DESIGNING AN IMAGE RETRIEVAL INTERFACE FOR ABSTRACT CONCEPTS WITHIN THE DOMAIN OF JOURNALISM

Ron Besseling

University of Amsterdam
Centrum Wiskunde & Informatica
The Netherlands

ABSTRACT

Research has shown that users have difficulties finding images which illustrate abstract concepts. We carried out a user study that confirms the finding that the selection of search terms is perceived difficult and that users find the subjectivity of abstract concepts problematic. In addition, we found that abstract queries were mostly used to get inspiration for followup queries. Search terms became more concrete as the search task progressed and the idea of the final image took shape. Based on this, we specified three user requirements: (1) representation of multiple interpretations for a single abstract concept, (2) ability to include synonym image results based on the initial query and (3) the need to support the transition of abstract to concrete search queries. We translated the user requirements into an interface design and evaluated the mockup with end-users. All three corresponding functionalities were perceived useful by participants. The ability to select a sub-concept of an abstract concept was reported most useful, because it improves the precision of the returned images.

Index Terms— image retrieval, abstract concepts, user interface, journalism

1. INTRODUCTION

Image retrieval for abstract concepts is the search for image content that is not directly present in the image, but needs to be inferred from background knowledge and experience [13]. For example, the explicit image content of Figure 1 is a heart shape made with cereal. However, implicitly it may express the abstract concept *love*. Abstract concepts are not directly present in the image, but may be associated with it. Other examples of abstract concepts are: *freedom*, *passion*, *happiness*, *glamour* and *admiration*.

In general, image perception can be subdivided into three levels: (1) primitive features (e.g. color, shape, texture), (2) objects (e.g. person, location, event) and (3) inductive interpretation (e.g. symbolic value, emotional cue, abstraction) [17]. In the case of Figure 1, a Level 1 feature may be the color *brown*. The shape *heart* is an example of a Level 2 feature and the concept *love* can be categorized under Level 3. This paper focusses on the latter which can be defined as the user's subjective interpretation ascribed to the symbolism of an image [17]. For the remainder of this article we use the term *abstract concept* to refer to a Level 3 concept. In addition, we make a distinction between the *information need* and *query* of a user. An information need is the topic about which the user desires to know more, and is differentiated from a query, which is what the user conveys to the computer in an attempt to communicate the information need [24]. In the case of Figure 1, the information need may be an image that represents *a romantic breakfast* and a corresponding query may be *love*.

Fig. 1: Stock photo of a heart shape made with cereal.

Studies show that journalists attach great value to abstract concepts when selecting preferred image(s) [13, 25]. However, image retrieval for abstract concepts is considered problematic [25]. The selection of search terms for abstract concepts is considered difficult [25]. Journalists presume that there are images in the archive related to their queried abstract concepts, but they do not know the right search terms to retrieve them. They avoid search tasks with abstract concepts, because they do not expect the system to be able to interpret them adequately [13]. Overall, journalists find image retrieval for abstract concepts more time-consuming than searching for images of specific objects or persons [25]. Journalists report that they do not even consider illustrating articles with abstract themes when they do not have the time for a lengthy

search process [25].

Research on image retrieval within the domain of journalism mainly focuses on search and annotation behavior of end-users [13, 20, 25, 29]. In this paper we combine the results of these studies with a user study to produce the user requirements and design of an image retrieval interface for abstract concepts. The research question for this study is:

> how can we visually support users with image retrieval tasks for abstract concepts within the domain of journalism?

The main research question is subdivided into three subquestions:

1. what are the characteristics of image retrieval tasks for abstract concepts;
2. what are the user requirements for an image retrieval interface for abstract concepts;
3. what is the relative importance of functionalities that support image retrieval for abstract concepts?

In order to answer the research question, we need to understand end-users' search behavior and identify the main problems with image retrieval for abstract concepts. The study described in this paper is explorative, because existing research on image retrieval for abstract concepts is limited. No study within the domain of journalism focusses specifically on abstract concepts and no corresponding image retrieval interface has been designed yet.

We used a user-centered design approach [16] to ensure the usefulness of the final interface for end-users [28]. The study consists of five phases. First we performed a literature study to identify problems with image retrieval for abstract concepts. Next, we conducted a user study to verify the current problems and discover users' needs for performing image retrieval tasks. Based on the user study, we formulated the user requirements and determined the key features for an image retrieval interface for abstract concepts. We then used the user requirements as a basis for the mockup design of the interface. Finally, we evaluated the interface with end-users to test whether the functionalities, which support image retrieval for abstract concepts, were perceived useful.

This paper is organized as follows: related work is given in Section 2. The third section describes the user study that allowed us to derive user requirements described in Section 4. In Section 5 we propose an interface design, followed by an evaluation of the interface in Section 6. The study is discussed in the seventh section. Lastly, we present conclusions and future work.

2. RELATED WORK

No image retrieval interface for abstract concepts has yet been designed. A possible reason for this is because abstract queries are less frequently used than queries about persons and objects [25, 29]. Several studies have been performed on search and annotation behavior of end-users. These studies do not specifically focus on image retrieval for abstract concepts, but present several findings that are relevant to our study.

Eakins et al. questioned 125 experienced image searchers through an online questionnaire about the functions and modes of interaction that they currently use, and those they would like to see in future systems [13]. The results showed that the ability to retrieve images by their semantic content is a clear priority for users of image databases. Lower level issues are generally considered less important, which reinforces earlier studies [14].

Markkula and Sormunen performed a field study on journalist practices in requesting, searching for and selecting photos in the course of their daily work [25]. The results suggest that image retrieval for abstract concepts is considered problematic, as described in the introduction.

Greisdorf & O'Connor [17] and [12, 15, 22, 23, 27] provide hierarchies of perception in the searching and evaluation of images. These frameworks all place abstract concepts at the highest level of perception. The study of Greisdorf & O'Connor suggests that a system for capturing human interpretations derived from images could enhance the efficiency of image retrieval for abstract concepts.

Our conclusion from these studies is that image retrieval for abstract concepts is a task desired by users, but for which there is yet insufficient technical support.

3. PRELIMINARY USER STUDY

The preliminary user study was conducted with two goals in mind. Firstly, to identify the main problems that users face when they perform image retrieval for abstract concepts. Secondly, to derive the characteristics of image retrieval tasks for abstract concepts within the domain of journalism. These characteristics are used to define search tasks for the evaluation of the interface in Section 6.2.

3.1. Setup

We questioned six image editors at their working environment in an one hour semi-structured interview. The interviews consisted of four parts, starting with an introduction about the study and general demographic questions. We then asked the participant in open questions (1) which image search engines s/he uses, (2) which type of search task s/he performs the most (e.g. people, objects, landscapes etc.) and (3) how long a typical image search task takes to complete.

The third part of the interview focussed on image retrieval for abstract concepts. We asked questions 1 and 3 of the previous part again, but this time narrowed down to abstract concepts. We could thus compare the frequency, time length and

difficulty of object image retrieval tasks with abstract image retrieval tasks.

In the last part of the interview, the participant performed an image retrieval task for an abstract concept on a laptop we brought. The task was specified by the participant and was often a search s/he needed to perform for work. We asked the participant to think aloud and interviewed him/her while s/he carried out the task. Afterwards, we asked which problems s/he encountered and if s/he had any suggestions for the improvement of search engines.

In total, six image editors (three male, three female) of three different organizations were interviewed. The participants' average age was 31 years (SD=5.0). All interviews were screen and voice recorded for documentation.

3.2. Results

The participants' comments are divided into three categories based on their relative importance for our study, namely: general observations, reported problems and task observations.

3.2.1. General observations

Examples of frequently visited image search websites are: Getty Images [4], Corbis Images [3], Shutterstock [10], iStockphoto [7], Hollandse Hoogte [5], ANP [1], AP [2] and Reuters [9]. Of these websites, Getty Images and Corbis Images are reported most suitable for image retrieval for abstract concepts. This is mainly due to the advanced search and filter functionalities.

We asked participants which current functionalities they found important in image search websites. There was no consistency in the functionalities mentioned by the participants. For example, one participant emphasized the ability to download descriptions and keywords with images, while another pointed out the ability to filter on composition and color. The mentioned functionalities are not further used in this study, because they were not specifically useful for abstract concepts.

Participants reported that the time spent to complete one image retrieval task differs greatly per task: between five and ninety minutes. There is also a large variation in the percentage of image retrieval tasks for abstract concepts (Level 3) compared to objects (Level 2): between 20% and 80%. All six participants stated that image retrieval for abstract concepts is more difficult and time-consuming than regular image retrieval. These statements confirm the studies reported in Eakins et al. [13] and Markkula & Sormunen [25].

Most participants did not have any suggestions for the improvement of image search engines. However, one participant mentioned his/her ideal situation for image retrieval for abstract concepts. S/he described this as a single textfield where s/he types the person, object, location and mood of the image, followed by a result page with relevant images only. The participant mentioned that this is not a realistic situation, mainly because manual tagging of images is subjective by nature.

3.2.2. Reported problems

Several participants mentioned the subjectivity of abstract concepts as the main problem with image retrieval for abstract concepts. The comments related to this issue are subdivided into three categories based on similarity.

1. Abstract concepts often have different meanings and the retrieval system does not know which interpretation the user is searching for. For example, one participant mentioned that jealousy in a relationship is different from materialistic jealousy. Images belonging to both interpretations are nonetheless mixed on the result page, which causes a number of images to be irrelevant.

2. One abstract concept can be expressed by different words. Participants mentioned that they repeatedly try different queries to search for the same abstract concept, because the right search term could make a huge difference in the number of relevant results that are found. This confirms the study of Markkula and Sormunen [25], that journalists often do not know the right search terms to retrieve the images visualizing the abstract concept that they are looking for.

3. The person who tags the image can have a different interpretation of the image from the user who is searching for it. Participants reported that they regularly encounter images with tags that seem illogical to them.

3.2.3. Task observations

The task observation partly confirmed the answers given in the interview. For example, participants who mentioned having trouble with finding the right search term used synonym and translation websites to adjust their query. Most participants did not use the concept filters on the website. Instead, they added search terms to their query to specify their request. In general, search terms became more concrete as the search task progressed. Participants used abstract search terms to get inspired. They specified their query when they saw an image they liked, sometimes using the tags of the image. For example, one participant used the search term *shame* and encountered an image of a man with a bag over his head. S/he looked at the tags of the image and selected the tag *obscured face* to be added to his/her search terms.

4. USER REQUIREMENTS

We specify three user requirements for an image retrieval interface for abstract concepts. The requirements are based on

reported problems 1 & 2 (Section 3.2.2) and the task observation (Section 3.2.3). Reported problem 3 is currently unsolvable due to the subjective nature of the task. Other results did not directly contribute to the user requirements, but may be useful for further research (Sections 3.2.1). In addition, requirements such as the ability to view pricing, buy the image and print a preview are intentionally left out. They apply to image retrieval in general and not specifically to image retrieval for abstract concepts.

4.1. Multiple interpretations

Most abstract concepts can have multiple interpretations. For example, the concept *love* may refer to a romance between two lovers, a heartwarming relationship between a father and son, or a solid friendship between two girls. Depending on the context, each interpretation may be what the user is looking for when using the search term *love*. This problem was mentioned by three participants in the interview and observed by four during the task observation. Images belonging to the different interpretations are nonetheless mixed on the result page, which causes a number of images to be irrelevant

Based on these findings, we can conclude that there is a need for the representation of multiple meanings for a single abstract concept. This allows users to immediately discard irrelevant interpretations of an abstract concept, which can potentially improve the *precision*[1] of the image retrieval system.

4.2. Synonyms

As argued in the introduction, the selection of search terms for abstract concepts is considered difficult [25]. Journalists presume that there are images in the archive related to their queried abstract concepts, but they do not know the right search terms to retrieve them. Some participants confirmed this during the task observation, by repeatedly trying different queries to search for the same abstract concept. There is a mismatch between the query of the user and the synonym(s) associated with an image, which causes the image to be left out the results even though it may be relevant.

Based on these findings, we can conclude that there is a need for an image retrieval system that includes synonym results based on the initial query. This may result in a better user experience since the user finds more relevant images, improving the *recall*[2] of the image retrieval system.

4.3. Concept exploration

The task observation showed that abstract search terms are often based on the theme or mood that the image editor has in mind after reading the article s/he needs to illustrate. Image results of abstract concepts are used as inspiration for followup queries. The search terms become more concrete as the search task progresses and the idea for the final image takes shape. This can be defined as a transition from abstract (Level 3) to object (Level 2) queries, as described in the introduction. However, sometimes an image editor immediately has an idea of the preferred final image after reading the article. In this case, abstract concepts are skipped and the search terms are concrete from the start.

Current popular image search engines do not support the transition of abstract to concrete search queries. Users can manually add search terms to their query, but they are not guided during their search task. Based on the task observation, we can conclude that there is a need to support this transition. If we visually assist the user with this process, s/he would not have to go through numerous images to find related concrete search terms to their abstract concept. Ultimately, this may result in more efficient searches.

5. DESIGN

The above user requirements form a basis for an image retrieval interface for abstract concepts. With the use of HotGloo [6] we designed a mockup of the interface. We kept the mockup low-fidelity to ensure that the user would focus on the functionalities during the evaluation and not on the esthetics of the interface.

The interface consists of six areas (Fig. 2): a search field (A), search options (B), concept specification (C), related search terms (D), a filter (E) and the image results (F). The search field, search options, filter and image results are default functionalities for a stockphoto search engine. Concept specification (C), related search terms (D) and the ability to include synonym results (B1) are novel functionalities and based on the user requirements.

In general, the interface shows much similarity with popular stock image search engines. This is done on purpose to provide consistency with image search engines that endusers are currently used to. The most popular stock photo search engines share the same global lay-out and changing this could have a negative effect on the learning-curve of the search engine. Additionally, changing the common lay-out can influence the evaluation of the new functionalities as participants may focus on design changes instead of the usefulness of functionalities. We will now go further into each new functionality and describe the intention of its design.

5.1. Concept specification (Fig. 2, C)

Based on the requirement *multiple interpretations* (Section 4.1), we designed a medium sized area where the user can select the *sub-concept* s/he is looking for (C). The query (A) and image results (F) change when the user clicks on an image or its corresponding label (C2). In this example three subconcepts are shown, but this number depends on the queried abstract search term.

[1] Precision is the percentage of retrieved items that are relevant [18].
[2] Recall is the percentage of all relevant items that are retrieved [18].

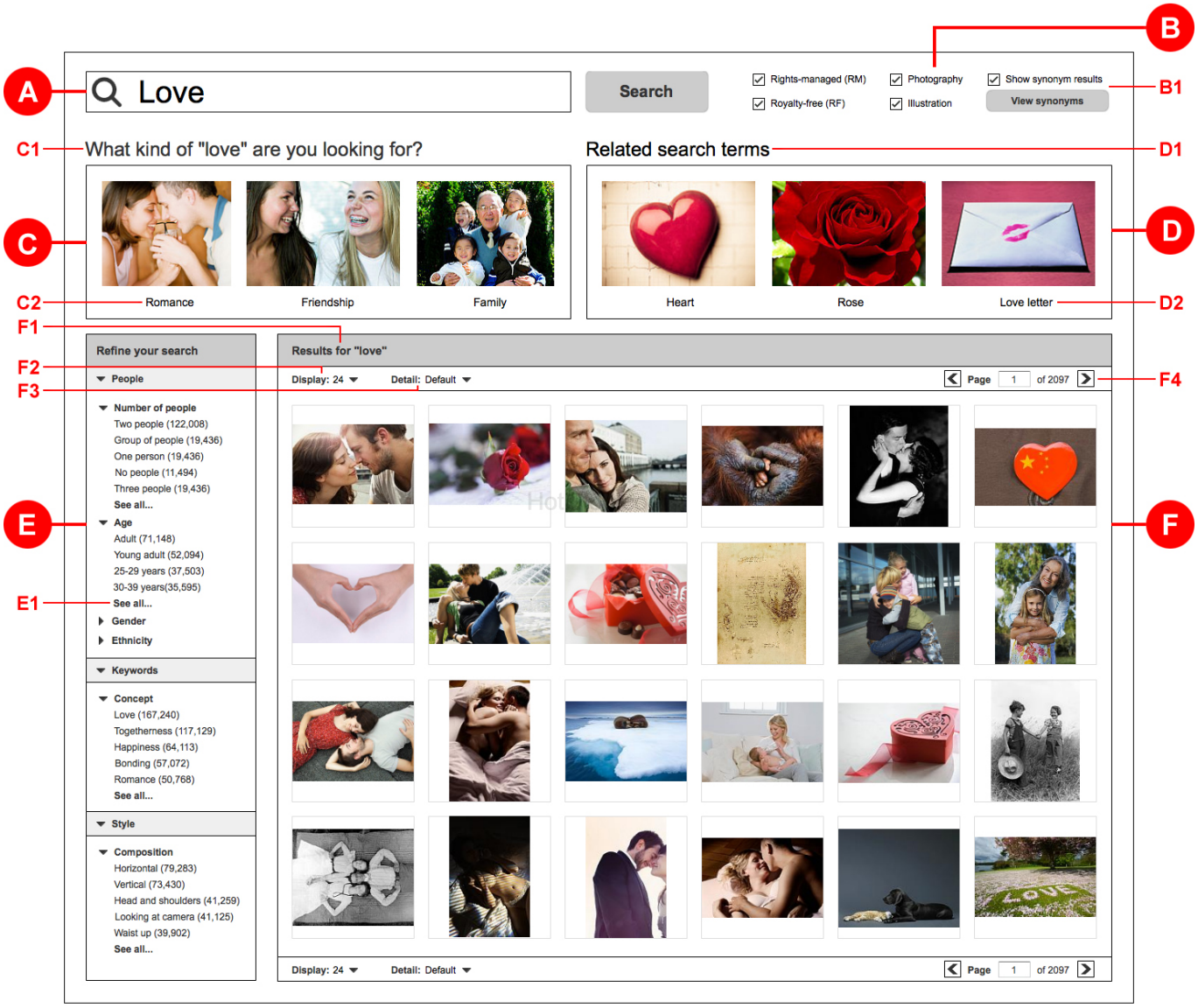

Fig. 2: Mockup of the image retrieval interface for abstract concepts, with the example query *Love*.

Main functionalities: (A) search field, (B) search options, (C) concept specification, (D) related search terms, (E) filter and (F) image results.

Details: (B1) show synonym results, (C1) concept specification title, (C2) concept specification label, (D1) related search terms title, (D2) related search terms label, (E1) see all filters, (F1) image results title, (F2) number of image results, (F3) detail of image results and (F4) page navigation.

The question *What kind of "love" are you looking for?* (C1) is to trigger the user to choose a sub-concept. A title is more passive and in this specific case more complex. For example, the titles *Concept specification*, *Sub-concepts* or *Categories* do not cover the functionality in a clear manner. The font size is larger than the default, but smaller than the search terms to emphasize its relative importance. This is also the reason why images belonging to sub-concepts are bigger than the image results. The corresponding labels (C2) are relatively small, because the images themselves visualize the sub-concepts. The labels have a supporting role in case the image is not clear enough.

We considered using text only to express sub-concepts, because images are more subjective than words. For example, it may be the case that the user does not like the image that represents a sub-concept. This can prevent him/her from choosing a sub-concept, while the image set belonging to it may contain relevant results. We chose images in combination with labels, so we can inspire the user with images for followup queries and clearly define the sub-concepts.

5.2. Show synonym results (Fig. 2, B1)

Based on the requirement *synonyms* (Section 4.2), we included a function to show synonym results based on the initial query. The checkbox is checked by default, but can be unchecked manually. The button *View synonyms* prompts a screen where the synonyms are shown with the ability to individually disable them. This may be useful as search tasks progress and search terms become more concrete.

The functionality is located within the search options, because it does not have to attract attention. Based on the result that many image editors use synonym queries (Section 3.2.3), we assume that users only have to occasionally interact with it when the synonym image results are not desirable. In this case, the checkbox will remain unchecked until the user checks it again.

5.3. Related search terms (Fig. 2, D)

Based on the requirement *concept exploration* (Section 4.3), we designed a medium sized area for concrete query suggestions related to the queried abstract search term (D). The query (A) and image results (F) change when the user clicks on an image or its corresponding label (D2). In this example three suggestions are shown, but this can be expanded with an interactive interface by adding a horizontal scrollbar.

The title *Related search terms* (D1) explains that the suggestions are related, but not that they are concrete. We knowingly left this out, since we think it will cause confusion because the end-user may wonder what is meant with the term *concrete*. The font size is equal to the concept specification title (C1) to emphasize its relative importance. However, the title is not in question form since there is no action required from the user before s/he looks at the image results.

As with the concept specification, we considered using text only to express the query suggestions. We chose images in combination with labels for the same reason we chose them for the concept specification: the ability to inspire end-users for followup queries is more important than reducing the chance of wrongly declining a related search term.

6. EVALUATION STUDY

An evaluation study was performed with two goals in mind: to test the usefulness of the interface and determine the relative importance of functionalities that support image retrieval for abstract concepts.

6.1. Setup

We questioned three image editors at their working environment in an one hour semi-structured interview. The interviews had a similar setup to those in the user study and consisted of three parts, starting with an introduction about the study and general demographic questions. We then asked the participant to perform an image retrieval task for abstract concepts on a laptop we brought. The task was specified by the participant and performed with any image search engine of his/her choice. It was often a search task s/he needed to perform for work. We asked the participant to think aloud and questioned him/her while s/he carried out the task. The main goal of this part was to verify that the participant was familiar with image retrieval for abstract concepts and to get him/her used to thinking aloud. Additionally, observations were used to confirm the results of the preliminary user study.

In the second part, participants were asked to perform predefined tasks using static mockup interfaces. The mockups were presented without explaining the interface upfront. The participant was asked to think aloud and express which actions s/he wanted to perform by moving the mouse over the mockup. The interviewer "played the computer" by providing feedback and describing the changes of the interface. If the participant went quiet, we motivated him/her with questions such as *"What do you see?"*, *"What would you do?"* and *"What stands out?"*.

In the last part of the evaluation we explained the new functionalities followed by a semi-structured interview about the mockup. We asked how useful and complex they perceived a functionality, if and when they would use it and if they had any suggestions for improvement.

In total, three image editors (all female) from three different organizations were interviewed. The participants' average age was 28 years (SD=2.9). We interviewed different image editors than the ones from the user study, because their knowledge of our research and influence on the user requirements could have resulted in a biased evaluation. All interviews were screen and voice recorded for documentation.

6.2. Tasks

Each participant performed three search tasks that were presented in random order. The tasks were based on those chosen by participants during the preliminary study, namely to search for an image to illustrate an article about:

1. shame that comes with abortion;
2. children's fear of nightmares;
3. relationships with large age differences.

The search tasks are not identical to those used by participants during the preliminary study, but the main abstract concepts are. For example, the search task about shame that comes with abortion is based on a task where the participant needed to illustrate an article about the usefulness of shame.

The image results used in the mockups were obtained by using the same query in Getty Images [4]. We defined the sub-concepts based on these results and used them for the concept specification. Additionally, we checked the image results for concrete tags and used them for the related search terms (see Table 1). The advantage of this approach is that the image results are not biased in our favor.

Table 1: Search task values

Task	Search term	Sub-concepts	Related terms
1	shame	embarrassment guilt	overweight nudity failure
2	fear	danger anxiety spooky	murder spider ghost
3	love	romance friendship family	heart rose love letter

6.3. Results

We report the results corresponding to the three parts of the interview: task observation, mockup walkthrough and closing interview.

6.3.1. Task observation

The task observation confirmed results of the user study (Section 3.2). Getty Images [4], Corbis Images [3], Shutterstock [10] and iStockphoto [7] were the most frequently used stock photo search engines. All participants' queries became more concrete as the search task progressed and some used synonyms to find more related image results.

6.3.2. Mockup walkthrough

Two of the three participants reported that they noticed the search field with the predefined search term first. Next, they quickly scanned the image results above the fold[1] and shortly thereafter noticed the concept specification and related search terms.

All participants understood the function of concept specification. One participant was confused by the sub-concepts at first, but quickly figured out the meaning after reading the title *What kind of ... are you looking for?* Concept specification was used by all participants, but not during every task. This was mainly because not every abstract concept had distinctive sub-concepts for the particular search task. For example, most participants did not know if the sub-concept *embarrassment* or *guilt* best fitted an article about shame that comes with abortion. However, the opposite was the case with an article about relationships with large age differences, where all participants immediately chose the sub-concept *romance*.

All participants understood the function of the related search terms. One participant thought that they were sub-concepts at first, but quickly figured out the difference when s/he noticed the title *Related search terms*. Participants hardly used the suggestions during the search tasks, because they were not useful to the predefined search tasks. For example, one participant mentioned that the search terms *overweight*, *nudity* and *failure* were not very helpful for the illustration of an article about shame that comes with abortion. Note that no sub-concepts are selected in the mockup, so the suggestions are based on the abstract concept *shame* in general.

No participant noticed the functionality to show synonym image results. This functionality was discussed with the participants in the closing interview.

6.3.3. Closing interview

Concept specification was reported useful by all three participants. They appreciate the ability to specify the meaning which they are looking for, thus making the image results more relevant. One participant mentioned that it creates the feeling that the computer thinks with you by helping you explore the queried concept. Another participant mentioned that the functionality is only useful for abstract search terms where one sub-concept is relevant to their information need. All three participants would only select a sub-concept when they are sure that the other sub-concepts do not contain images relevant to their search task. This confirms the finding of Markkula & Sormunen [25], where journalists stated that very narrow queries might exclude the best photos from the result set. One participant suggested to support queries with multiple abstract concepts, by creating sub-concepts that combine the different abstract search terms. However, s/he realized

[1] Above the fold is the part of a web page that is visible without scrolling. [8].

afterwards that s/he hardly uses queries with more than one abstract search term in it.

The synonym results were reported useful by all participants. They appreciate that the search engine automatically finds related synonyms, because most participants are used to searching for synonyms themselves. All participants mentioned that they had not noticed the functionality during the mockup walkthrough, because it did not stand out. Most of them thought it is a good choice to activate it by default. However, one participant emphasized that s/he would probably turn it off as the search task becomes more specific and the idea of the final image becomes more clear.

The related search terms were reported useful by all participants, although no one used it during the mockup walkthrough. Most participants appreciate the suggestions because the images may inspire them for followup queries. However, they said that this was not the case with the predefined tasks we observed. On the followup question whether wrongly suggested search terms may be annoying, participants answered that this was not the case because they just ignore them after a quick glance. Furthermore, one participant mentioned his/her concern that the related search terms may be clichés only, because they are based on popularity.

6.4. Key findings

The evaluation confirmed the usefulness of the functionalities. However, it also brought up unforeseen issues.

Concept specification (Section 5.1) was frequently used during the mockup walkthrough and reported useful in the closing interview. During the mockup walkthrough it became clear that the distinctiveness of sub-concepts is key to the usefulness of the functionality. A sub-concept has to share minimum similarity with other sub-concepts, because otherwise users do not know which sub-concept best represents the idea in their head.

The ability to include synonym results (Section 5.2) was not noticed during the mockup walkthrough, but reported useful in the closing interview. This result was not surprising, since the functionality was deliberately designed to not attract attention. Participants appreciate that it is on by default, but emphasize the importance of being able to turn it off as the search task becomes more specific.

Related search terms (Section 5.3) were not used during the mockup walkthrough, but reported useful in the closing interview. This was because the images used in the mockups were not manually adjusted to fit the search tasks as, but obtained by using the same query in Getty Images [4] as explained in Section 6.2. As a result, the suggestions in the mockup were often not closely related to the search tasks. This issue is reduced if the user chooses a sub-concept, so the search engine *knows* which interpretation of the abstract concept s/he is looking for. One participant confused the related search terms with concept specification, probably due to the strong resemblance in lay-out and design. The final interface design should create better distinction between the two, by varying the size, color and shape for example.

7. DISCUSSION

In Section 7.1 and 7.2 we discuss the choices we made during the research which could have affected our result. In section 7.3 we discuss the technical feasibility of the novel functionalities described in Section 5.

7.1. Used images

The images used in the mockups were obtained by using the same query in Getty Images [4], as described in the evaluation (Section 6.2). The advantage of this approach is that the image results are not biased in our favor. One can assume that the functionalities were perceived more useful if we adjusted the image results to fit our predefined tasks. However, it also showed us that the related search terms can be useless if the context of the search task is not defined. For example, the search terms *overweight*, *nudity* and *failure* seem related to the abstract concept *shame*, but are not very helpful for the illustration of an article about shame that comes with abortion.

In the three mockups, only one of the nine related search terms may have been useful to the corresponding search tasks. This is probably the reason why participants did not use the related search terms during the mockup walkthroughs. Concept specification can improve the precision of the related search terms by providing the interpretation of the abstract concept. Nevertheless, the issue will probably also concur in practice, since the context of a search task is not known to the search engine at the beginning of the task.

7.2. Qualitative data

The study is explorative and based on qualitative data only. We considered using a survey for gathering user requirements, but selected a user study with the assumption that users will have a hard time expressing their needs in a predefined survey. We still support this decision since most participants expressed their issues only imprecisely in the interview, but the task observation gave us a clear understanding of the problem and inspired us towards possible solutions. For example, one participant mentioned that s/he tried different queries to find the ”right” search term for an abstract concept. During the task observation we noticed that s/he used Google to find synonyms, which inspired us to include synonym results. The user requirement *multiple interpretations* (Section 4.1) was not mentioned at all in the interviews, but observed during the task observation. It turned out to be the most useful perceived functionality, as described in the evaluation (Section 6.4).

The results of this study are not representative due to the small sample size. However, the consistency in responses from different participants indicates that the selected functionalities would be useful to other users as well.

7.3. Technical realization

The technical details of the described functionalities (Section 5) are not part of this study. Nevertheless, we give an indication how they may be realized in practice.

A lexical database, such as WordNet [11], is needed to distinguish abstract (Level 3) from object (Level 2) words which are associated with the resulting images. The sub-concepts may automatically be determined by selecting the most frequent tags which are classified as abstract concepts in the lexical database and associated with resulting images that contain the queried abstract concept. In addition, the related search terms may be determined the same way by automatically selecting the most frequent object tags.

The image chosen to represent a sub-concept may be selected by using the most clicked image of that image set. This method cannot be used to determine the image that represent a related search term, because this will probably result in a cliché image that needs to be avoided, as described in Section 6.3.3. Presumably, a manual selection has to be performed, preferred by image editors. Several techniques can be used to learn more about the context of the search task as it progresses [19, 21, 26, 30]. This can improve the precision of the related search terms. Additionally, the lexical database can be used to determine synonyms related to the queried abstract concept and extend the query with these synonyms.

8. CONCLUSION & FUTURE WORK

Research question 1 - We carried out a preliminary user study to understand current problems with image retrieval for higher level concepts and to give us insights in the search behavior of end-users. The selection of search terms was perceived difficult and participants found the subjectivity of abstract concepts problematic. We found that image retrieval for abstract concepts is mostly used to get inspiration for followup queries. In addition, search terms became more concrete as the search task progressed and the idea of the final image took shape.

Research question 2 - Based on the user study, we specified three user requirements: (1) representation of multiple interpretations for a single abstract concept, (2) ability to include synonym image results based on the initial query and (3) the need to support the transition of abstract to concrete search queries.

Research question 3 - The user requirements were translated into an interface design which we evaluated with end-users. All three novel functionalities were perceived useful by participants. The ability to select a sub-concept of an abstract concept was reported most useful, because it improves the precision of the returned images. Furthermore, the ability to include synonym results was reported second most useful, because it improves the recall. The related concrete search terms were perceived least useful in comparison with the other two functionalities, because they often did not fit the context of the search task.

The study described in this paper is based on qualitative data only. Future quantitative research is needed to confirm the user requirements. The interface design needs to be converted into an interactive prototype and evaluated with end-users, e.g. by comparing it with current stock photo search interfaces. More research is needed to study the selection of sub-concepts, since we found that it influences the perceived usefulness of concept specification.

Current stock photo search engines can integrate the described functionalities into their interfaces to support image retrieval for abstract concepts. Our study suggests that this will improve the user experience and efficiency of image retrieval tasks for abstract concepts.

9. ACKNOWLEDGEMENTS

We thank Frank Schallmaier and Heike Gülker of de Volkskrant; Marijn van der Meer, Gaby Baas and Niels Broekema of Quest Magazine; Nienke Bekking of Kassa Magazine; Marloes Engelhart of Vara Magazine; Liselot M'charek-Rijsdijk of Supertrash Magazine and Sara Mok for their time regarding the user interviews. We also thank Spinque for the origin of this study. Last, but certainly not least, we thank Lynda Hardman, Vera Hollink and Arjen de Vries for their support and useful feedback.

10. REFERENCES

[1] Anp - www.anp.nl.

[2] Ap - www.ap.org.

[3] Corbis images - www.corbisimages.com.

[4] Getty images - www.gettyimages.com.

[5] Hollandse hoogte - www.hollandse-hoogte.nl.

[6] Hotgloo - www.hotgloo.com.

[7] istockphoto - www.istockphoto.com.

[8] Marketing terms - www.marketingterms.com.

[9] Reuters - www.reuters.com.

[10] Shutterstock - www.shutterstock.com.

[11] Wordnet - wordnet.princeton.edu.

[12] B. Burford, P. Briggs, and J. P. Eakins. A taxonomy of the image: On the classification of content for image retrieval. *Visual Communication*, 2(2):123–161, 2003.

[13] J. Eakins, P. Briggs, and B. Burford. Image retrieval interfaces: A user perspective. *Lecture Notes in Computer Science*, 3115:49, 2004.

[14] P. Enser. Pictorial information retrieval. *Journal of Documentation*, 51:126–170, 1995.

[15] P. Enser and C. Sandom. Retrieval of archival moving imagery - a step too far for cbir? *Multimedia Content-Based Indexing and Retrieval Workshop*, 2001.

[16] J. Garrett. *The Elements of User Experience: User-Centered Design for the Web*. Pearson Education, 2002.

[17] H. Greisdorf and B. O'Connor. Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation*, 58(1):6–29, 2002.

[18] M. A. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.

[19] S. C. H. Hoi, W. Liu, M. R. Lyu, and W. Ma. Learning distance metrics with contextual constraints for image retrieval. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:2072–2078, 2006.

[20] L. Hollink, A. T. Schreiber, B. Wielinga, and M. Worring. Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61(5):601–626, 2004.

[21] H. Jegou, H. Harzallah, and C. Schmid. A contextual dissimilarity measure for accurate and efficient image search. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[22] C. Jörgensen. Access to pictorial material: A review of current research and future prospects. *Computers and the humanities*, 33(4):293–318, 1999.

[23] L. Keister. User types and queries: Impact on image access systems. *Challenges in Indexing Electronic Text and Images*, pages 7–22, 1994.

[24] C. D. Manning, P. Raghavan, and H. Schütze. *An Introduction to Information Retrieval*. Cambridge University Press, April 2009.

[25] M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.

[26] N. Rasiwasia and N. Vasconcelos. Image retrieval using query by contextual example. *Proceeding of the 1st ACM international conference on Multimedia information retrieval*, pages 164–171, 2008.

[27] S. Shatford. Analyzing the subject of a picture: A theoretical approach. *Cataloging and Classification Quarterly*, 6(3):39–62, 1986.

[28] K. Vredenburg, J. Mao, P. Smith, and T. Carey. A survey of user-centered design practice. *CHI '02 Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 471–478, 2002.

[29] S. Westman and P. Oittinen. Image retrieval by end-users and intermediaries in a journalistic work context. *Proceedings of the 1st international conference on Information interaction in context*, pages 102 – 110, 2006.

[30] X. Xing, Y. Zhang, and B. Gong. Mixture model based contextual image retrieval. *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 251–258, 2010.