During the construction of the prototype we thought of a slide bar that could iconize the photo to retrieve a more conceptual image. This could be three buttons as well, one at a detailed level, one iconized and one in between. We suggest future research to study this subject. Further, to study the possibility to let the user disconnect the color algorithm in the system and only go for grayscale, or let the user define priorities between shape, texture, color, and size.

Will the user be presented a sufficient number of hits, or does a photo matching system need to be complemented with Web pages that doesn't contain any images but information of the photo? A suggestion for a future study is to determine if the need of an image as an information element in a result list is vital for the user to accept the information that is presented, or if not presenting such information, if the user is aware that there might be more information but with no images. Probably there is not the same amount of images as there are words on the Web. The amount of matched hits could therefore be much lower when searching with photos than when searching with words.

If a Web page witch contains the sought information is not high ranked, a user has to browse several result pages before finding it. On the other hand, if the Web page with the sought information is high ranked, it seems like Search modes is not as vital. An assumption is that types of Search modes where Web pages probably have a high ranking like Shopping mode, is not as rewarding as for Search modes where the sought information only can be found in less ranked Web pages like personal Web pages.

As respondents had some difficulties in navigating (scrolling horizontally and vertically) inside a Web page, a suggestion that may ease the navigation inside a Web page is to let the browser automatically focus on the part of the Web page where the matching image is located. A proposal for future research is to see if this actually helps users to find needed information faster or if it just confuses.

Respondents commented in negative aspects that there were several hyperlinks to the same Web page in the result list. Perhaps handling of replications of hits is of greater importance on a mobile device where space is valuable and can be used to present another unique hit instead.

Conclusion

We found that photo search in such is very satisfying and enjoyable to use. We found that Search modes seems to be more rewarding when returning a large result set than the system not using Search modes. The results indicate that using a photo search on unknown unmarked objects is more rewarding than using a text-based search. With marked objects, it seems that users prefer text-based searches to some extent. In conclusion, the Second round of prototyping revealed some issues that weren't found in the Initial phase or in the First round of prototyping and gave greater insight into others.

Discussing the study

Comparing the Initial phase with the First and Second rounds of prototyping, we found that if respondents don't have a prototype at hand, their imagination about what can be done with the system is wider and wilder. Even if that met the purpose of the Initial phase, it surprised us that users in the two following studies where they actually were testing a prototype were less imaginable. It seems that because they had a prototype in front of them, they did not let their imagination take them very far away from it. This can be due to that in the Initial phase respondents were asked to come up with any ideas, whereas for the other two studies we never expressed that. Further, they didn't have that time to fully imagine what the system could do in the prototyping rounds. We asked respondents face to face and waited for a direct response, whereas the respondents in the Initial phase had 24 hours to come up with innovative ideas.

With the selection of respondents in this study, it is precarious to generalize the results to a whole population, even to generalize to other similar groups. E.g. it's possible that a similar selection of Japanese students would generate results not at all similar to the ones we found here.

Another issue with the results is that the respondents didn't take the photo themselves, we just encouraged them to pretend to take the photo. This maybe influenced them in ways that affected the results but this we don't know. Maybe they didn't get that "feel" for the idea of photo-search we we're trying to achieve. From the Initial phase, the photos of the objects differed very much from each other. Questions that arose were how users would take photos if they were given the chance? From what angle? How much time would they spend taking the photo? This has not been accounted for in our test sessions.

Conclusions and future research

The study was setup using three sequential parts, Initial phase, First roud of prototyping, and Second round of prototyping. They all had differencies in methods and purposes. But together they all strived to the purpose of the thesis that is to explore the area of general mobile photo-based Web searches to find what issues and phenomena it contains. As we carried out three studies, where the later is build on the previous we had to choose a path that was based on the results from the previous study, due to the characteristics of the whole study as such. Less significant issues and discussions about the issues can be found under each parts of the study chapter. Here we present an overall view of the results.

The results of the study indicate that respondents are positive to the new way of searching using photos. Further, the results of the study indicates that in some contexts when one doesn't know what one is standing in front of, the photo matching system is advantageous compared to a text-based system. With marked objects photo-based search is not always the best way, there are several alternative strategies to find the needed information. If one is used to a certain search strategy and method, it may be easier to stick with that. It would be interesting to study what strategy people use to find information of unmarked objects today, and which is the most frequent. Such a study could also tell if it really is necessary to provide another way of searching like photo matching. However, 10-15 years ago people didn't ask for text-based search engines, today it is one of the most popular ways of finding information. As we've noted earlier, we believe the existence of tools can change search behaviour.

The study indicates that there is no consensus of what users think is the most important information element in a search result list. Is it via the thumbnail of matched image, or is it through information elements with the characteristics that the user can evaluate the reliability of Web pages? We found that it might be a combination of these. That is, we can't say that just Search modes help users to find what they are looking for. Matching image algorithms and automatic filtering may also play its role, but we are not sure at what level. The results indicate that the matching images are the most valuable information element when evaluating the image-matching algorithm, URLs are valuable when evaluating Web pages, and it seems that snippets are valuable to find the sougth information directly or when evaluating the content of Web pages. By having one element of each such evaluation type, we think users can get a higher evaluation precision. As to lessen the users memory load, text-based search engines present the text-query that the results are based on, even in the result pages. By presenting this, users can directly compare the results with the typed text query. Even if it may be easier to remember a photo than text, this could be appropriate to do with photo searches as well. Presenting the taken photo in the

background or somewhere else in the result page could be one way of keeping the search input as a reference and feedback to the user.

A subject for future studies is to test giving users the possibility to filter results of a photo search by tweaking parameters. If the user could specify what features of an image she thinks is important, the result set might be better suited to the users information need. For example, the user could be very interested in the color and texture of the object in the photo but not the shape, or perhaps the shape is very important, but not the color.

A standpoint of the study was to study a way of providing better information that allows the user to evaluate the results more quickly, not how to return higher quality search lists like index coverage and ranking algorithms. We found contradictory to Woodruff et al. (2001) that these two in some extent are inseparable when it comes to photo search. An exclusion of the algorithms from our study could not be made as the results from the Initial phase led us to study a way of helping the user through a search process and to present high quality result list, a way to filter a result list. As respondents revealed certain issues, we had to involve aspects of more technical features like algorithms. One example of this is the suggestion to let users tweak parameters controlling the precision of the image algorithm.

As today, text-based search engines use booleans to narrow or widen search results and it could be an interesting subject for future research to see how this could be applied to photo searches. The recipe search from the Initial phase is an example of where a user could take several photos and combine them to one search using booleans.

One of the main issues revealed in the Initial phase was the issue that users wanted each hit to be caterogized in the result list. The outcome of this was the idea of Search modes, a way of combining the photo search with some kind of categories to help a user filter a search result. A general photo-based search system has difficulty of knowing what kind of information in need the users want to satisfy with each photo. The system can only return as good hits as possible based on the photo and with Search modes we were trying to overcome this issue. An interesting idea would be to make creation of Search modes available to the users of the search engine. Users could design their own Search modes using configurable building blocks of filters, rules and presentation templates, and then publish their Search modes for others to use in their own mobile photo-based searches. The results in the study show that Search modes are considered useful, but they may need to be configurable to personal needs.

Our contribution to the research area of image-based information retrieval is that to be able to sort the amount of Web pages that contains matched images it is not enough to increase the precision of image matching algorithms to satisfy users' different information needs. Users need additional ways of specifying their information need. Specialized photo-search systems do not need this as the information need is specified already in the choice of service as in the IDeixis project that is an example of this. However, a general photo-based Web search is different and from reactions among our respondents, Search modes or another form of categories, automatic or manual, seems to be an interesting starting point when trying to solve the issue of specifying information need with limited input devices.

References

Aslandogan, Y. A. & Yu, C. T. (2000). Evaluating Strategies and Systems for Content-Based Indexing of Person Images on the Web. In *Proceedings of the eighth ACM international conference on Multimedia* (pp. 313–321). Marina del Rey, CA.

Beale, R., & Lonsdale, L. (2004). Mobile Context Aware Systems: The intelligence to Support Tasks and Effectively Utilise Resources. In *Proceedings of the 6th International Symposium on Mobile HCI* (pp. 240-251). Glasgow, UK.

Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic Web. Scientific American 284, 5, 34–43.

Biever, C. (2006, January, 7). See it, snap it, and let your phone find it on the web. *New Scientist*, p. 20.

Bosswell, W. (n.d.). What is the most used search engine? (online). 2005, May 20. URL http://websearch.about.com/od/faqs/f/most_pop_search.htm.

Bryman, A. (1995). Quantity in Social Research. London: Routledge.

Burton, M., Pupons-Wickham, D., Phelps, L., Spain, K., Crews, J., & Rich, N. (1999). Secondary navigation in software wizards. In *Proceedings of CHI '99 extended abstracts on Human factors in computing systems* (pp. 294-295). Pittsburgh, Pennsylvania.

Campbell, C. & Tarasewich, T. (2004). What can you do with three pixels? In *Proceedings of the 6th International Symposium on Mobile HCI* (pp. 1-12). Glasgow, UK.

Cohen, L. (2005, February). Boolean Searching on the Internet. (online), 2005, May 20. URL: http://library.albany.edu/internet/boolean.html.

Delisle, J. (2004, July 27). Google quick help. (online), 2005, May 21. URL http://faculty.valencia.cc.fl.us/infolit/Google/help.htm.

Dictionary.com (n.d.). Search engine, (online). 2005, May 20. URL http://dictionary.reference.com/search?q=search%20engine.

Dryer, D. C. (1997). Wizards, guides, and beyond: rational and empirical methods for selecting optional intelligent user interface agents. In *Proceedings of Intelligent User Interfaces 97* (pp. 265-268). Orlando, FA.

Etemad, E., Newth, J. D. (2004). Pocket-Sized Design: Taking Your Website to the Small Screen. Alistapart issue 187.

Glover, E. J., Lawrence, S., Gordon, M. D., Birmingham, W. P. & Giles, C. L. (2001). Web search– your way. Communications of the ACM, (pp. 97-102). New York, NY.

Google.com (2005). Search Results Page, (online). 2005, May 12. URL http://www.google.com/intl/en/help/interpret.html.

Gudivada, V. N. & Raghavan, V. V, (1995). Content-based image retrieval systems. IEEE Computer, 28 (9), 18-22.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. Interacting With Computers, 12(4), 409–426.

Jones, M., Buchanan, G. Marsden, G. & Thimbleby, H. (2000). User Interfaces for Mobile Web Devices. Position paper for W9 Mobile Workshop [WWW document]. URL http://www9.org/w2-mobileweb/web9pos.PDF/.

Jones, M., Buchanan, G. & Thimbleby, H, (2002). Sorting Out Searching on Small Screen Devices. In *Proceedings of the 4th International Symposium on Mobile HCI* (pp.81-94). London, UK.

Kaikkonen, A. & Roto, V. (2003). Navigating in a Mobile XHTML Application. In *Proceedings of CHI 2003* (pp. 329-336). Lauderdale, FL.

Kjeldskov, J., Skov, M. B., Als, B. S. & Høegh, R. T. (2004). Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context Aware Mobile Systems in the Field. In *Proceedings of the 6th International Symposium on Mobile HCI* (pp.61-73). Glasgow, UK.

Knowthis (2005). Leading Search Engines and Web Search Tools. (online). 2005, May 17. URL http://www.knowthis.com/research/searchtools.htm

Koskela, T. & Vipola, I. (2004). Usability of MobiVR Concept: Towards Large Virtual Touch Screen for Mobile Devices. In *Proceedings of the 6th International Symposium on Mobile HCI* (pp. 252 – 263). Glasgow, UK.

Lazonder, A. W., Biemans, H. J. A. & Wopereis, I. G. J. H. (2000). Differences Between Novice and Experienced Users In Searching Information on the World Wide Web, Journal of the American Society for Information Science, 51(6), 576–581.

Lempel, R. & Soffer, A. (2001). Picashow: Pictorial authority search by hyperlinks on the Web. In *Proceedings of the 10th Int. World Wide, Web Conf* (pp. 438-448). Hong Kong, China.

Loeb, S. (1992). Architecting Personalized Delivery of Multimedia Information, Communications of the ACM, 35(12), pp, 39-48.

Maybury, M. T. & Wahlster, W. (1998). Introduction. In Readings in Intelligent User Interfaces. In *Proceedings of the 4th international conference on Intelligent user interfaces* (pp. 3-4). Los Angeles, CA.

McDonald, S., Lai, T. S. & Tait, J. (2001). Evaluating a Content Based Image Retrieval System. In *Proceedings of SIGIR'01* (pp. 232-240). New Orleans, LA.

McDonald, S. & Tait, J. (2003). Search strategies in content-based image retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 80–87). Toronto, Canada.

Miles, M. B. & Huberman, A. M. (1994). Qualitative Data Analysis. New Bury, CA: Sage.

Mizzaro, S. (1997). Relevance: The whole history. Journal of the American Society for Information Science 48, 9, 810–832.

Notess, G. R. (2005). Strategy 2: Subject Directories for Organizations. (online), 2005, May 21. URL http://www.searchengineshowdown.com/strat/.

Paepcke, A., Chang, C., Garcia-Molina, H. & Winograd, T. (1998). Interoperability for Digital Libraries Worldwide. Communications of the ACM, 41(4), 33-43.

Paepcke, A., Garcia-Molina, H., Rodriquez, G. & Cho, J. (2000). Beyond Document Similarity: Understanding Value-Based Search and Browsing Technologies, SIGMOD Records, March 2000, 29 (1), 80-92.

Pazzani, M. & Billsus, D. (1997). Learning and revising user profiles: The identification of interesting Web sites. Machine Learning, 27, 313-331.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (2002). Human-Computer-Interaction. Harlow: Pearson Education Limited

Raisamo, R., (1999). Multimodal Human-Computer Interaction: a constructive and empirical study, (Ph.D. dissertation, University of Tampere), [online]. Available at http://wwwhome.cs.utwente.nl/.

Rogers, I. (n.d). The Google Pagerank Algorithm and How It Works (online), 2006, January 9. URL http://iprcom.com/papers/pagerank/.

Sagarmay, D. (2004). Multimedia Systems and Content-Based Image Retrieval. Arlington: Information Resources Press.

Schofield, E. J. (2003). Language-Models for Questions, In Proceedings of the 10th Conference EACL Workshop on Language Modeling for Text Entry Methods. Budapest, Hungary.

Sullivan, D. (2002, October 14). How Search Engines Work. (online), 2005, May 20. URL http://searchenginewatch.com/webmasters/article.php/2168031.

Sweeney, S. & Crestani, F. (2003). Supporting Searching on Small Screen Devices using Summarisation. In *Mobile and Ubiquitous Information Access: Mobile HCI 2003 International Workshop* (pp. 187-201). Udine, Italy.

Tollmar, K., Yeh, Y. & Darrell, T. (2004). IDeixis – Searching the Web with Mobile Images for Location-Based Information. In *Proceedings of the 6th International Symposium on Mobile HCI* (pp. 288-299). Glasgow, UK.

Web pros group. (n.d.). Types of Search Engines & Opportunities (online), 2005, May 20. URL http://www.webprosgroup.com/search-engine-types.htm.

Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J. & Pirolli, P. (2001). Using Thumbnails to Search the Web. In *Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems* (pp. 198-205). Seattle, Wash.

Yin, R. K. (2002). Case Study Research: Design and Methods. Thousand Oaks, CA: Sage.

Zhao, R. & Grosky, W. I. (2001). Bridging the semantic gap in image retrieval. Distributed Multimedia Databases: Techniques and Applications (pp. 14-36). Hershey, PA.

Appendices

Appendix 1: Interview guide (Initial phase)

What did you think of when you took the photo?

What is the question you asked yourself to get information of?

Which of the information elements below do you think would help you evaluating the hyperlinks to find the needed information?

Would different types of information elements help you for the five photos?

	Photo 1	Photo 2	Photo 3	Photo 4	Photo 5
URL					
Text size (KiloBytes)					
Image size (KiloBytes)					
Last updated					
The owner of the website.					
Domain name					
Summary					
The Web page purpose					
An image of an image on a web page					
An image of the web page					
Category					
Statistics- amount of hits					
Other, what?					

Appendix 2: Results (Initial phase)

Respondent One

Photo taken	**Questions asked**	**Information wanted**	**Comments**
Bottle of Ketchup	What is the price in other nearby stores?	Price, name of the stores, distance	One doesn't know all the prices.
Book	Information about the book?	Abstract, resumes, links to other similar books. Where to buy	Want a direct link to http://www.amazon.com. Price doesn't matter!
Different food products.	Give me some recipes? What can I cook?	Recipes, cuisines	
A street corner	Where am I?	Map	Sites with maps that the respondent knew were: http://www.mapquest.com and http://www.kraks.dk
The registration number of a car	Who owns this car?	Telephone number, address	Wanted the phone to ring up the owner.

	Ketchup	**Book**	**Food**	**Street corner**	**Car**
URL	X	X	X	X	
Text size (KiloBytes)	Difficult to know				
Image size (KiloBytes)					
Last updated	X	-			
The owner of the website.		-	Depends		
Domain name		-			
Summary	Short	-			X
The Web page purpose				X	
An image of an image on a web page		X	X		X
An image of the web page	Very useful	X	X		X
Category	Color symbols	X	X		X
Statistics- amount of hits	Fun	X	X		X
Other, what?					

Comments

URL is always good

If the Web page is Nestlé, they probably only list their own products.

Test on Google

Typed Ketchup prices, + Helsingör, Heinz. Couldn't find it

Typed name of author. Found on first hit

Typed Onion Potatoes Paprika salmon chilli. Found hits on restaurants.

Typed Stengade Fiolgade. Couldn't get hit on http://www.mapquest.com.

Typed registration number. Couldn't find it.

Respondent Two

Photo taken	**Questions asked**	**Information wanted**	**Comments**
His own hair	Where and to whom can I sell hair?	Address, Telephone number.	Curiosity!
A logo of a football team seen on television	Want information about the team? Where are they from?	Country, city	
Picture of the Swedish King and Queen.	When did they marry?	A year	
A trashcan contain a bag of Burger King	Where is the closest Burger King?	Distance, location like Address.	Wanted a map with a distance arrow.
Ashtray	Who is the designer of this ashtray? When did the person design it?	Name, year	

	Hair	**Brand**	**King &Queen**	**Burger King**	**Ashtray**
URL		X		X	X
Text size (KiloBytes)			X		
Image size (KiloBytes)					
Last updated					
The owner of the website.					
Domain name		X			
Summary	X	X	X		X
The Web page purpose	X				
An image of an image on a web page				X	
An image of the web page				X	X
Category	X				
Statistics- amount of hits					
Other, what?					

Comments

No comments

Test on Google

Tried with many keywords. Couldn't find where to sell

Typed Kaizer chiefs soccer. Result: South Africa (ten hits down)

Typed Karl XVI Gustav+Silvia. Result- 1976 (snippet)

Typed Burger King. Result- a webpage with a map, no distance.

Typed Kulan Designer, + ashtray, Kulan design, Kulan+Gustavsberg. Result-Gunnar Larsson Year not found.

Respondent Three

Photo taken	**Questions asked**	**Information wanted**	**Comments**
A Bush	What is the name of it? How will it look during wintertime? When and how long is the Blossom period? Where can I buy it? How much?	Name, Photo, A period of days, address, price.	Lots of questions!
A Car	What is the brand of the car? What is the name of the color? Where can I buy it?	Name of car and of color. An address.	Wanted to buy a used car, his suggestion pointed at http://www.blocket.se
Konsum Supermarket at Mårtenstorget (Lund, Sweden)	Open Hours? What other stores have open? "This, if the Supermarket is closed "	Open Hours, names and distance, a map	Suggest a rating system like http://www.download.com
The Mårtenstorget (a square in Lund)	What is the day of the market? What else happens?	Dates, hours and Information of special happenings	Would like to know what happens on other squares.
A Cinema	What films are they showing?	Titles of films, abstracts, time	Of course, respondent wanted to be able to book.

	A Bush	**A Car**	**Supermarket**	**Mårtenstorget**	**A Cinema**
URL		X	X	X	
Text size (KiloBytes)					
Image size (KiloBytes)					
Last updated			X	X	
The owner of the website.					X
Domain name					
Summary	X	X	X	X	
The Web page purpose	X				
An image of an image on a web page	X				
An image of the web page	X				
Category	X	X	X	X	
Statistics- amount of hits					
Other, what?					

Comments

Hard to generalize if a title of the page is needed. It depends on how well the search results are categorized.

A rating system is a good way of deciding the quality.

If it is a company I trust I never check when the page is last updated. I trust that the web page is up to date.

Test on Google

Did not complete test

Appendix 3: Interview guide (Round one)

After the first search ("Kulan") not using Search modes (Taking photo > Search result > Webpage)

- How does searching by taking a photo and using it as input to the search feel?

After the second search ("Myran") using Search modes (Taking photo > Select mode > Mode settings > Search result > Web page)

- What was your thought on being presented a choice of Search modes?
- What did you think of going through several steps to come to the search result list?
- Did you feel that the result became more precise when you used the Search modes?
- Can you think of other types of Search modes that could be handy with mobile searches?
- How did you make the decision of which hyperlink to follow in the result list?
- Are there any information elements in the result list that you find unnecessary?
- Is there any information element that you think is missing in the result list that would allow you to evaluate the hyperlinks better?
- Was it worth the extra workload to get a more specific result page?

Appendix 4: Result matrix (Round one)

Question	**Female age 26**	**Male age 25**	**Male age 25**	**Male age 24**	**Summary and comments**
1. How does searching by taking a photo and using it as input to the search feel? Compare with text-based search.	Good if you don't know the name of what you're looking for, if you know the name the respondent probably rather searches by words if not the image search is faster. If you look for other images, image search is natural.	Feels strange. Never thought of this before. Doesn't know any practical use of this but understands that you can get info on anything you see.	1. Very good, in this ideal scenario at least. Good when you don't know the wording for a search. Wants to do text-based search but if one doesn't know the object respondent wants to do image search. Worried about copies, how can one be sure that the image search doesn't find cheap copies from IKEA? The image in the result is very important as it verifies the result. Respondent did find all information in the snippet and this is his normal search strategy on Google. If respondent wants the full story respondent follows hyperlinks, but facts can often be found in the snippet if the search is good enough.	1. Hates writing things, likes photo, just click a button.	Good if you don't know the name of the object, when you don't know wording for a search. Feels strange. Two of them rather want to search with keywords.
2. What was your thought on being presented a choice of Search modes?	Hard to choose. But if Shopping doesn't work the respondent planned to go back and use General.		3. Good, plan was to go into General first to read and then Shopping mode for prices.	3. Shopping – intressant. Hopes to finding price in Shopping mode.	Good, but hard to choose. Why is it so? "Combine?" "Caching searchlist?" Misunderstanding word "General"

3. What did you think of going through several steps to come to the search result list?	Even more steps would have been hard. Automatic sensing of location would be desirable.	Makes the search more precise, good if it works. Respondent was a bit disappointed when ending up at a Norwegian site when looking in Sweden.	4-5. good in this case, not sure about others. If the Search modes are good the system will be usefull. Expects General mode to exclude the hits from other Search modes like Shopping. So, no webshops in General.	4. Very good. Many small steps better than all on one page. Likes the possibility of narrowing the search.	No comments about where you are in the search process eg 2/5. Positive if they really narrow the search.
4. Did you feel that the result became more precise when you used the Search modes?	Yes. Old hyperlinks not good though	Likes the idea but the prototype did not make the search result specific enough for him.		5. Absolutely. Hard to tell if the hyperlinks really were from lund though.	Prototype not specific enough. Strengthens the positive result of question 4.
5. Was it worth the extra work load to get a more specific result list?	Few steps with few options makes it ok.				
6. Can you think of other types of Search modes that could be handy with mobile searches?	Tourist info would be good. General mode too general. Shopping mode hard, as different products have different shopping habits.	Education mode? Used products mode.	8. Geographic mode (Paris Hilton issue: are you searching for the hotel or the person?) wants auto context sense. Keyword search (like IDeixis) could be good sometimes.	8. Review mode – what others think about e.g. restaurant. Search modes should not be too specific though, you don't want to exclude too much when choosing a mode.	

7. How did you make the decision of which hyperlink to follow in the result list?	Snippet is main. Pictures makes you exclude bad hits, like the book cover.	Title is irrelevant. Domain is important URL not. Respondent recognized some domains in the search like blocket.se. The snippet is the core, very important. Image works as a confirmation that the search worked.	10. The image important, confirms the search. If the search is good and you can trust it, the images are not needed. Snippet second important. Domain not URL.	10/11. Snippet is good. But different from search to search. In review domain would be important. Image confirms search. Titles can't be trusted. Looks at all info and does assessment. To find price you can enter a page. Rather follows a hyperlink then having to scroll a lot.	Image is confirmation 4/4. Snippet most important 2/4. Snippet important 2/4. Domain is important not the URL 4/4. Title not important ¾. Title gives you a hint ¼. How will they respond if we exclude Title or give them a substitute?
8. Are there any information elements in the result list that you find unnecessary?	Title gives you a hint of what kind of page it is but can't be trusted, Snippet relevant, Image verifies search, Domain but not the complete URL is good.		11. Title not needed in this case. It did not draw his attention in this case.		
9. Is there any information element that you think is missing in the result list that would allow you to evaluate the hyperlinks better?		Date of last update and nr of images on the page. If too many, the connection is slow.	12. Size of page in bytes. Because of cost and speed. Probably less important in the future.		All answers consider connection speed, and are out of our jurisdiction. Date of last update interesting.

Observation 1	Noticed the number of hits Read the snippet of first hit Found the designer in the snippet of the second hyperlink Followed the hyperlink, but liked the snippet better. Found a place to shop in third hit Found the price in link as well Commented that the image confirms the link	Observed the nr of hits. Had some trouble reading on the screen. Some characters are corrupt. Goes through the complete list of search results Comment on different use of different sites, like this is a shopping site and so on. Got the answers from the first search hit	Saw the number of hits. First hit, confused, no image, doesn't know if it is the correct product. Second hit, with image, weird characters. This hit is related to the first (gustavsberg mentioned in both) so respondent assumes the first hit to be correct as well.	Saw 9 hits. Tries to enter first hit. Doesn't work. Tries second, broken as well. Changes strategy to just look at the searchpage. Then found all info in the three first hits. Commented that his strategy usually is to open many hyperlinks.	4/4 noticed number of hits. Sooner or later they all used the strategy to only look in the snippet for info. Disturbed of strange letters, 2 gave Comment of this. 2/4 followed hyperlinks, they didn't work so the respondents were disappointed. Confirmation through combination of two snippet where one searchresult didn't have an image. Image good, the users tend to use the image for confirmation. One didn't dare use a hyperlink that didn't have an image.

Observation 2	"If you choose Shopping mode you probably get all the answers" Chose Sweden because the respondent thought Lund to be too specific (they probably don't even sell it here), and Europe too large, to find this chair. Clicked a hyperlink, didn't like to scroll horizontally to see all text on the page. Hard to follow linebreaks. Comment on choosing the hyperlinks. Hard to tell which one is selected.	Knows the chair and the designer Goes into Shopping because respondent expects to find used products there and exclude info about chair. Expects General mode to include Shopping. Chooses Sweden because it's a good level of search and probably gives you Swedish prices Gets a price of 10000 kr and doesn't know if it's Danish or Swedish price. And respondent doesn't know if its one ore more chairs for that price. Ads in images are repulsive.	Choose General, expecting to find the general information asked for. It was too general, goes back and into Shopping – lund as respondent wanted to buy it in lund. Commented on that respondent was not "genuinely interested in the chair" and that it lowered his effort. Commented on the slow loading of pages.	Plays along when snapping the photo. Chooses Shopping – lund, close to pick up a chair. Says if electronics, respondent would choose Europe. Finds the designer in the search result. Finds the price in the list, 6000 for four chairs.	2/4 chose Lund, they wanted to shop close. 2/4 because they wanted good hits. Distortions, hard to tell which hyperlink selected. Ads in images are repulsive.

Comments from respondent 3

- Because of the scroll screen you are forced to look through all information. On google, respondent quickly scans the page without reading everything.
- Wants to add words manually to the image search.
- General mode should exclude the other Search modes. "Show all hits" should be a different thing.
- Idea about putting the number of hits at bottom of the page because you always look through the 10 first hits anyway. The decision to go to the next page of results is not made until at that point.
- If you get 10000 results you may want to take another picture to better specify the search.
- If less than 10 hits – no use choosing a mode.
- The mode selector page could specify how many hits each mode will give.
- Doesn't really believe in keyword extraction from pages. Has an idea about a manually created database of words and categories to search in instead.
- The domain-name tells more about the page than a thumbnail. Thumb probably too small in mobile screen. Not even on google, why on mobile. Good thing though to quickly find pages visited before.

Appendix 5: Interview guide (Round two).

0. Ask yourself a question about the object in front of you? (This is a start up question to conduct the test)

1. How does it feel to take a photo and search with that instead of searching with words?

2. Did you find it easy to find a Web page that contained the information you needed?

3. For what would you use a system like this?

4. What is it the limit of what questions that can be answered?

5. What do you think about combining photo search with key words?

6. Would you find the sought information easier using another way of searching?

Questions for B-group only (using Search modes)

7. What was your thought on being presented a choice of Search modes?

8. What do you think of going through several steps to come to the search result?

9. How did you find the relation between the alternatives ”All” and ”Shopping”?

Appendix 6: Clustered data (Round two)

We clustered data from the interviews and observations that we felt had something in common or were related in other ways. This is our translation of the original data.

To track answers and observations back to the video all data is marked in one of two ways:

5.2 Test Person 5, question 2

TP6 Observed behaviour or a spontaneous comment of Test Person (Respondent) 6

Each cluster has been assigned a letter. This is for tracking purposes and the ordering of clusters has no other meaning.

A- The mobile Device.

2.3 The device limited more than the actual system.

2.10 Not used to PDA, had some problems.

4.9 Photos can be insufficient. The camera resolution, sun light and so on.

4.6 Camera quality.

TP9 First thought it was no problems with scrolling, then thought it was really heavy. Bad designed web pages are more attentive

7.10 Camera quality does it affect the search?

TP 5 Wants to have a horizontal screen.

TP9 Wants to scroll by dragging the whole page.

1.1.1 Not used to search with a mobile device.

B- Go to.

TP8 Hard to locate the image, once inside a webpage.

2.4 Difficult to find on the web pages.

TP4 Wants to get to the image on the web page directly.

C- General Attitudes and feelings

2.8 It was easy to find.

1.5 Felt really cool

2.2 Good with unknown objects.

1.10 Good when searching unknown objects.

3.10 Good when one doesn't know the object.

1.8 Can search on unknown objects.

1.9 Useful, especially for unknown objects.

3.10 Difficult to see myself using it.

2.10 It was really easy.

2.7 Yes! It was easy!

2.4 Easy to find the web pages.

1.1 Ohh, even more unnecessary mobile functions.

1.9 There is so much shit on Google.

6.8 Photo search is unbeatable!

3.4 People will over use this system at the beginning. Everyone will take photos of everything.

2.6 Absolutely easy to find information.

1.10 Feels fun, new and its exciting but also strange and not used to it.

D- Prototype

TP3 Many links to the same web page.

1.3 Same web page several times in the search result.

1.5 The user didn't take the picture. Therefore couldn't say anything of user experience.

TP6 Thinks there is many links to the same web page.

TP8 Believe that there are many links to the same web page.

TP8 Hard to find the "change mode" button.

E - Search with text queries and photo queries

Text

4.6 On Google, one can change the query; this is not possible with photos.

4.3 With text-based search, it is possible to change and develop new queries; this is not possible with image search.

1.10 Before the study the TP said that it's hard to understand what to use the system for. After trying it out, the TP was much more confident.

TP10 Felt unease with the system, used a text-based query search in Google instead.

2.1 "One is used to search with words."

2.5 "One can specify more with text."

1.5 Thinks words are more useful.

1.2 Before the study, the TP thinks it's faster with text-based search, afterwards didn't know which was faster.

3.5 Thinks that words are more useful than photos.

4.2 Text-based search can filter through adding words.

Image

3.3 Some situations is faster with images.

TP1 Taking a photo is much faster than to find out what text queries to use.

3.8 Same context as the use of text queries.

1.9 It's not more difficult than Google.

2.9 Same as 1.9

1.6 Feels good, not a big difference from Google.

Other

3.2 If one knows more of the object one can search with text.

4.6 If one won't get any hits, one can't do anything about it. A negative aspect!

4.7 Image is just one search criteria.

1.7 Image is just one search criteria. Wants to filter the results more than Search modes do.

Text + image

5.10 Image + text or searching with more than one image.

5.2 Spontaneously, the person wants to combine image and text.

5.4 Same as 5.2

5.2 Combine images.

1.7 Word looks for words on the web pages, image looks for images on the page, much easier.

2.8 If one searched with text like "bowl", one might get results from wrong bowl.

F – Algorithms and accuracy

1.6 High expectations, to get better results.

1.9 Thought that the system would be more precise.

2.5 "As with text-based queries, one can not get a perfect search result"

1.9 Wants a perfect matching "imitations" of the photo, and then related below.

1.8 Can be difficult to get exact the same object.

1.3 Not that many hits as with text-based search.

TP9 Wants to have a ranking on the result pages. The best matched first.

TP4 Wants to define how accurate the image matching should be, with a sidebar. In one end exact matching and in the other end loose matching.

G- Image, Text, and a combination of these

4.3 Many pages don't have an image of the object, but has information in textual form.

1.6 Assumes that there is always text to an image, therefore if one finds the image one finds information of it in textual form.

1.6 "One really finds web pages with images, and not just boring text pages".

H- Mode

8.10 Its faster with fewer steps.

8.10 Search modes are unnecessary.

8.4 No problem. If one searches often one might be disturbed of the steps that one has to go through.

7.6 Was not ready, changed mode when respondent was not satisfied with the result list.

7.8 Assume that "All" is about more general subjects. Good that one can limit the search list.

8.6 No problem! Great with categories. Three steps is ok to go through, 10 steps would be too many. Four would be ok if one really gets appropriate hits.

7.4 Wanted to buy, therefore the choice "Shopping".

TP6 Felt comfortable that respondent would find related shopping Web pages under the mode "Shopping".

8.2 Wants to choose mode before, instead of after the search engine has done the search procedure.

7.2 Chose "All" because respondent didn't want to buy anything, but didn't know what "All" meant.

8.8 One prospers on it: Little bit more work load, and decision making but one gets better hits.

TP8 A tendency that the respondent wants to stay in one mode. "It takes long time if one enters wrong page and then change".

I- Mode requests

8.8 One should be able to choose to not using Search modes. Because it is more work load using Search modes.

TP7 The majority of the web pages are in English, but wanted Swedish. A mode about language would be great.

4.7 Wants to limit just to Sweden or price search.

J- Verification

1.2 When one searched, one didn't think that one used a photo.

TP6 Evaluated the matching images on the result page to see if they matched.

TP5 In the beginning, the respondent in general didn't look at the matched images.

TP4 Verifies with the matching image: thinks it's nice.

TP6 Thinks it can be difficult for the system to verify real and fake objects.

1.2 Verifies through the matching images.

1.6 Verifying in the search result is relevant, not really important inside the Web page.

TP6 It's more difficult with unmarked objects.

2.1 Unknown object, more active search. Easier with a known object.

TP3 Learned that the images in the result list was the matched images, didn't care about the other images that was not correctly matched.

2.8 Matching images verify the search, in the result list and in the web page.

TP9 Enters several web pages, thinks that the matched images are not true.

TP7 Looks first at the links, and then looks at the matched images.

K- Image features

4.9 How does one chose objects inside a photo to be searched on?

TP8 Size does matter.

1.2 Is it correct size?

TP3 The scale was a problem, the TP was really thoroughly when measure the physical object (iittala) comparing with the matched image.

L- Modality

TP5 Asked if the system could find information even if the web page doesn't contain that image. When we said how IDeixis worked, that one of their prototypes had the functionality that after taking a photo, a set of keyword was returned to be used for a next search. Respondent thought this was great.

TP10 Saw Iittala in the snippet, then used this to evaluate other links.

TP2 Found Info in snippets (year of publication), two years were presented. Chosed Amazon, because respondent trusted this more than the other.

TP1 Change modality, finds the name in the snippet and searches with that instead of images.

TP8 Found the word "Discography" in the snippet.

TP9 "images are another way to express" It's another spectrum of information search.

M- Objects

4.5 One can not photograph if not standing in front of it.

1.2 A need of finding a physical object that one wants information of.

N- Prerequisite knowledge

TP8 Says that one can find discography on the web shop.

TP3 Knowledge about what pages contain what, is an advantage.

O- Areas of use

TP 6 Wants to take photos and then at home search with them.

4.1 A complement to other search methods.

1.10 If one has a physical photo of something that one doesn't know what it is.

TP 7 "One always has the mobile phone at reach.

P- What's on the Internet

4.2 A risk that their might be too many images on the Internet.

1.7 There is a risk for private tourist images.

4.8 The image has to be on the Internet.

2.6 The images must be on the Internet, the search is therefore limited.

4.3 The system only works if one supposes to buy. Would like to have a mode of finding persons.

2.6 The more web pages that has similar images, the more likely to find a web page that fits.

4.6 The result depends on the amount of images on the Internet.

4.10 The amount of images and what kinds of images on Internet affect the search.

4.2 Limits what kind of images there is on the Internet.

4.8 The freshness of the images.

Appendix 7: Questionnaire (round two)

This is the original questionaire given to respondents after conducting the test in the Second round of prototyping. It is presented here in Swedish, the original language. Translations of the questions can be found in the analysis of the questionnaire data starting at page 82.

Du ska nu sätta betyg och jämföra mellan ”vanlig” Google-sökning med text på mobiltelefon och sökning med hjälp av ett foto som du just har provat. Ringa in det svarsalternativ du tycker passar bäst in på de olika sökalternativen (1 är negativt, 5 är positivt).

Hur bra anser du söksystemet kan hjälpa dig hitta information...

1.	...om okända föremål du har inom synhåll?				
Google (text)	1.	2.	3.	4.	5.
Fotosökning	1.	2.	3.	4.	5.
2.	...om föremål av samma typ som det du har inom synhåll?				
Google (text)	1.	2.	3.	4.	5.
Fotosökning	1.	2.	3.	4.	5.
3.	... om kända föremål, t.ex. föremål märkta med text?				
Google (text)	1.	2.	3.	4.	5.
Fotosökning	1.	2.	3.	4.	5.
4.	s...om pris/inköpsställe för föremål?				
Google (text)	1.	2.	3.	4.	5.
Fotosökning	1.	2.	3.	4.	5.
5.	...tidsmässigt tillfredställande?				
Google (text)	1.	2.	3.	4.	5.
Fotosökning	1.	2.	3.	4.	5.