A survey on visual information search behavior and requirements of radiologists

D. Markonis
HES–SO, Switzerland

M. Holzer,
Medical University of Vienna, Austria

S. Dungs,
University Duisburg–Essen, Germany

A. Vargas,
HON, Switzerland

G. Langs,
Medical University of Vienna, Austria

S. Kriewel,
University Duisburg–Essen, Germany

H. Müller
HES–SO, Switzerland

Keywords: user behavior, survey, image retrieval, radiology

Abstract

Objectives: The main objective of this study is to learn more on the image use and search requirements of radiologists. These requirements are then to be taken into account to develop a new search system for image and associated meta data search in the Khresmoi research project.

Methods: Observations of the radiology workflow, case discussions and a literature review were performed to construct a survey form that was given online and in paper form to radiologists. In addition to the survey, eye tracking was performed on a radiology viewing station to analyze typical tasks and to complement the survey.

Results: In total 34 radiologists answered the survey online or on paper. Image search was mentioned as a frequent and common task, particularly for finding cases of interest for a differential diagnosis. Sources of information besides the Internet, are books and discussions with colleagues. Search for images is unsuccessful in around 25% of the cases, stopping the search after around 10 minutes. The most common reason for failure is that the target images are considered rare. Important additions for search requested in the survey are filtering by pathology and modality, as well as search for visually similar images and cases. Few radiologists are familiar with visual retrieval but several desire the option to upload images for searching similar ones.

Conclusions: Image search is common in radiology but few radiologists are fully aware of visual information retrieval. Taking into account many unsuccessful searches and the time spent for this, a good image search engine could improve the situation and help in clinical practice.

1 Introduction

Image retrieval has been an active research domain for the past 25 years [8] but so far there have been few commercial applications of image search using visual information such as LookThatUp[1] or since mid 2011 Google images. The medical domain has often been mentioned as an important domain for content–based image retrieval (CBIR) [26], but again, only a few clinical applications have been evaluated in clinical routine [1]. The Khresmoi project[2] 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 the specifications of such a system.

To develop applications and products based on real user needs has been a standard procedure in many other fields [19] including medical information retrieval [9, 25]. As is stressed in [11], "...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 very 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 the information needs and the use of information retrieval (IR) systems. In [13], a systematic framework for evaluating the use of medical IR systems is proposed. While the full framework is out of the scope of our study, important concepts, such as user satisfaction and search failure, were taken into account.

A theoretical analysis on information needs is attempted in [29]. Moreover, a model for information seeking behavior is proposed using four categories. Categories *a* and *b* include search strategies by the user, that do depend or not on a mediator/IR system. Categories *c* and *d*, include the search strategies employed by the mediators/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 [18], on the behaviour of journalists when searching for an image. For this purpose, observation of the journalists in their work, interviews with them and analysis of a sample of their queries were used. Analyzing queries is a common approach for investigating visual information search needs and behavior in specialized fields, such as history [5] and art [17], or general large scale studies [10]. However, as it is well explained in [17], "the use of image retrieval systems varies in different fields because users have their own specific information-seeking behavior and need unique features designed for their tasks".

Similar approaches have been followed in the medical domain. Some studies rely on log files from either media search engines [20] or MedLine [22] to find out more information on how clinicians search for images. Others [21, 12] perform interviews and surveys among clinicians, but these were done on a small scale and were not focused on radiologists specifically. The study described in this paper aims to shed light onto the search for images in the field of radiology and identify the requirements for a specialized image search engine.

In general, a variety of methods exist for obtaining information on system use:

[1] `http://www.ltutech.com`
[2] `http://www.khresmoi.eu`

- observation of the behavior of users (which can include direct observation [18, 14] or the analysis of log files that record the behavior of users [18, 5, 17, 10, 20, 30]);
- interviews with stakeholders [18, 12, 14, 6];
- surveys [19, 21, 30].

While observation can provide useful insights on the behavior of the subjects, it often lacks to provide 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 and 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.

In addition to direct observation, another way to monitor image use behavior is observation using eye–tracking equipment. Eye gaze tracking has been widely used in user interface design and evaluation [14, 2, 7, 4] 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, 15, 16, 27]. 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. The number of responses is low as only known radiologists were targeted in the institutions to guarantee a high quality of responses.

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. These observations were used to help constructing the survey and the questions. 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. This includes a description of the workflow through diagnosing cases and preparing the radiology report.

2. Case discussions were followed where interesting cases are explained including the reasoning process, imaging data required, and evidence provided by several exams. This process also described the workflow and the use of external knowledge to support the workflow.

3. Eye tracking experiments were performed, where radiologists diagnose cases while being eye–tracked to obtain information on areas of the images that are of particular interest and about the way a clinician browses through an image.

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. The eye–tracking software and hardware were installed and patient studies were imported to the work station via a CD.

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. The detailed analysis had the goal to find problematic parts and analyze whether the forms were understandable for radiologists.

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 [21]. 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:

1. clinical work on patients;
2. work regarding teaching, as in the preparation of lectures;
3. research work that can include a variety of tasks.

Besides the search requirements, basic demographic data on the radiologists was acquired to better interpret results. The final version of the survey consisted of 4 sections: general data, clinical work, teaching, and research. In the general section there were questions regarding:

- age;
- gender;
- specialization within radiology;
- country of radiology specialization;

- type of hospital;
- years of experience in radiology;
- activity distribution between teaching / research / clinical work.

A common set of questions was used for the three activity domains. The first part is focused on the current image search behaviour of radiologists:

- tasks where images other than those of the patient treated, need to be found;
- sources searched for images;
- how they are searched for;
- how relevance of an image can be determined;
- how often search for images fails;
- why the search fails;
- how much time taken before stopping;
- how much time it taken when relevant images are found.

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. Questions were:

- What are useful additions for search systems?
- What would a perfect search system be like?
- How can visual information of images be exploited?
- Which tools for an automatic annotation of images would be useful?
- Are medical terminologies or ontologies being used?

3 Results

This section describes the results obtained in the study.

3.1 Observation

The radiology workflow starts usually with opening a case for which an imaging exam was requested. The images are then transferred from the PACS server to the local viewing workstation, and then the viewing process can start. 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 such as 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 up in group discussion about possible pathologies, corresponding to "information exchange" as in [29]. 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 and differential diagnosis the medical finding is completed. If, from the scientific or teaching point of view, the study is interesting for the radiologists it is being marked for future reference.

In case discussions, the workflow is similar to the description above 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 cases occurring rarely. Important steps in the diagnosis process are the comparison of findings with the state of the art in the literature and 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 practising 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, measuring for example the growth of a tumour. Computer–aided detection, such as highlighting particular abnormal regions or visualizing results are also important.

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. As the system was on a separate workstation these were chosen cases and not cases they would have worked on anyway. 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 (Figure 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 giving examples of expected answers could be better;
- many persons store images locally on their computers for future use and this has to be taken into account (although this is not a desired practice in most hospitals where data acquisition needs to be validated by an ethics committee);
- radiologists found it hard to separate between the three proposed tasks of teaching, research and clinical work and mixed things when filling the survey, sometimes mentioning the overlap between the tasks;

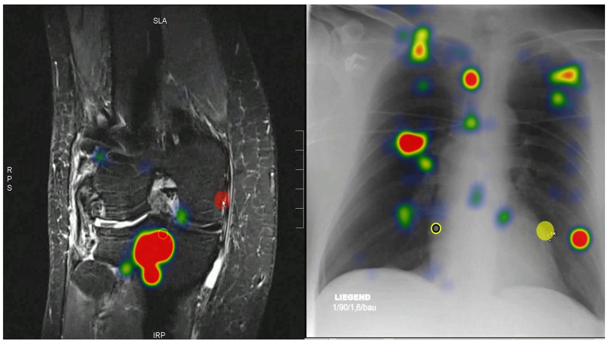


Figure 1: Heat map images for knee MRI (left) and a chest x–ray (right), showing that often a particular region attracts the main attention as in the knee MRI or several distinct regions as with the chest x-r-ay.

- many formulations were modified as clinicians referred to it 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 the survey to which 34 radiologists responded. Ten persons filled in the paper form after a seminar at the University of Geneva, whereas all other participants filled in the electronic form. The ten paper forms were transcribed into the electronic form for a homogeneous treatment.

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 30–55 years. Two thirds are male and one third is female, highlighting that radiology is one of the few domains with a majority of men in medicine. As could be expected from performing the survey mainly in Vienna and Geneva, most persons have had their radiology education in either Austria or Switzerland. All other countries are from Western Europe or the US, meaning that education is comparable.

In terms of radiology specializations there are no surprises. 23 persons specialize in general radiology, 1–2 persons each in musculoskeletal radiology, thoracic radiology, radiology informatics, neuroradiology, orthopaedic radiology and body imaging. One person mentioned to specialise in CT and another one is still a student. 28 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.

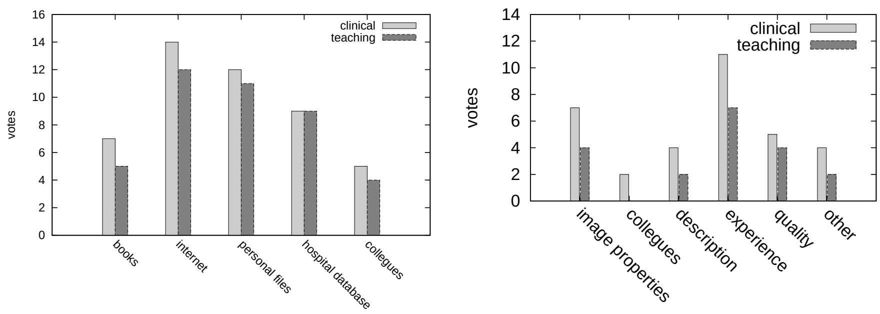
For work time distribution it was possible to weight the time spend on clinical work, teaching and research on a scale from 0–5. This allows estimating the percentage spent on each of the activities. Most persons perform all three activities, few have no teaching or no research and all clinical work.

3.3.2 Clinics — Teaching — Research:

All 34 participants mention tasks where they search for images other than of the person being diagnosed. There are several reasons for searching images apart from the ones they are currently assessing. 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 tumours or lesions in brain, liver or other structures. A task mentioned, where images can be useful is also the grading of a disease.

For teaching, 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 tumours 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 (Figure 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 desktop PCs to search images (sometimes with keywords), as mentioned by 12 . The local patient record is queried using the patient 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 more organized and larger personal databases. Generally keyword search is mentioned and no CBIR systems.



(a) (b)

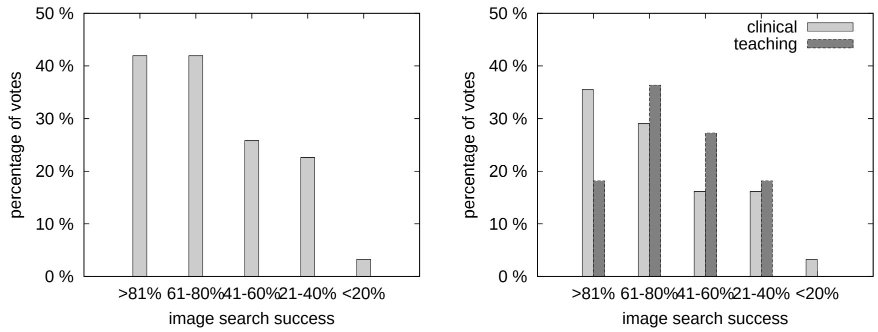
Figure 2: (a)Information sources used for finding images. (b) Defining relevance or suitability of images

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, record date, 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 of comments on the image also has an influence. Asking colleagues for their opinion is another option, mentioned by two persons. Figure 2b depicts the responses for clinical work and teaching. Personal experience is the most important criterion followed by image properties that can include for example the modality.

A notion mentioned by clinicians was the trust that they have in the image or the diagnosis attached to it. If a diagnosis was, for example, only taken based on the images there is less trust than when a biopsy confirmed the diagnosis. For images found on the Internet this information is not always documented.

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 not found. In Figure 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 success rates below 40% are in this category. This may highlight that clinical work is less well defined and has harder search tasks than for example teaching. Based on the question on the success rate of searches, in clinical work the time taken for searches is found to be lower than for teaching and research, which also explains the higher failure rate.



(a) (b)

Figure 3: Self-assessment of the success rate of image searches, (a) the overall percentages and (b) a comparison between clinical work and teaching.

The clinicians think that most of the time the desired images are available but cannot be found due to various reasons. The main reason for not finding a relevant image is that the topic or pathology is rare, too new and sometimes also too general. It needs to be noted that not all storage systems are fully searchable (e.g. scanned reports). Time pressure has a negative impact on finding relevant images as well. Figure 4 compares the responses for clinical work and teaching but both categories lead to similar results.

When comparing search times of successful image search, it becomes clear that over 70% of successful searches finish after ten minutes or less (Figure 5). Only a few persons search longer to successfully find images. For unsuccessful image search 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 important of image search in the workflow. It also highlights the room for

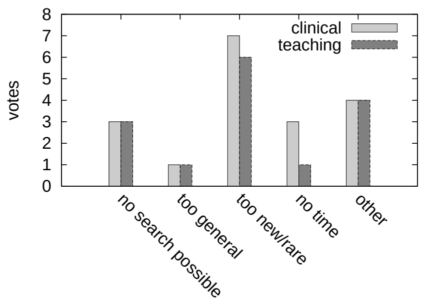


Figure 4: Reasons given for unsuccessful image search.

improvement with optimized search tools to find relevant information quickly.

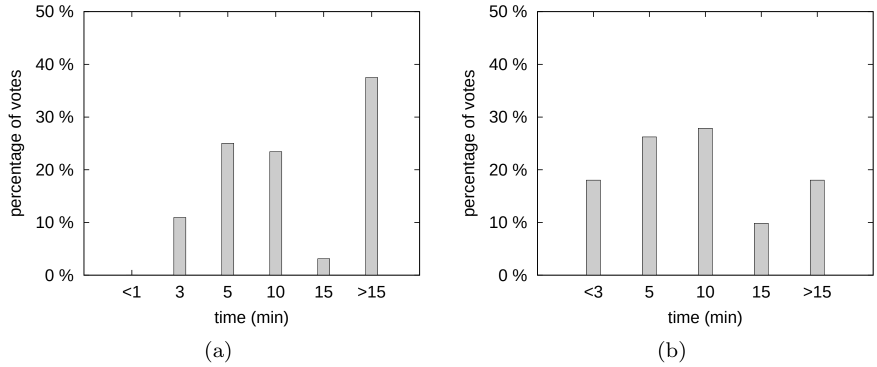
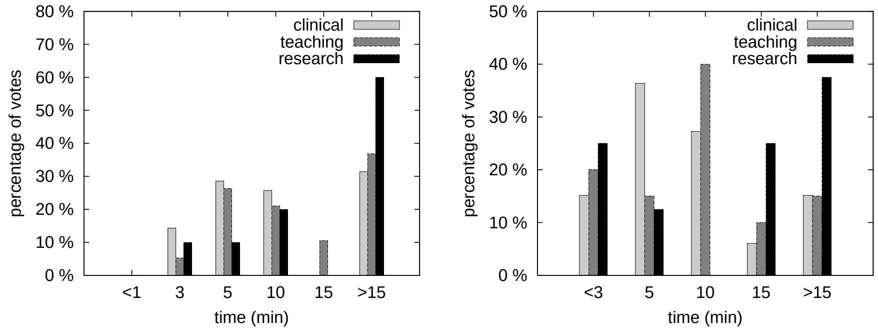


Figure 5: Time taken to find relevant images (a) and times before an unsuccessful search is abandoned (b)

Figure 6 compares the search time for clinical work, teaching and research. The length of successful search for clinical work 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 larger and many search for 10 minutes or more and still find relevant images.

For unsuccessful search, times for research are often more than 15 minutes. During clinical activities, the average time before quitting the search is significantly shorter, 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 more easily find out whether or not relevant images even exist can be very helpful and reduce the amount of time lost.

23 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 even symptom classification

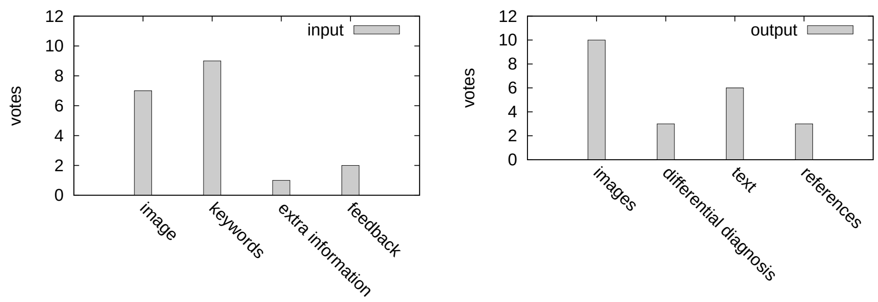


(a) (b)

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

(using for example an ontology), query by text and image 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 to judge the confidence of a diagnosis.

19 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 (Figure 7a). Keywords could vary strongly, from describing the anatomical structure, the pathology and histology, up to more demographic information like patient age.



(a) (b)

Figure 7: Desired (a) input and (b) output data for a medical image search.

The output should include image examples and a corresponding description (Figure 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 would be useful such as a biopsy, 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, indicating 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 judgements of other radiologists on the similarity (mentioned by two persons). 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 at all. Medical Subject Headings (MeSH) was mentioned most frequently (10). There is no major difference in terminology use between clinical work, teaching and research. Five participants mention to use RadLex and SNOMED CT was mentioned once. The dominant PACS system mentioned is AFGA IMPAX (8),followed by Fuji Synapsis (3) with one participant each mentioning GE, McKesson, CareStream, PAC/RIS and Siemens Syngo. One participant answered to use 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 behaviour several persons added comments about the perfect search system. Most comments mentioned here in their raw form (the survey as well as the responses were entirely in English), not ordered by domain (teaching, clinical work, research):

- like Google but including DICOM images and text combined;
- structured information on a case including histopathology, images, structured data;
- confidence score in the diagnosis, e.g. backed by biopsy or other exams; search by diagnosis;
- simple to use, minimize need to play with 3D stuff;
- differential diagnosis, multiple views possible, feedback of others possible;
- pathology chosen by radiologists, search among images in a similarity cluster;
- like Google but with references to the literature such as Goldminer;
- marking the ROI in an image and search with this, search for anatomical structures, for normal and abnormal cases;
- keyword and image as input for search, selection of case and then search for further information on the found cases;
- quantification of the size of structures to search for;
- search for differential diagnosis, location, organs and particular conditions;
- search by diagnosis, the opposite to clinical reading;

- look for a certain pathology and find cases with it including images;
- having research databases and research PACS linked, search by keywords rather than by image;
- having the entire patient documentation searchable by keywords; yottalook is of good quality for this;
- search by image description, pathology and histology would be useful;
- "show me a similar image for which a final diagnosis is available";
- free text search in radiological reports;
- Be able to search by various key words. E.g., By diagnosis: "pneumonia" would retrieve many different types and appearances of pneumonia.You should be able to limit your retrieval by additional terms, e.g., "Klebsiella pneumonia" would limit the retrieval to pneumonia caused by that pathogen;
- By manifestation: "interstitial disease" would retrieve many different types and appearances of interstitial diseases. You should also be able to limit further using additional descriptors;
- By appearance: e.g., "round lung lesions" would retrieve all types of diagnoses that would fit that description;
- It should also have a "more like this" function;
- It should have a way of circling a region of interest and retrieve more images that are similar to the region of interest that was indicated.

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 even for an image–based domain such as radiology. 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. Similar to search engines like Google used by all clinicians 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 unproblematic 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 of important for the search.

4 Discussion

While the introduction of filmless imaging and PACS has improved the performance and efficiency in radiology [19, 23, 24], many problems still exist in the current situation in terms of information retrieval. Correctly describing the pathology can be a difficult task, but is essential for retrieving usable results. Visual retrieval is not possible with the clinical system 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.

Hardware limitations of the eye tracking system lead to a reduced experimental setup where only one monitor was tracked (instead of the two that are otherwise often used in radiology). Therefore, it was not possible to capture the precise workflow of radiologists using the specialized hardware. 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 but this first study was mainly aimed at obtaining first ideas. Further studies are foreseen within Khresmoi to observe actual image search behavior once the first prototypes are available. Substantial differences in viewing behaviour between image types were found. In some cases single areas of high concentration of fixation can be determined while other image types show broad scanning paths in viewing behaviour. 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. This will be solved with a new version of the hardware although the main findings can also be obtained with a single screen;
- 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 their activity on and thus search by regions constitutes one of the important requirements for future medical CBIR systems.

Obtaining a very large number of responses from often busy and overloaded radiologists was difficult. 34 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 5 years experience. This has the advantage to have persons who grew up with 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 [28] — 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 in 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 modality, 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 analysing 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 experience 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, which is currently uncommon. 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 a first step to better understand the requirements of radiologists in handling images and searching for visual data that can help them in daily tasks. It has to be mentioned that many radiologists are not familiar with visual retrieval, so 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 for planning the next steps to reach an impact of the technologies in radiology image search.

6 Acknowledgements

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

References

[1] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori. Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. *Radiology*, 228(1):265–270, July 2003.

[2] M. S. Atkins, A. Moise, and R. Rohling. An application of eye gaze tracking for designing radiologists' workstations: Insights for comparative visual search tasks. *ACM Transactions on Applied Perception (TAP)*, 3(2), April 2006.

[3] D. V. Beard, R. E. Johnston, and O. Toki. A study of radiologists viewing multiple computed tomography examinations using an eyetracking device. *Journal of Digital Imaging*, 3(4):230–237, 1990.

[4] T. Beckers and N. Fuhr. User–oriented and eye-tracking–based evaluation of an interactive search system. In *4th Workshop on Human–Computer Interaction and Information Retrieval*, 2010.

[5] Y. Choi and E. M. Rasmoussen. Searching for images: The analysis of users' queries for image retrieval in american history. *Journal of the American Society for Information Science and Technology*, 54(6):498–511, April 2003.

[6] K. W. Cogdill, C. P. Friedman, and C. G. Jenkins. Information needs and information seeking in community medical education. *Academic Medicine*, 75(5):484–486, May 2000.

[7] E. Cutrell and Z. Guan. What are you looking for?: an eye–tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007.

[8] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, April 2008.

[9] W. M. Detmer, O. G. Barnett, and W. Hersh. Medweaver: integrating decision support, literature searching, and web exploration using the umls metathesaurus. In *Proceedings of the AMIA Annual Fall Symposium*, pages 490–494, 1997.

[10] A. Goodrun and A. Spink. Image searching on the excite web search engine. *Information Processing & Management*, 37(2):295–311, March 2001.

[11] P. Gorman. Information needs of physicians. *Journal of the American Society for Information Science*, 46(10), January 1995.

[12] W. Hersh, H. Müller, P. Gorman, and J. Jensen. Task analysis for evaluating image retrieval systems in the ImageCLEF biomedical image retrieval task. In *Slice of Life conference on Multimedia in Medical Education (SOL 2005)*, Portland, OR, USA, June 2005.

[13] W. R. Hersh and D. H. Hickam. How well do physicians use electronic information retrieval systems? *Journal of the American Medical Association*, 280(15):1347–1352, 1998.

[14] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *Proceedings of the 9th ACM/IEEE–CS joint conference on Digital libraries*, 2009.

[15] H. L. Kundel, C. F. Nodine, E. F. Conant, and S. P. Weinstein. Holistic component of image perception in mammogram interpretation: Gaze–tracking study. *Radiology*, 242:396–402, February 2007.

[16] H. L. Kundel, C. F. Nodine, and E. A. Krupinski. Using gaze–tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology*, 15(7):881–886, July 2008.

[17] H. liang Chen. An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology*, 52(3):260–273, 2000.

[18] M. Markkula and E. Sormunen. Searching for photos – journalists' practices in pictorial IR. In J. P. Eakins, D. J. Harper, and J. Jose, editors, *The Challenge of Image Retrieval, A Workshop and Symposium on Image Retrieval*, Electronic Workshops in Computing, Newcastle upon Tyne, 5–6 1998. The British Computer Society.

[19] A. Moise and M. S. Atkins. Design requirements for radiology workstations. *Journal of Digital Imaging*, 17(2):92–99, 2004.

[20] H. Müller, C. Boyer, A. Gaudinat, W. Hersh, and A. Geissbuhler. Analyzing web log files of the Health On the Net HONmedia search engine to define typical image search tasks for image retrieval evaluation. In *MedInfo 2007*, volume 12 of *IOS press, Studies in Health Technology and Informatics*, pages 1319–1323, Brisbane, Australia, 2007.

[21] H. Müller, C. Despont-Gros, W. Hersh, J. Jensen, C. Lovis, and A. Geissbuhler. Health care professionals' image use and search behaviour. In *Proceedings of the Medical Informatics Europe Conference (MIE 2006)*, IOS Press, Studies in Health Technology and Informatics, pages 24–32, Maastricht, The Netherlands, August 2006.

[22] H. Müller, J. Kalpathy-Cramer, W. Hersh, and A. Geissbuhler. Using Medline queries to generate image retrieval tasks for benchmarking. In *Medical Informatics Europe (MIE2008)*, pages 523–528, Gothenburg, Sweden, May 2008. IOS press.

[23] R. O. Redfern, S. C. Horii, and E. Feingold. Radiology workflow and patient volume: Effect of picture archiving and communication systems on technologists and radiologists. *Journal of Digital Imaging*, 13(2):97–100, May 2000.

[24] B. I. Reiner, E. L. Siegel, and C. Flagle. Effect of filmless imaging on the utilization of radiologic services. *Radiology*, 215:163–167, April 2000.

[25] D. Revere, A. M. Turner, and A. Madhavan. Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. *Journal of Biomedical Informatics*, 40(4):410–421, August 2007.

[26] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content–based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[27] G. D. Tourassi, M. A. Mazurowski, and B. P. Harrawood. Exploring the potential of context–sensitive cade in screening mammography. *Medical Physics*, 37(11):5728–5736, November 2010.

[28] A. J. A. M. van Deursen and J. A. G. M. van Dijk. Internet skills performance tests: Are people ready for ehealth? *Journal of Medical Internet Research*, 13(2), 2011.

[29] T. D. Wilson. On user studies and information needs. *Journal of Documentation*, 37(1):3–15, 1981.

[30] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003.