Towards User-Centered Retrieval Algorithms

Manuel J. Fonseca
Department of Computer Science and Engineering
INESC-ID/IST/Technical University of Lisbon
R. Alves Redol, 9, 1000-029 Lisboa, Portugal
mjf@inesc-id.pt

ABSTRACT

Nowadays almost all retrieval algorithms (for text, images, drawings, etc.) are mainly concerned in achieving good system-centered measures, such as precision and recall. However, these systems are used by users, who try to achieve goals through the execution of tasks. To better satisfy the users' needs we must involve them in the development process of the retrieval systems.

In this paper, we argue that a user-centered approach, where users are included in the development cycle of the overall retrieval system, can lead to improved retrieval algorithms and also to a better user satisfaction while using the system.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - Graphical user interfaces (GUI)

General Terms

Design, Human Factors

Keywords

User-Centered Design, User-centered approach, Retrieval algorithms

1. INTRODUCTION

The majority of the retrieval algorithms, whether they are for text, images, drawings, 3D objects, audio, video, etc., are mainly interested in performing well for system-centered measures, like for instance precision and recall. However, these systems are used by users who want to perform specific tasks and achieve specific goals. We can develop a good retrieval system, that performs well against a predefined ground truth, but when we delivery it to users they may not be able to find what they want or they may not even be able to submit a query to the system.

For illustration purposes let us consider the following hypothetic scenario: "We developed a system for retrieving generic complex vector drawings, like for instance technical drawings, architectural plants or clipart drawings. We evaluated it using query-by-example and a set of predefined drawings, achieving a good precision and recall measure. Afterwards, when we delivered the system to users, we noticed that they were not able to use it, because they could not find the (first) drawing that they must use as query to find the desired drawing. Moreover, users do not want to search for the complete drawing, but only by a subpart of the drawing."

This scenario could be avoided if before we developed the retrieval system we asked users what were their needs, what did they want to perform on the system and how they want to do it. To collect all this information we need to apply a user-centered approach where users are involved in the development of the retrieval system and algorithms.

In this paper we defend an user-centered approach as a way to create better retrieval algorithms and improve the overall retrieval system. We start by shortly describe the user-centered approach and the iterative cycle used in the user interface design. In Section 3 we describe our application of the user-centered approach in the development of retrieval algorithms. Finally, we present some conclusions.

2. USER-CENTERED DESIGN

The user-centered design (UCD) is a design methodology, where the needs, skills and limitations of the users are taken into account during all stages of the development of the system. The key premise of the user-centered design is that the active involvement of the users in the development process as well as in the evaluation of the interactive products can lead to well-designed systems that best meet the desired usability goals. These systems will take advantage of users skills, will be relevant to their work and activities, and will help them rather than constrain their actions.

One of the principles from the UCD [4] states that we first need to identify who the users will be (profile, skills limitations, etc.) and what tasks they perform and/or wish to perform. The second principle mentions that the systems should be exposed to users in the early stages of development to collect feedback from them. Finally, the third principle is iterative design. The results and feedback from user testing should be used to fix and improve the system. The UCD assumes an iterative cycle with identification of the users' needs, design of the solution and evaluation, repeated as

EuroHCIR '11 Newcastle, UK
.

often as necessary, as depicted in Figure 1.

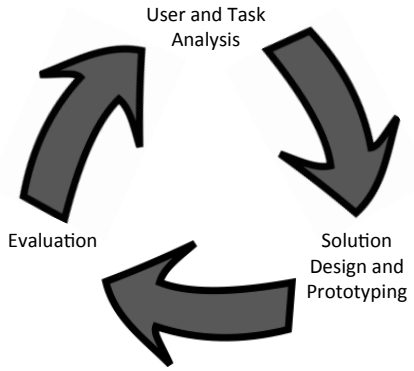

Figure 1: User-centered design iterative cycle.

3. USER-CENTERED RETRIEVAL

Typically when we want to develop a new retrieval approach, we look at the media to retrieve (text, audio, video, drawings, images, etc.), identify the features that better describe the media, create a matching algorithm and finally we compute precision and recall. Although this methodology allows us to create retrieval systems, we believe that by including the user in the development cycle will allow us to deliver better and more usable retrieval systems, that will allow users to achieve their goals and not only systems that have a good precision and recall performance.

Moreover, we should not develop retrieval systems, and that includes descriptor computation, matching algorithms and presentation of the results, without first identifying a set of user needs and functional requirements (first step in the user-centered design). We need to know our users, their skills, their background, their profile. We must identify their needs and requirements, their goals and how they achieve them. In summary, we need to do an user and task analysis before we start developing our retrieval system. User and task analysis should not only influence the design of the user interface, but also the design of the retrieval approach or algorithm.

For instance, users could use various strategies to perform a search in a drawing retrieval system. They could use a drawing that they already have, in a file, to search for similar drawings using query-by-example, or they could draw a sketch of the drawing that they want to find. As we can see, the retrieval solution (feature extraction, indexing and matching algorithms) will be different on each case. While in the first case we only need to compare two drawings of the same complexity and with the same characteristics (sets of lines and polygons), in the second case we need to compare complex drawings with sketches (typically simpler and with less elements). Thus, the way users perform the task to achieve their goal influence the retrieval approach that we should develop.

After developing the retrieval solution based on the user requirements, we should evaluate the retrieval system, using not only system-centered measures, but also user-centered measures, such as time to complete tasks, error rates, satisfaction, etc. As in the user-centered design of interactive systems, results from the evaluation of the retrieval system (system and user centered measures) should be used to improve the system and to refine the user and functional requirements of the retrieval system.

One of the things that we observed in one evaluation session with users, was that users did not care about where in the order of retrieval the intended drawing appears, the important fact being that it was there. One of the users produced this comment "It [the system] found it [the drawing]! That is what counts!" However, when we evaluate retrieval systems, the majority of the existing measures and ground truth datasets privilege precision. Of course this system-centered evaluation is important, but we should also take into account the users perspective, where they privilege recall.

3.1 An Example

Involving the users can affect the way we develop the retrieval algorithms. In recent years we developed a generic approach for complex vector drawing retrieval, based on the topology and geometry of the elements present in the drawing. These two features were used to describe the content of the drawings, and during matching, we first compare the drawings using topology and them we compare the geometry of those with similar topologies, giving the same weigh to both features (for more details see [1]). This generic retrieval approach was used to develop one system for retrieving technical drawings [3] and another for retrieving clipart drawings [2].

Before we developed this solution and the two retrieval systems, we performed user and task analysis to understand how users wanted to make queries to this type of systems. We notice that they prefer to draw sketches of the drawing that they were looking for than to submit an existing drawing to perform a query-by-example. Moreover, most of the times they do not have a drawing similar to the one that they are looking for.

The two systems were both evaluated with users, and from those evaluations we observed that the way users search for technical drawings was different from the way they search for clipart drawings [6]. While in the case of technical drawings users draw more complete sketches with several visual elements, and consequently defining a richer topological con-

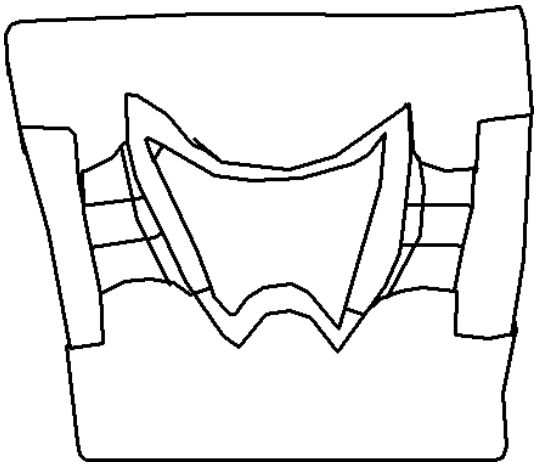

Figure 2: Sketch specifying a query to find a technical drawing.

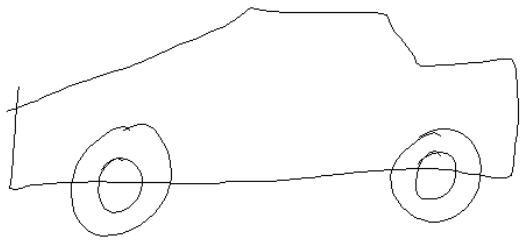

Figure 3: Sketch specifying a query to find a clipart drawing.

figuration, as illustrated in Figure 2; for clipart drawings, users produced simpler sketches, with fewer elements and with a poorer topological description (see Figure 3).

Due to this observation during tests with users, we refine our retrieval algorithm for retrieving clipart drawings [5], putting more emphasis on the geometry than on topology. With this change we were able to achieve a better precision and recall measure for clipart drawings, and we adapted our retrieval system to the users' way of sketching queries.

3.2 Discussion

We can not develop our retrieval algorithms without involving our users into the development cycle. As in the design of interactive systems, also in the development of retrieval systems we must involve the users.

They must be involved in the initial phase, so we can understand how they search for the information, what are their knowledge, what are their limitations and what is their profile. With this we are able to identify users needs and functional requirements.

Later on, during the development of the algorithms we should take into account this input and adapt the algorithms to provide "good results" for "our" users, and not for the users in general, or for the system.

Finally, during the evaluation stage, besides computing the traditional system-centered measures, for a set of datasets defined as ground truth, we should also involve users in the evaluation to collect quantitative and qualitative measures. Information gather during evaluation should be used to improve the retrieval algorithms and the overall retrieval system, in the next iteration of the iterative cycle of the user-centered approach.

4. CONCLUSIONS

In this paper we defended a user-centered approach for the development of retrieval systems. As in the case of user interfaces design, also for retrieval systems is important to know our users, adapt the algorithms to them, and involve the users in the evaluation of the system.

We believe, and we had confirmed, that the involvement of the user in the development cycle of retrieval systems can conduct to better systems that satisfy users needs and are more adapted to them.

5. ACKNOWLEDGMENTS

This work was supported by FCT through the PIDDAC Program funds (INESC-ID multiannual funding) and the Crush project, PTDC/EIA-EIA/108077/2008.

6. REFERENCES

[1] M. J. Fonseca. *Sketch-Based Retrieval in Large Sets of Drawings*. PhD thesis, Instituto Superior Técnico / Technical University of Lisbon, July 2004.

[2] M. J. Fonseca, B. Barroso, P. Ribeiro, and J. A. Jorge. Retrieving clipart images by content. In *Proceedings of the 3rd International Conference on Image and Video Retrieval (CIVR'04)*, volume 3115 of *Lecture Notes in Computer Science*, pages 500–507. Springer-Verlag, Dublin, Ireland, July 2004.

[3] M. J. Fonseca, A. Ferreira, and J. A. Jorge. Content-based retrieval of technical drawings. *International Journal of Computer Applications in Technology (IJCAT)*, 23(2–4):86–100, 2005.

[4] J. D. Gould and C. Lewis. Designing for usability: key principles and what designers think. *Commun. ACM*, 28(3):300–311, 1985.

[5] P. Sousa and M. J. Fonseca. Geometric matching for clip-art drawing retrieval. *Journal of Visual Communication and Image Representation (JVCI)*, 20(2):71–83, February 2009.

[6] P. Sousa and M. J. Fonseca. Sketch-based retrieval of drawings using spatial proximity. *Journal of Visual Languages and Computing (JVLC)*, 21(2):69–80, April 2010.