A Classification Scheme for User Intentions in Image Search

Mathias Lux
Klagenfurt University
Universitätsstr. 65-67
Klagenfurt, Austria
mlux@itec.uni-klu.ac.at

Christoph Kofler
Klagenfurt University
Universitätsstr. 65-67
Klagenfurt, Austria
christoph.kofler@itec.uni-klu.ac.at

Oge Marques
Florida Atlantic University
777 Glades Rd.
Boca Raton, FL 33431 USA
omarques@fau.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Searching for images on the web is still an open problem. While multiple approaches have been presented, there has been surprisingly little work on the actual goals and intentions of users. In this poster we present our classification scheme for user goals in image search and describe our ongoing work focusing on identification and classification of user intentions during image search tasks.

Keywords
User intentions, image retrieval

ACM Classification Keywords
H.1.2 User/Machine Systems: Human factors.

General Terms
Human Factors, Theory, Experimentation

Introduction
Human behavior is typically not random, undirected or unintentional. We postulate that nearly every conscious interaction of a human agent with a computer is driven by an intention. For computer science, the concept of "intentions" is arguably fuzzy, context-dependent, vague, and hard to measure. However, the concept of a "goal" can be defined easily: *a goal is a state of affairs*

that a user tries to achieve. Likewise, the (partial) success of achieving a goal can be measured. An example illustrates the idea: someone needs an image to design a printed brochure. At a certain page, the user has the intention to add an image. As soon as the user starts an action to retrieve an adequate image, a clear goal can be defined by the objective (find an image), constraints (adequate for the brochure in size, color and quality), and success criteria (image was successfully downloaded and added to the brochure).

Within our work we assume that knowing the goal of users allows for better support of their actions. So, if the goal is to find an image for a brochure, the search might be constrained by factors such as license issues or color schemes for the overall brochure design. On the other hand, if the goal is simply to find an image for entertainment purposes, such constraints are not necessarily present.

The detection of user goals in text retrieval has been addressed by several researchers in the past decade. Broder [1] and Jansen et al. [2] introduced a taxonomy for classification of user goals in text-based web search. This taxonomy consists of three top level classes:

- *Informational*: retrieve content to obtain new knowledge.
- *Navigational*: locate a specific web site.
- *Transactional*: locate and visit a specific web site for further use, e.g., for undertaking transactions such as purchasing a product.

Rose and Levinson [3] have introduced a further subclassification of the taxonomy to classify intent more precisely. In addition to defining *directed*, *undirected*, *advice*, *locate*, and *list* gradations for *informational* intentions, they have also introduced *download*, *entertainment*, *interact*, and *obtain* as subclasses for the *transactional* class.

In our work we focus on the retrieval of images. In a first step we have tested the validity of Broder's taxonomy as well as Rose and Levinson's refinements for image search [4][6] and found that they cannot be applied without adaptation. In this poster we present a novel and adapted classification scheme of user intentions in image retrieval. We discuss the employed methodology in a broader context and outline the process of adaptation and validation of the found classes. We then explain and discuss limitations and impact of the novel classification scheme.

Approach / Methodology

The ideas presented in this paper emerge from a broad framework for "User Intentions in Multimedia", whose main steps are depicted in Figure 1. It starts from 'Fundamental questions' such as: "What are the users' intentions when they search for multimedia items, e.g., images, on the Web?", "How can those intentions be captured?", "Are there patterns of multimedia search that apply to many users?", and "If so, how can such patterns be determined and experimentally tested?". Based on those questions, a set of 'plausible hypotheses' is produced and used as a starting point for a series of 'principled methods and experiments' to test them out. Such experiments generate valuable raw data, which can in turn be used to extract 'promising cues, insights and patterns' which will either confirm the hypotheses or lead to their refinement for subsequent investigation. Since this work is grounded on computer science and focused on ultimately improving the overall user experience when searching for multimedia data, the bottom row of the diagram in Figure 1 depicts how the patterns and insights acquired from the experiments described in this paper can be used to design and build 'clever tools' (e.g., an automatic way to select the best viewing option for *Flickr* results, described in [6]), which – after a series of meaningful refinements – can be used to 'add value' to the end user.

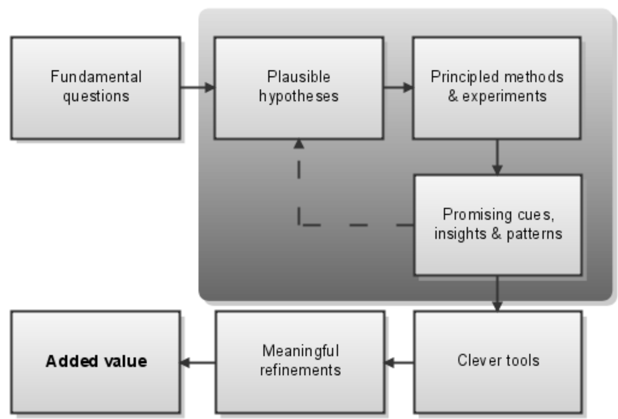

Figure 1. A framework for research on "User Intentions in Multimedia".

While Figure 1 describes the overall path of our research, in this poster we focus on iterations in the shaded area of Figure 1. Our hypothesis - "Broder's classification scheme is appropriate for image retrieval" – has been rejected based on our experiments [4].

We then adapted the taxonomy twice and tested the scheme again as described in the following section.

Classification Scheme

In [4][6] we presented experiments on the *Flickr* platform with Broder's taxonomy. Previous work [5] indicated that in text-based web search a large percentage of queries contained intentional structures. We wanted to analyze whether this is true for image search as well. We defined ten tasks, reflecting typical search problems for Broder's classes. We observed user behavior in a quantitative study and interviewed users in a qualitative study as well. Our test population consisted of 20 students and co-workers from Klagenfurt University, who had to solve the tasks on *Flickr* by searching and viewing their chosen result image(s). In general, we were not able to find intentional structures on the query level. However, we hypothesize that people express their intentions implicitly through *search behavior*: we found relationships between our task classification and the data generated. User intent can be measured in terms of *number of clicked result images* and *duration to solve a task*. To introduce an example of how intent can be used to add value in real-world applications, we developed a tool which automatically classifies the user's intent based on her search behavior and dynamically adapts *Flickr*'s result view [6]. The preliminary study led to the insight that Broder's taxonomy is not adequate to describe user intent for image search. Due to the text-based web search context of the taxonomy, not every goal in image search can be assigned to one of the classes. Furthermore, the semantics of some classes are not applicable in image retrieval. We defined an analysis- refinement-process which was iteratively performed to adapt the introduced taxonomy and to create a classification scheme for image search.

In a subsequent set of studies, we conducted interviews with nine people who use *Flickr* (and similar platforms) extensively. An architect, four photographers, a research assistant at Klagenfurt University, and three librarians at Florida Atlantic University have been interviewed. These interviews have led to new ideas, clarification of open problems, and served as an evaluation of the work completed so far. A detected problem was *class overlapping* – a task may belong to more than one class. Another challenge is to find the characteristics of search results which are more likely to indicate success or not. Successful completion of a task is possibly due to the interplay between text/concepts, (mental) image, (non-)visual analysis, and semantics of image and text. Several difficulties with the *informational* class – obtaining new knowledge by finding a set of images – have been identified, e.g., "Does the result of a search session consist of one, two, or more images?", "Is the obtained knowledge generated directly or indirectly through the image?", "Is the result an image or is it associated with an image?"

The *informational* class was renamed to *knowledge orientation,* the *navigational* class to *navigation*, the *transactional* class to *transaction*, and a new class named *mental image* was introduced. Moreover, we redefined the classes in a more detailed way including the *user's need prior to search,* the *user's need during search*, and the characteristics of search results *(based on)* explained above. Table 1 presents the classification scheme of intentions for user goals in image search, including an example task for each of the main classes. These classes, coded *ET1*, *ET2*, *ET3*, and *ET4* fall nicely into one (and only one) class in the Venn diagram presented in Figure 2.

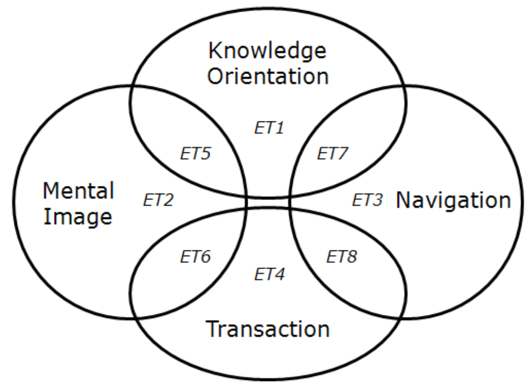

Figure 2. Venn diagram showing the classification scheme and the overlapping of classes.

In *ET1*, the user has to extract knowledge from his search results to list *three Indian beer brands*. In *ET2*, the user already has a mental image of how a *crop circle* looks like. In contrast, in *ET3*, the user knows about the existence of *mathias' juggling pictures*, but has no idea about their exact visual content, since they could have been taken anywhere within the university as well as from any possible point of view. In *ET4*, the user wants to find and *download* a picture showing *joy* to *add value* to her presentation.

It can be easily seen that there is no class overlap between the *mental image* and the *navigation class*. A search goal is either to find an image by contents or by semantics. Also the non-existence of overlap between the *knowledge orientation* and the *transaction* class can be explained easily: if a user has to add value or convey meaning, value and meaning have to be defined beforehand, and cannot be learned within the same task.

	Knowledge Orientation	Mental Image	Navigation	Transaction
User's need prior to search	The user wants to learn something by looking at (a) picture(s).	The user knows how the image content looks like.	The user knows about the existence of the image, but its content is unknown.	The user wants to find a picture for further use.
Based on	Text/Concepts	Mental image	Text/Concepts	Non-visual concepts
User's need during search	Extract knowledge from the picture(s) to answer a question.	Match the retrieved images to mental image.	Find images and topics, to restrict the number of possible result images.	Find image to convey meaning and/or to add value to something.
Based on	Image semantics and Text semantics	Visual analysis and Image semantics	Image semantics and Text semantics	Visual analysis and Image semantics
Example task	*"Find out the names of three different beer brands in India." (ET1)*	*"Find a typical photo of a crop circle." (ET2)*	"Look at the pictures of 'mathias' *juggling clubs at Klagenfurt University." (ET3)*	*"Find a picture expressing 'joy' for a presentation and download it." (ET4)*

Table 1. A Classification Scheme for User Intentions in Image Search including sample tasks.

If the user gains knowledge from downloading the images the knowledge gain would still not be part of the actual goal.

Figure 2 also shows the overlap of certain classes in the defined scheme which can be explained by the sample tasks *ET5* to *ET8*, described next. *ET5* – *"Find out the differences between the old and new Starbucks brand logo."* – is a combination of the *knowledge orientation* and *mental image* classes and therefore defined in the intersection between them. The user has already a perception of the *Starbucks logo* in mind, but wants to obtain additional knowledge as well. In *ET6* – *"You are searching for a background image for your cell phone. Find a picture showing a beautiful sunset and download it."* – the user has to *search* and *download* an image showing a *sunset*, which she probably already has in mind. In *ET7* – *"Your friend 'XY' bought a new car* *which you haven't seen yet. She took two photos of that car. Take a look at the photos."* – the user knows where to find the pictures, so the search task is *navigation*, but also knowledge about *the new car* is extracted by looking at the photos. In *ET8* – *"Find the latest photo published in group XY to use it as your desktop background"* – the user knows that there are pictures published within this user group and where to find them. While navigating towards the desired content the user also wants to download it.

Conclusions

We have presented a novel classification scheme for user intentions in image search. Our work is based on multiple user studies and interviews and shows interesting characteristics, the most notable being the definition and explanation of class overlap. The classification scheme is composed of two pairs of disjoint classes (see Figure 2) and therefore allows for implementation by the elegant use of two binary classifiers.

Current and ongoing work focuses on completing the surveys and user study as well as the interpretation of those. Furthermore, we will take a closer look at possible search behavior features for intent classification. In addition we will investigate the real life relevance of each of the tasks to get an idea of the overall practical value of the classification scheme for further refinements. We are also working on novel tools to support user interaction by view adaptation as well as adaptation of relevance functions in image retrieval. Moreover we aim to analyze whether *search behavior* is an adequate indicator of user's intentions as well as to investigate if there is a better research methodology to test intent expression besides the use of tasks.

Acknowledgements

We'd like to thank Markus Strohmaier for the insightful discussions on user intentions. This work was supported by the Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant 20214/17097/24774.

Citations

[1] A. Broder. A Taxonomy of Web Search, *SIGIR Forum*, 36(2):3-10, 2002.

[2] B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of Web queries. *Inf. Process. Manage.*, 44(3):1251-1266-2008.

[3] D. E. Rose and D. Levinson. Understanding user goals in web search. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 13–19, New York, NY, USA, 2004. ACM.

[4] C. Kofler and M. Lux. An Exploratory Study on the Explicitness of User Intentions in Digital Photo Retrieval. In *I-KNOW '09: Proceedings of the 9th International Conference on Knowledge Management and Knowledge Technologies*, 2009.

[5] M. Strohmaier, M. Lux, M. Granitzer, P. Scheir, S. Liaskos, and E. Yu. How Do Users Express Goals on the Web? – An Exploration of Intentional Structures in Web Search. In *We Know'07 International Workshop on Collaborative Knowledge Management for Web Information Systems, in conjunction with WISE'07, Nancy, France*, 2007.

[6] C. Kofler and M. Lux. Dynamic Presentation Adaptation based on User Intent Classification. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 1117–1118, New York, NY, USA, 2009. ACM.