A Survey on Visual Information Search Behavior and Requirements of Radiologists

D. Markonis; M. Holzer; S. Dungs; A. Vargas; G. Langs; S. Kriewel; H. Müller

1. Introduction

Image retrieval has been an active research domain for the past 25 years [9] but so far there have been few commercial applications of image search using visual information such as LookThatUp\textsuperscript{*} or since mid 2011 Google images. The medical domain has often been mentioned as an important domain for content-based image retrieval (CBIR) [30], but again, only a few applications have been evaluated in clinical routine [1, 5]. The Khresmoi project\textsuperscript{b} aims to develop a multi-lingual, multi-modal search and access system for biomedical information and documents. One of the target user groups in Khresmoi are radiologists and thus understanding the image search behavior and needs of radiologists is vital for designing such a system.

To develop applications and products based on real user needs has been a standard procedure in many other fields [22] including medical information retrieval [10, 29] and clinical decision support [13]. As is stressed in [12], “... by asking from the perspective of the user, 'what should a successful system do', relevant variables can be identified”. For CBIR systems such user studies were initially only rarely performed. Approaches were rather technology driven in terms of applications than based on real user requirements. On a wider scale, many studies exist on information needs and the use of information technology in the medical field [16]. In [15], a systematic framework for evaluating the use of medical IR systems is proposed.

A theoretical analysis on information needs is attempted in [33]. Moreover, a model for information seeking behavior is proposed using four categories. Categories a and b include search strategies by the user that do or do not depend an IR system. Categories c and d include the search strategies employed by the IR systems to satisfy the user’s demands for information. Our study focuses on the first two categories in the field of radiology, while the two latter will be the subject of user testing and system evaluation.

One of the first user analyzes for image retrieval was [21], on the behavior of jour-
A variety of methods exist for obtaining information on system use:

- observation of the behavior of users (which can include direct observation [17, 21] or the analysis of log files that record the behavior of users [6, 11, 20, 21, 23, 34]);
- interviews with stakeholders [7, 14, 17, 21];
- surveys [22, 24, 34].

While observation can provide useful insights on the behavior of the subjects, it often lacks the provision of clear and in-depth information about important concepts such as user satisfaction, unmet needs and desired functionalities. On the other hand, interviews assist in obtaining such information but mainly on a qualitative level as such interviews are usually time-consuming. Finally, structured surveys with stakeholders can give quantitative results, but need careful design, as questions should be easy to understand and on target. They need to be performed on a relatively large scale in order to have statistical significance. As the role of these methods is complementary all of them were applied in this study.

Direct observation provides comprehensive information about how radiologists use images during their clinical work. However, the speak-aloud-process that is usually used may deviate from the natural, unobstructed clinical workflow. A way to monitor image use behavior without interfering with the task is observation using table-mounted eye–tracking equipment. Eye gaze tracking has been widely used in user interface design and evaluation [2, 4, 8, 17] but also in the radiology field, eye–tracking analysis was used as input for the design of workstations [2, 3]. Moreover, studies of eye movement tracking were used to analyze image interpretation and decision making of radiologists [3, 18, 19, 31]. However, most of these studies either concentrated on specific anatomy locations or on certain modalities.

2. Methods

This section describes the techniques that were used for obtaining information on the image use and search behavior of radiologists. All tasks were performed in the radiology departments of the Vienna University hospitals, Austria and the University hospitals of Geneva, Switzerland, two large teaching hospitals.

2.1 Observation

In order to learn more about information behavior of radiologists, watching them perform standard tasks and then analyzing information needs at specific moments was the first step. This information was then used to give better insight for the creation of the questions for the survey and to complement the data from the survey. To obtain information on the workflow three steps were taken:

1. Listen to experienced radiologists describing the main steps of the radiology image analysis process.
2. Case discussions were followed where interesting cases are explained including the reasoning process, imaging data required, and evidence provided by several exams.
3. Eye tracking experiments were performed, where radiologists diagnose cases while being eye–tracked.

The setup of the eye tracking consisted of a stand-alone workstation at the University hospital of Vienna, which was not connected to the Picture Archival and Communication System (PACS) network. There was one workstation PC connected to one 23” LCD monitor.

2.2 Interviews

After a first survey form was constructed using the observation results and the literature review, several structured interviews with the draft survey form were performed. The goal was to learn whether the questions were understandable and whether responses correspond to the study’s target of interest.

Three detailed interviews were performed in Geneva with successive versions of the survey form, where a clinician filled in the form explaining aloud how the questions were understood and why a particular answer was given. Each time the form was adapted based on the comments of the previous interview. In Vienna, two rounds of structured interviews with the survey forms were performed and the form adapted accordingly.

2.3 Survey

Starting point for the survey questions was a user study previously performed in Portland, Oregon, USA, and then later in Geneva, Switzerland [24]. Based on the questions in this survey a form was adapted to comments from local radiologists to correspond to the specific group of radiologists instead of clinicians in general as the first surveys did. Three main tasks were identified to evaluate the specific needs: clinical work on patients, work regarding teaching, as in the preparation of lectures, and research work.

Besides the search requirements, basic demographic data on the radiologists was
acquired to better interpret results. The final version of the survey consisted of four sections: general data, clinical work, teaching, and research.

A common set of questions was used for the three activity domains. The first part is focused on the current image search behavior of radiologists. In the second part the participants were asked to propose services and tools useful for their search, and imagine a perfect image search system for their needs. A detailed list of questions is included in [26].

3. Results

This section describes the results obtained in the study.

3.1 Observation

The radiology workflow usually starts with opening a case for which an imaging exam had been requested. The images are transferred from the PACS server to the local viewing workstation and then the viewing process starts. The viewing options are set depending on requirements such as size and number of views per screen. The setup depends on the imaging modality and on the radiologist’s preferences and can be changed during the analysis process. Before starting to analyze the images, the patient’s medical history and anamnesis are reviewed. The radiologist then analyzes the images by adjusting the brightness/contrast and scrolling through the slices. The sets of images can be changed using thumbnail previews. Tools for measuring sizes are available for specific organs and pathologies. Once a pathology or abnormality is found there are two possibilities:

- **The abnormality is known:** potential diagnosis and differential diagnosis are given and the medical finding is described.
- **The abnormality is unknown:** search for additional information is needed.

A common way to handle unknown abnormalities is to ask an experienced colleague for help. This sometimes ends with a group discussion about possible pathologies, corresponding to ”information exchange” as in [26]. Often, the radiologist has to search through the literature (Internet, books, scientific articles). For this, the pathology needs to be described as well as possible. With the potential diagnosis the medical finding is completed. If, from the scientific or teaching point of view, the study is interesting for the radiologist it is often being marked for future reference.

In case discussions, the workflow is similar, starting with the anamnesis and history of the patient. One of the important aspects when presenting cases is sharing the experience between radiologists, particularly for rare cases. The main steps in the diagnosis process are the comparison of findings with the state of the art in the literature and the request of additional exams (imaging, laboratory, etc.) to assure that the probability of a correct interpretation is high. This means that access to and knowledge of the literature is important in practicing evidence-based medicine also in radiology. Justifying decisions is important and links to related cases are essential.

Other important aspects mentioned in the seminars are the temporal nature of images, for example comparing images of the patient over time. Computer-aided detection, such as highlighting particular abnormal regions or visualizing results, is also used.

For the eye tracking, the system was calibrated for each participant. When calibration was finished successfully, the participants were performing their image viewing and analysis tasks described above. The study included 3 sessions, each with a different radiologist. Two persons were working on the same studies (head CTs, mammography, chest x-rays) and the third one on a knee MRI. The task was to perform the usual analysis for diagnosis. The radiologists were explaining their tasks while they were performing the actions and the results of the experiments were visualized among others as heat map images (Fig. 1).

3.2 Interviews

The main outcomes of the structured interviews to adapt the forms were:

- radiologists are not familiar with visual retrieval (search using visual characteristics of an image) and examples for expected answers should be given to facilitate answering the questions;
- when an interesting case is found, many persons store images locally on their computers in a simple folder structure for future use; this has to be taken into account (although this is not a desired practice in many hospitals where data acquisition needs to be validated by an ethics committee); current hospital information systems often do not provide this functionality or need an additional...
effort for anonymizing the case and making it accessible in the desired format;
● radiologists found it hard to separate between the three proposed tasks of teaching, research and clinical work, mixing them when filling the survey, sometimes mentioning the overlap between them;
● many formulations were modified as clinicians referred to them as computer science jargon that could be hard to understand.

The results of the structured interviews were used for modifying the forms.

3.3 Survey

This section describes the outcomes of a survey to which 34 radiologists responded. 10 persons filled in the paper form after a seminar at the University hospitals of Geneva, whereas all other participants filled in the electronic form.

3.3.1 Demographics

As many questionnaires were filled in seminars, about half of the population is under 30 years. The other half is evenly distributed between 30 and 55 years. Two thirds are male and one third is female. Most persons had their radiology education in either Austria or Switzerland, which was expected from performing the survey mainly in Vienna and Geneva. All other participants come from Western Europe or the US, meaning that education is comparable.

In terms of radiology specializations, 23 persons specialize in general radiology, 1–2 persons each in musculoskeletal radiology, thoracic radiology, radiology informatics, neuroradiology, orthopedic radiology and body imaging. One person mentioned to specialize in CT imaging and another one is still a student. Twenty-eight persons work in public hospitals and two in private clinics, with four persons mentioning to mainly work in research at the University. The rather junior sample is seen in the years of experience, with eleven persons having less than two years of radiology experience. Otherwise, the distribution is relatively even between 3 and over 20 years experience.

For work time distribution it was possible to weight the time spend on clinical work, teaching and research on a scale from 0 to 5. Most persons perform all three activities; few have no teaching or no research but all have clinical work.

3.3.2 Clinics – Teaching – Research

All 34 participants mention tasks where they search for images other than of the person being diagnosed. The main reasons to search for images are finding material for presentations (mentioned by 8), differential diagnosis during a medical finding for difficult cases or in case of an unclear pathology (mentioned by 13) or performing clinical research (mentioned by 3). Specific examples listed are lung fibrosis, brain or bone tumors or lesions in brain, liver or other structures. Another task mentioned where images can be useful is the grading of a disease.

When teaching work is involved, the main focus of clinicians is to find similar cases. Depending on the class they are teaching, they look for easy, advanced or tricky cases. The image type depends on the current topic and ranges from plain x-rays, CT scans to typical pathologies such as primary brain tumors or lesions. It also includes differential diagnosis. Links with images of the scientific literature were also mentioned as useful.

While performing the search, the most frequently used source is the Internet using keywords (Fig. 2a) mentioned by 14. Google (5) is used as well as public medical databases (PubMed, Goldminer, e-anatomy, Eurorad were mentioned each by 1–3) and Wikipedia (mentioned by 2). Most of the clinicians also have personal files stored on PCs to search images (sometimes with keywords), as mentioned by 12. The local patient record is queried using the patient
When an image is found, it needs to be decided whether or not it is useful based on experience and comparison with a reference case. This corresponds to the notion of relevance in information retrieval. The correct image properties (e.g., modality, contrast, patient age/gender, mentioned by 7) as well as the quality of the images and the reliability of the sources define suitability of the found images. The availability of a detailed description or comments on the image has an influence as well. Asking colleagues for their opinion is another option. Self-assessment of the success rate of searches, in clinical work the importance of image search in the workflow is not always documented or available. When comparing search times of successful image search, it becomes clear that the topic or pathology is rare, too new and sometimes too general. It needs to be noted that not all hospital information systems are fully searchable (e.g., scanned reports). Time pressure has a negative impact on finding relevant images as well. The room for improvement with optimized search tools to find relevant information quickly.

On average, the clinicians have a 75% success rate searching for images (based on self-assessment). This can be an overestimation as people might not be aware of all available data and potentially more relevant items were not found. In Fig. 3a the percentage of persons with a success rate below 40% is low. When comparing teaching and clinical work it becomes clear that clinical work has a higher risk of failed search as all persons mentioning success rates below 20% are searching for clinical work. This may highlight that clinical work is less well defined and has harder search tasks than for example teaching. In terms of success rate of searches, in clinical work the time taken for searches is found to be lower than in teaching, which may also explain the higher failure rate.

The clinicians think that most of the time the desired images are available but cannot be found. Few think that the search images are really not available. The main reason for not finding a relevant image is that the topic or pathology is rare, too new and sometimes too general. It needs to be noted that not all hospital information systems are fully searchable (e.g., scanned reports). Time pressure has a negative impact on finding relevant images as well. The room for improvement with optimized search tools to find relevant information quickly.

When comparing search times of successful image search, it becomes clear that over 70% of successful searches finish after ten minutes or less (Fig. 5a). Only a few persons search longer to successfully find images. For unsuccessful image search (Fig. 5b) a few persons already stop after 5 minutes or less, but most often 10 minutes or even over 15 minutes are mentioned before stopping. This highlights the importance of image search in the workflow and the room for improvement with optimized search tools to find relevant information quickly.

Figure 3 compares the responses for clinical work and teaching but both categories lead to similar results.

name or ID, which are sometimes stored on PDAs or local PCs regarding interesting or typical cases (mentioned by 9). Other options for finding information is looking at books (8) and asking colleagues (4). There is no significant difference between clinical and teaching activities. Clinicians focusing on teaching seem to have larger and more organized personal databases. Generally keyword search and not CBIR systems are mentioned.
is rather short and considerably below 10 minutes. For research only few responses were obtained but for teaching the time to successfully find images is much longer and many persons search for 10 minutes or more and still find relevant images.

For unsuccessful search, times for re-search are often more than 15 minutes. During clinical activities, the average time before stopping the search is significantly shorter\(^6\), between 5 and 10 minutes, probably due to time pressure. This difference balances the overall failed search time distribution. Since about one fourth of all searches are unsuccessful it is clear that having tools to find out more easily whether or not relevant images even exist can be very helpful and reduce the amount of time lost.

Twenty-three of 34 persons responded to the question on potentially useful additions for image search, sometimes in several categories such as teaching, research and clinical work with slight differences between the categories. The functionalities most often suggested are search by pathology (23) and modality (19), followed by search for similar images (17) and patient demography (8 times). Apart from the predefined options, the radiologists mentioned the need for multilingual retrieval and proposed other additions such as pathology or symptom classification (for example, using an ontology), query by text and images together, and semantic retrieval based on image characteristics. The search for reconstructed 3D images was also mentioned as was the need to connect radiology images with histopathology or other criteria allowing judging the confidence of a diagnosis.

Nineteen responses were obtained for desired input possibilities and 22 for desired result formats. As there were no major differences between clinical work/teaching/research, the three are combined. The perfect image search system should use images (with possible regions of interest) as well as keywords as input (Fig. 7a). Keywords could vary strongly, from describing the anatomical structure, the pathology and histology, up to more demographic information like patient age.

The output should include image examples and a corresponding description (Fig. 7b). If available, differential diagnosis could be provided by the search engine. More detailed information including references would be helpful. A few people mentioned that information supporting the diagnosis such as a biopsy would be useful to raise the level of trust in the information supplied.

There was a relatively small amount of feedback regarding the exploitation of visual information for information search, indicat-

\(^6\) The significance of this difference was verified by performing a t-test on the two distributions, resulting in \( t_{\text{diff}} = 6.32 \), which is over the limit value \((3.55)\) for \( df = 40 \) and \( p = 0.001 \).
ing that the radiologists are not familiar with the concept, the state of the technology and research. However, there were interesting suggestions: the search for similar images and similar cases (mentioned by 5), search for similar regions of interest (3) and the possibility to search for similar images and have social judgments of other radiologists on the similarity (mentioned by two persons), also to increase potential trust into a diagnosis. This can be in a way similar to web platforms for image sharing such as MedTing\(^4\). The importance of not only visual information but the connection with other patient data was mentioned three times, full text search also three times. The possibility to have statistics on the diagnoses for similar images was mentioned once.

In total, 24 radiologists responded to the question about possible goals of automatic annotation. Anatomic region is mentioned 24 times as being important and modality 11 times. For research and teaching the modality is mentioned more often than for clinical work but otherwise differences between categories are small. Another annotation target mentioned was the quantification of the size of structures (6 persons). It was also mentioned that all the extracted information should be made available as free text for image search. Few radiologists (mostly the more experienced ones) mention to use systematic terminologies for image search or image descriptions. Of 20 persons who responded to the question, 7 mention to not use any terminology. Medical Subject Headings (MeSH) was mentioned most frequently (10). There is no major difference in terminology use between clinical work, teaching and research. Five par-

\(^4\) http://www.medting.com/

Fig. 6 Time taken for (a) successful and (b) failed image search compared for clinical work, teaching and research.

Fig. 7 Desired a) input and b) output data for a medical image search.
Participants mention to use RadLex and SNOMED CT was mentioned once. The dominant PACS system mentioned is AFGA IMPAX (8), followed by Fuji Syn- ipsis (3) with one participant each answering GE, McKesson, CareStream, PAC/RIS and Siemens Syngo. One participant mentioned to be using several systems.

3.3.3 The Perfect System

As the questions on how visual information can be exploited and how the perfect system should look like are of great importance for the design of medical image retrieval systems, we analyze the responses with a focus on the text given by the respondents. It is clear that imagining the perfect search system is hard when no example system is known. Still, besides the current image use and search behavior several persons added comments about the perfect search system. Most comments are included in their raw form in [26].

The results show that perfect search systems are more concerned with structured data than they are with visual data, although several people mention search for similar cases or images. On the one hand, this could indicate the lack of knowledge about current research prototypes for visual retrieval, but on the other hand it highlights the importance of structured clinical data in the process. This means that even for visual retrieval combinations with clinical data are essential. Some text search systems such as Goldminer and Yottalook are known to the clinicians and used by them.

It can be important to base a new search tool on what is known and used at the moment and then add functionalities such as visual similarity search. This is not problematic since it includes the continuation of design problems of current systems but it can avoid a rejection by the clinicians. Social interactions with others and comments of others are also mentioned as important for the search.

4. Discussion

While the introduction of filmless imaging and PACS has improved the performance and efficiency in radiology [22, 27, 28], many problems still exist in the current situation in terms of information retrieval. Describing the pathology correctly can be a difficult task but is essential for retrieving usable results. Visual retrieval is not possible with the clinical systems currently in use. Books are not instantly available for all pathologies and sometimes have to be found in the library. Searching in books is easy when ideas about pathologies exist. The Internet is a commonly used resource. Besides standard medical databases (such as MedLine), search engines (e.g. Google) are primarily used for the search. Search engines are powerful but not designed for medical purposes and often return undesired and low quality results.

When radiologists analyze an image substantial differences in viewing behavior between image types can be found. In some cases single areas of high concentration of fixation can be determined while other image types show broad scanning paths in viewing behavior (Fig. 1). About 30 minutes of the process were recorded. When discussing the data it became apparent that it might be better to clearly select a number of reference cases for such a study and then compare this between radiologists. In general, a larger number of cases might have lead to more conclusive results. Further studies are foreseen within Khresmoi to observe actual image search behavior once the first prototypes are available.

Administrative difficulties such as the anonymization of videos when viewing patient data were found during this test session and these will help to improve the recordings for the next eye tracking session:
- tracking two screens at the same time lead to bandwidth problems with the eye tracking system;
- videos recorded had to be anonymized as well before they could be analyzed, so all patient names were removed using a low pass filter; in future (large scale) eye-tracking tests the anonymization process needs to be automated.

The analysis of the observations was mainly used for the construction of the survey. The eye tracking confirmed the belief that very small regions of interest are what clinicians really focus most of their activity on and thus search by regions constitutes one of the important requirements for future medical CBIR systems. These findings were confirmed by the interest in searching by ROI expressed in the survey.

Obtaining a very large number of responses from often busy and overloaded radiologists was difficult. Thirty-four radiologists responded to the questions asked within the three months of the online and paper survey. Most persons responding were junior (below 30 years) and with less than five years of experience. This has the advantage of having persons who grew up using the Internet and digital image handling but the inconvenience that they might not question current practices and might have had fewer situations where they were lacking crucial information in clinical work. The current Internet generation is also plagued by the problem that they often believe to be competent information searchers, which does not always correspond to reality [32] – particularly in terms of how to use the information found.

Here, the most important aspects are listed to give a complete picture.
- **Role of image search:** The search for images and similar cases is an essential part of the radiology workflow. During the assessment of clinical data they use information from other images obtained from multiple sources: reference books, communication with other radiologists, personal files, the hospital database, and the Internet (both specialized databases such as PubMed, or general search engines such as Google). Radiologists allocate a significant amount of time to searching but fail in a substantial number of cases (around 25%).
- **How and what to search for:** Keyword search is currently the dominant search type, including Internet search, and access by patient ID in clinical records, or the oral communication with colleagues. During result selection, experience plays a dominant role when analyzing and choosing images. This indicates that substantial prior knowledge is necessary to perform efficient and successful search. Communication among colleagues is used to share knowledge not only during training but also in clinical practice. Past cases store experi-
ence of other colleagues and can make this experience available in a systematic way. Trust in the information found and evidence for a particular diagnosis are important. This can be more easily confirmed in communication with colleagues than when searching other sources such as the Internet. The scientific literature has an advantage over general Internet sources. Visual retrieval is little known although first prototypes exist such as IRMA (Image Retrieval in Medical Applications) and MedGIFT (Medical GNU Image Finding Tool).

- **Limits of current search:** There is clearly room for improvement considering the allocated time and the success rate of current image search. This is consistent with the perception of radiologists who conclude that the obstacle for finding images is not the availability but the limits of search technology or novelty of the data. Keyword search is perceived to have limits as an accurate prior assessment of the present case is required before formulating a query. A tedious selection of results based on individual inspection of potentially ambiguous candidate images is then necessary. This is limiting in the case of rare diseases, where search and comparison with other examples might be most relevant but little prior assessment is feasible. A related limitation is the lack of comprehensive keyword assignments in reference databases. Keywords are ambiguous and only using terminologies can help in this respect. Many radiologists build their own personal reference databases to compensate for searchability. In many institutions the lack of institutional archives requires this. Another way is storing patient IDs of interesting cases with short textual annotations in files that allow finding cases. The dominant role of experience, the emergence of scattered personal reference databases and the culture of communication among colleagues suggest that facilitating the sharing of knowledge and removing requirements of prior assessment and keyword identification can have significant impact on the radiologists’ clinical work, teaching and research.

- **Wishes for future systems:** Suggestions for future search are consistent with limits of current search. Radiologists name search for pathology as a goal. However, they value images equally to keywords as potential query inputs and suggest the use of ROIs to obtain more specific search results. Other functionalities include limiting the search to modalities and to include textual data into the visual retrieval, which could be achieved by faceted filtering of the search results. The idea of trust or confidence in a diagnosis of a case was also mentioned to be important. An important aspect is to link search results and cases with the literature. The peer-reviewed literature offers a level of trust.

### 5. Conclusions

The results of this survey are another step to better understand the requirements of radiologists in handling images and searching for visual data that can help them in daily tasks. Some of the results confirm previous studies highlighting the dominant role of image search in radiology the lack of use of visual retrieval [20] or the interest in searching by pathology, anatomy and modality [21]. Others, shed light into different aspects of this topic by giving a more detailed analysis of the notion of trust in the retrieved results and a more in-depth analysis of the reasons that the search for images often fails with the existing tools. The following list contains the most important results of this study:

- Radiologists often search for images or cases containing images and fail around one out of four times; search for visually similar images or cases with visually similar images is a desired functionality.
- The main reasons for this failure are the lack of time, the novelty or rareress of the search target and the shortcomings of the current search technology.
- The level of trust, the image quality and other colleagues’ opinions are important for the decision for potential relevance of the results.
- Time available for search and the types of searches are different between teaching, research and clinical work.
- A perfect search system would allow searching visually for regions of interest and would combine visual and structured or free text information.
- Very few radiologists currently use terminologies for annotating their cases or regions of interest.
- Structured data need to accompany images as desired output for search systems as images out of context are often of little help.

As many radiologists are not familiar with visual retrieval, being able to show them prototypes and having them work with the prototypes will most likely help them understand problems and potential and will make it easier to formulate desires for a perfect search system. Such tests can be analyzed with eye tracking to see how actual systems are used in practice by radiologists. After the first user tests in the Khresmoi project, a similar survey among the participants will take place. This might provide more concrete feedback regarding the existing system and planning the next steps to reach an impact of the technologies in radiology image search.

### Acknowledgments

This research has received funding from the European Union under grant agreement number 257528 (KHRESMOI).

### References


