

# **A Qualitative Task Analysis of Biomedical Image Use and Retrieval**

William Hersh, MD<sup>1</sup>, Jeffery Jensen, BS<sup>1</sup>, Henning Müller, PhD<sup>2</sup>,

Patrick Ruch, PhD<sup>2</sup>, Paul Gorman, MD<sup>1</sup>

<sup>1</sup>Oregon Health & Science University, Portland, OR, USA

<sup>2</sup>University & Hospitals of Geneva, Geneva, Switzerland

Oregon Health & Science University

3181 SW Sam Jackson Park Rd., BICC

Portland, Oregon, USA

Phone: +1-503-494-4563

Fax: +1-503-494-4551

e-mail: [hersh@ohsu.edu](mailto:hersh@ohsu.edu)

## **Abstract**

Although increasing numbers and types of biomedical images are available electronically and used for diverse purposes, little is known about the actual uses of these images or how users search for them. One way to accelerate the development of image retrieval systems might be to develop test collections for use in an image retrieval system evaluation. To develop topics for a real-world test collection, we performed a qualitative task analysis of real users from faculty at an academic medical in their roles as researchers, clinicians, and educators. This will guide the development of the test collection as well as establish a foundation for further analysis of tasks for which image retrieval will be used.

## **1 Introduction**

Image retrieval is a poor stepchild to other forms of information retrieval (IR). Whereas a broad spectrum of Internet users, from laypeople to biomedical professionals, perform text searching routinely [1], fewer (though a growing number) search for images on a regular basis. While development of image retrieval approaches and systems began as a research field 20 years ago, progress has been stalled for multiple reasons. One problem is the inability of image processing algorithms to automatically identify the content of images in the manner that information retrieval and extraction systems have been able to do so with text [2].

A second problem is the lack of robust test collections and realistic query tasks that allow comparison of system performance [2, 3].

The lack of useful test collections is one of the motivations for the ImageCLEF initiative, which aims to build test collections for image retrieval research. ImageCLEF is a part of the Cross-Language Evaluation Forum (CLEF, [www.clef-campaign.org](http://www.clef-campaign.org)), a challenge evaluation for information retrieval from diverse languages [4]. CLEF itself is an outgrowth of the Text Retrieval Conference (TREC, [trec.nist.gov](http://trec.nist.gov)), a forum for evaluation of text retrieval IR systems. TREC and CLEF operate on an annual cycle of test collection development and distribution, followed by a conference where results are presented and analyzed.

The goals of TREC and CLEF are to build realistic test collections that simulate retrieval tasks and enable researchers to assess the performance of their systems and compare their results with others [5]. The goal of test collection construction is to assemble a large collection of *content* (documents, images, etc.) that resemble collections used in the real world. Builders of test collections also seek a sample of realistic *tasks* to serve as *topics* that can be submitted to systems as *queries* to retrieve content. The final component of test collections is relevance judgments that determine which content is relevant to each topic. A major challenge for test collections is to develop a set of realistic topics that can be judged for relevance to the retrieved items. In the TREC Genomics Track, a part of TREC devoted to evaluation document retrieval in the genomics domain, topics were developed based on interviews with real-world biomedical researchers [6]. This provided a better fidelity for the retrieval task and serves as a basis for the task analysis performed in this work.

Test collections are imperfect, reflecting a trade-off of real-world system use versus experimental control [7]. In order for the results of experiments to be reproducible, these collections must be static, despite the fact that real-world databases are not. There are also challenges in obtaining content, since many holders of content (e.g., health systems and publishers) may be reluctant to allow use of content for various reasons (e.g., privacy and economic concerns respectively). Even with good test collections, experiments using them are typically done in “batch” mode, simulating (to various degrees) real users. Finally, another issue is that relevance judges may not agree, with typical values of kappa indicating only “fair” (e.g., 0.4-0.6) levels of agreement [6, 8].

Image retrieval test collections have additional challenges. One reflects the nature of the image retrieval task, in that some systems are designed to handle visual (image-based) queries, while others are focused on semantic (text-based) queries [9, 10]. The former queries are challenging due to limits in the ability of systems to recognize objects in images, while the latter are stymied by the tendency of image interpreters to incorporate varying amounts of inference in their assessments, often describing the same feature along a finding-diagnosis continuum [11].

In the ImageCLEF 2004 medical image retrieval task, the retrieval task consisted of finding medical images similar to an index image with respect to anatomic region, modality, view direction, and radiological protocol [12]. This, of course, is only one type of task for which

an image retrieval system might be used. What other tasks image retrieval systems might be used for has not been studied systematically. While a body of informatics research has focused on the needs of various users [13, 14], few studies have looked at why or how people search for images [15], especially in the biomedical domain. The goal of the work described in this paper is to begin to understand the ways which different biomedical users in their various roles (e.g., clinicians, researchers, educators, students, librarians) use images and how they would search for them. An additional motivation for this work is to inform topic development for the ImageCLEF medical test collection in 2005 and beyond. The creation of realistic test collections will improve the ability of researchers and developers to build image retrieval systems.

## 2 Methods

We undertook a qualitative assessment of image use and retrieval tasks from potential real-world users. A qualitative approach was used because we wanted to develop a general categorization of image retrieval tasks that would enable us to develop topics for the ImageCLEF test collection.

### 2.1 Subjects

The study included a convenience sample of 13 subjects representing a spectrum of biomedical professional roles including clinician, researcher, educator, librarian, and student. Subjects were recruited from the staff of an academic medical center through personal contacts and referral by other subjects.

### 2.2 Data Collection

For each role the interviewee had, we used a semi-structured interview consisting of a series of questions related to how they used and retrieved images. A preliminary interview schedule was developed and a series of pilot interviews was conducted to test this instrument. Based on the results of these pilot interviews, the final, refined interview schedule was developed, as shown in Figure 1.

In your role as a                      Clinician              Researcher              Educator              Librarian, please answer each of the following: 1. Tell me the tasks that you perform in your work where images are helpful. 2. For each of these tasks, give me an example of the kind of image you would try to find. 3. When performing each of these tasks, where do you look for the images? 4. When looking for these images, how do you look for them? 5. When you find images, how do you decide whether one (or more) are suitable for your needs?
---

Figure 1 - Role-based survey about image use.

### **2.3 Data Analysis**

Content analysis was performed on data from semistructured interviews. Two investigators (WH, PG) independently reviewed the data. Subject responses were sorted into categories, and the results were combined by consensus.

## **3 Results**

A single interviewer (JJ) conducted the interviews with fourteen subjects. Several subjects listed two roles, such as clinician and educator, giving a total of 19 roles for 13 subjects. The distribution of specific roles is shown in Table 1. The transcripts of the interviews were independently analyzed by PG and WH to collapse and extract categories and data.

In general, image supported tasks fell into three broad groups: research-specific images for data analysis and presentation; patient-specific images for clinical decisions and management; and mainly generic diagrams for explanation and education. To a large extent, reported image seeking behavior corresponded to these categories. Table 2 lists the image-supported tasks described by the 13 subjects.

## **4 Conclusions**

This preliminary work gives us some insight into how various medical personnel use images in their different professional roles. It also establishes a foundation for how image retrieval systems might be built to support users in finding these types of images. As the ImageCLEF image retrieval initiative grows, we will aim to develop test collections that take these uses into account.

We plan to pursue further work in elucidating how researchers, clinicians, and educators use and seek images. We will extend the work described in this paper to a larger scale and more quantitatively oriented analysis. We also believe that the test collection itself will motivate the development of further image retrieval systems and approaches, leading to wider use and in turn getting users to think of other tasks for which they may want to use image retrieval systems.

## **Acknowledgements**

This work was supported by a supplement to National Science Foundation (NSF) grant ITR-0325160. We also acknowledge the European Commission IST projects program in facilitating this work.

Table 1 - Distribution of roles among subjects interviewed.

Role (number)	Researcher	Clinician	Educator	Librarian	Student
Researcher and educator (1)	1		1		
Clinician and educator (4)		4	4		
Clinician and researcher (1)	1	1			
Researcher (3)	3				
Clinician (1)		1			
Educator (1)			1		
Librarian (1)				1	
Student (1)					1
Total (13)	5	6	6	1	1

Table 2 - Image-supported tasks extracted from analysis of surveys.

Research-related	
Images as data: analysis and interpretation of images (photomicrographs, magnetic resonance imaging, etc.)	
Images for presentation and publication of research findings to research audience	
Patient care-related	
Check image test result in electronic health record	
Diagnosis of uncommon or unrecognized condition	
Illustration and explanation to patient	
Education-related	
Educational presentations to students, etc. (listed for all roles)	
Learning – librarian support of others' learning; clinician self-education on new technique	
Other	
General Information	Educators
Expert witness testimony	Clinician
Developing collections	Educators
Marketing (before and after)	Clinician (plastic surgery)

## References

1. Madden M and Rainie L, *America's Online Pursuits: The changing picture of who's online and what they do*. 2003, Pew Internet & American Life Project: Washington, DC.
2. Müller H, Michoux N, Bandon D, and Geissbuhler A, *A review of content-based image retrieval systems in medical applications-clinical benefits and future directions*. International Journal of Medical Informatics, 2004. 73: 1-23.

3. Horsch A, Prinz M, Schneider S, Sipilä O, Spinnler K., Vallée JP, et al., *Establishing an international reference image database for research and development in medical image processing*. *Methods of Information in Medicine*, 2004. 43: 409-412.
4. Braschler M and Peters C, *Cross-language evaluation forum: objectives, results, achievements*. *Information Retrieval*, 2004. 7: 7-31.
5. Sparck-Jones K, *Reflections on TREC*. *Information Processing and Management*, 1995. 31: 291-314.
6. Hersh W, Bhuptiraju RT, Ross L, Johnson P, Cohen AM, and Kraemer DF. *TREC 2004 genomics track overview. The Thirteenth Text Retrieval Conference (TREC 2004)*. 2004. Gaithersburg, MD: NIST. in press.
7. Hersh WR, *Relevance and retrieval evaluation: perspectives from medicine*. *Journal of the American Society for Information Science*, 1994. 45: 201-206.
8. Hersh WR, Buckley C, Leone TJ, and Hickam DH. *OHSUMED: an interactive retrieval evaluation and new large test collection for research. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1994. Dublin, Ireland: Springer-Verlag. 192-201.
9. Tagare HD, Jaffe CC, and Duncan J, *Medical image databases: a content-based retrieval approach*. *Journal of the American Medical Informatics Association*, 1997. 4: 184-198.
10. Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, et al., *Content-based image retrieval in medical applications*. *Methods of Information in Medicine*, 2004. 43: 354-361.
11. Greenes RA, McClure RC, Pattison-Gordon E, and Sato L. *The findings-diagnosis continuum: implications for image descriptions and clinical databases. Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care*. 1992. Baltimore, MD: McGraw-Hill. 383-387.
12. Clough P, Sanderson M, and Müller H, eds. *Overview of the CLEF cross language image retrieval track (ImageCLEF) 2004*. *Multilingual Information Access for Text, Speech, and Images: Result of the Fifth CLEF Evaluation Campaign, Lecture Notes in Computer Science*, ed. Peters C, et al. 2005, Springer-Verlag: Heidelberg, Germany.
13. Gorman PN, *Information needs of physicians*. *Journal of the American Society for Information Science*, 1995. 46: 729-736.
14. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al., *Analysis of questions asked by family doctors regarding patient care*. *British Medical Journal*, 1999. 319: 358-361.
15. Enser PGB, *Pictorial information retrieval*. *Journal of Documentation*, 1995. 51: 126-170.