**Grant Agreement Number: 257528**

**KHRESMOI**
**www.khresmoi.eu**

# D9.1: Report on image use behaviour and requirements

| | |
|---|---|
| **Deliverable number** | *D9.1* |
| **Dissemination level** | *public* |
| **Delivery date** | *18.5.2010* |
| **Status** | *final* |
| **Author(s)** | *Henning Müller* |

SEVENTH FRAMEWORK PROGRAMME

# 1 Executive summary

This deliverable document for deliverable D9.1 of the Khresmoi project describes the methodology for constructing a survey among radiologists on image search behaviour and analyses the results obtained.
The construction of the survey was based on an analysis of the literature on analysing medical image search and on an observation of radiologists diagnosing cases and discussing past cases. This includes the use of eye tracking equipment on a radiology workstation to find out more about areas of attention and the parts exactly analysed in images. Structured interviews were performed with the first version of the survey to make the questions well understandable and modify potentially misunderstandable parts.
In total, 26 radiologists answered the survey, mainly from Switzerland and from Austria. The persons answering the survey were in a large part below 30 years of age and little experienced, but several people with over 15 years of experience in radiology participated as well.
Questions were asked separately for the three main activities, notably clinical work, teaching and research. Several people mentioned that these are hard to separate and only answered once for the three domains.
Image search was mentioned as a common task, particularly for finding cases of interest for differential diagnosis to a current case. Sources of information besides the Internet, where Google with key words is frequently used, are books and discussions with colleagues to share experiences. Specialized search engines such as Goldminer or Yottalook are mentioned by a few people. Relevance is often determined by image criteria such as modality or resolution and most often is based on personal experience.
Search for images is successful in around 65% of the cases. When it is not, most often it is said that the target is rare or new and only rarely that the images might not exist. Successful image search takes in general 10 minutes or less whereas many unsuccessful searches are stopped only after more than 15 minutes. For clinical work the time available is 5 minutes whereas for teaching and research a little more time seems to be available.
Important additions for search were mentioned to be filtering by pathology and modality, and also search for similar images and similar cases was often mentioned.
Very few of the radiologists are familiar with visual search but many would like to have the possibility to upload images for searching similar ones. Terminologies to describe images are only very rarely used but several persons regard the exact formulation of a query as difficult and would like to be able to search for specific pathologies and modalities.
Image search is a common task in radiology but currently only few radiologists are fully aware of research in visual medical information retrieval. Taking into account many unsuccessful searches and the time used for this a good image search engine could well improve the situation. Particularly filtering of results by modality, pathology, anatomic region but also based on patient demographics are other aspects that a radiology image search engine would need to include to have real impact.

# 2 Table of Contents

# List of abbreviations

| | |
|---|---|
| CT | - Computed Tomography |
| DICOM | - Digital Imaging and Communications in Medicine (standard) |
| ICD | - International Statistical Classification of Diseases and Related Health Problems |
| IRMA | - Image Retrieval in Medical Applications |
| MedGIFT | - Medical GNU Image Finding Tool |
| MeSH | - Medical Subject Headings |
| MRI | - Magnetic Resonance Imaging |
| PACS | - Picture Archival and Communication System |
| ROI | - Region of Interest |

# 3 Introduction

Image retrieval has been an active research domain for the past 25 years [1], but so far there have been few real applications of image retrieval in routine use. The medical domain has often been mentioned as an important domain for content-based retrieval, but again, only a few clinical applications have been evaluated in clinical routine [2].
To develop applications based on real user needs has been a standard procedure in many other fields including information retrieval. For visual information retrieval this was initially only very rarely done. Approaches were rather technology driven in terms of applications than based on user requirements. One of the first user analyses for image retrieval was [3], analysing the behaviour of journalists when searching for an image. Similar requirements can also be obtained for clinicians searching for medical images. First studies exist [4,5,6,7], but most of these studies have so far been done on a small scale and often not concentrated on radiologists but on clinicians in general. Besides interviews these studies also rely on log files from either media search engines or MedLine to find out more information on how users search for images.
Various methods exist in general for obtaining information on system use:

- observation of the behaviour of users (which includes in one way or another the analysis of log files that record the behaviour of users);
- interviews with stakeholders to obtain in-depth information, mainly on a qualitative level as such interviews are usually time-consuming;
- surveys to gather quantitative information.

One of the target user groups in the Khresmoi project are radiologists, and thus the information search behaviour of radiologists is the target of this deliverable.
The main objective of this deliverable is the description of the methodology of analysing the requirements of radiologists in terms of image use and image search, then the execution of the survey and an analysis of the results. The results should subsequently be used to build the prototypes of the Khresmoi system aimed at radiologists at the end of year 2 of the project.
An important aspect that needs to be taken into account is the fact that radiologists are often very busy and getting detailed feedback from them is therefore often not easy in large numbers. This study aimed at obtaining responses from around 30 radiologists and 26 responses were finally obtained in time for the deliverable. In addition, there are currently very few visual retrieval systems used in practice so many radiologists are not aware of the possibilities and mainly know the data access by patient ID in clinical records and the free text search of Google that are commonly used by clinicians.
This text is structured as follows: Section 4 describes the first step in the process that consisted of observing the use of images by radiologists to better understand radiology workflow and places where visual search could be useful. Section 5 describes the main steps in constructing the survey including the main questions. Then Section 6 describes a small number of structured interviews with the survey form to adapt it to the needs of radiologists and make it as understandable as possible. Section 7 analyses the main results and section 8 concludes this deliverable.

# 4 Observation

The first step in learning more about information behaviour of radiologists is watching them perform standard tasks and then analysing which information is necessary at which point and which are the moments where a specific information need occurs and where information is missing. These observations were used to construct the survey and the questions of the survey in connection with a literature review.
To obtain information on the workflow three steps were taken. First, in the Vienna radiology department experienced radiologists described the main steps of the radiology image analysis process to the Khresmoi researchers. This included a description of the workflow through diagnosing several

cases and preparing the radiology report. Then, in the Geneva radiology two seminars were followed where interesting cases are discussed among radiologists including the reasoning process, imaging data required, and evidence provided by several exams. Again, this process described the workflow and the use of external knowledge to support the workflow in radiology. A last step consisted in an eye tracking experiment, where radiologists had to diagnose cases while being eye tracked. This procedure had the goal to have not only explicit knowledge given by the radiologists on the process but also understand better how image browsing is performed, which parts of an image are regarded and how this process could be supported by image retrieval or computer-aided detection applications.

## 4.1 Following the clinical workflow

In the clinical workflow the radiologists start off with choosing a case for which an imaging exam was requested. The images of the case are then fetched from the PACS (Picture Archival and Communication System) server and transferred to the local viewing workstation. When the process is complete, the radiologist can start with the image analysis process.

The viewing options are set up depending on requirements as the software allows various settings, such as size and number of views per screen. The setup depends on the modality of the medical record and on the radiologist's preferences. It can be changed during the analysis process. Before starting to analyse the images the patient's medical history and anamnesis is reviewed.

The radiologist then starts analysing the images by adjusting the brightness/contrast and scrolling through the slices using the mouse wheel. The modalities (different series highlight different pathologies) are switched using thumbnail previews provided by the software. Also a distance measurement tool is available for evaluating size of specific organs and pathologies.

When a pathology or abnormality is found there usually are two possibilities:

1. The abnormality is known:
   diagnosis and differential diagnosis are given and the medical finding is completed.
2. The abnormality is unknown:
   a common way to handle an unknown abnormality for a radiologist is to ask a more experienced colleague for additional information. This sometimes ends up in a group of radiologists discussing possible pathologies.
   Most of the time the radiologist has to search through the available literature (Internet, books, databases). To do so, the pathology needs to be described as good as possible. This is a critical and difficult task, as the quality of the research depends on the chosen search phrases and verbally describing visual information is prone to errors.

With the final 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 as reference for the future. The entire workflow is also illustrated in Figure 1.1.

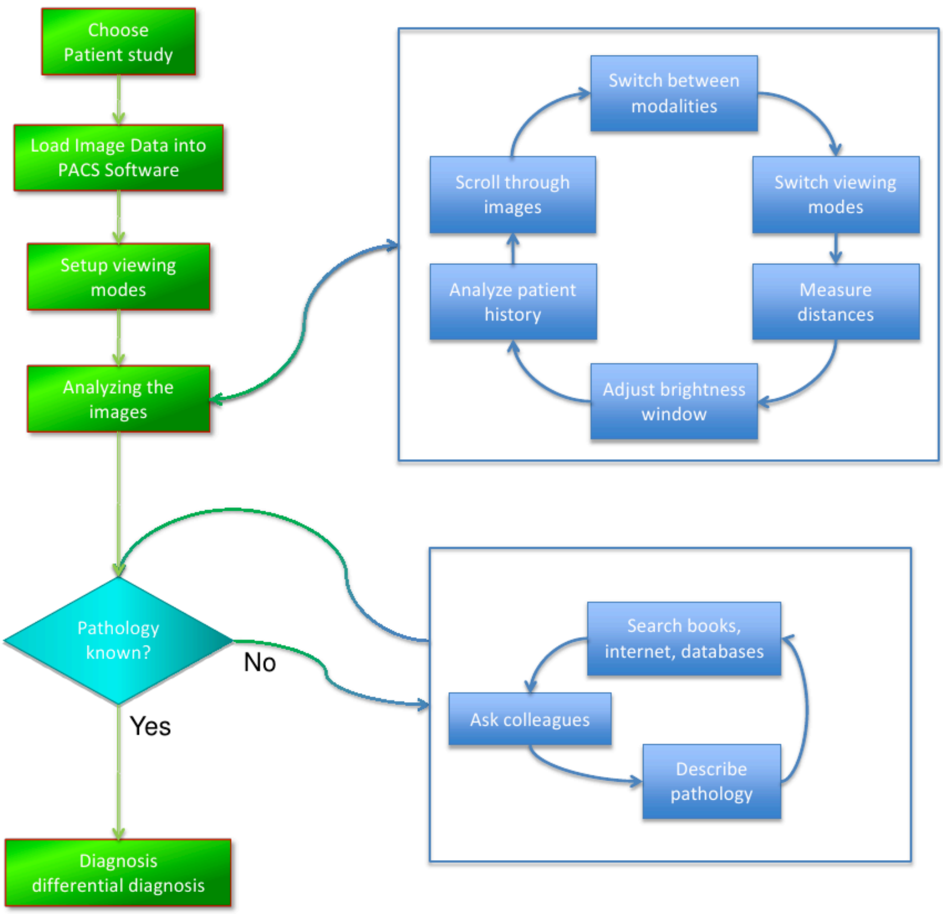


*Figure 1: Standard radiology workflow when analysing images.*

Many problems arise in the current situation in terms of information retrieval. As mentioned, correctly describing the pathology can be a difficult task, but  is essential for retrieving usable results. Visual search (searching with images) 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 first. Searching in books is easy when ideas about potential 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 research. These search engines are powerful but not designed for medical purposes and often return undesired and low quality results.

In seminars, the workflow is similar to the workflow described above. 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 comparing the findings with the state of the art in the literature and requesting additional exams (imaging, laboratory, etc.) to assure that the probability of a correct diagnosis is high. This means that access to and knowledge of the literature is extremely important in doing evidence-based medicine also in radiology. Justifying decision is important and links to related cases can be regarded as extremely important.

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 detecting, such as highlighting particular abnormal regions or visualizing results are also mentioned as being important.

## 4.2 Eye tracking on a clinical work station

Through observation of clinicians diagnosing images and through speak aloud protocols much information can already be obtained. Still, having a clear idea about areas of the images that are of particular interest and about the way a clinician browses through an image can give additional information.

Eye tracking equipment [8,9] can thus help to obtain this additional information on the use of images in radiology. Eye tracking is being used in many domains such as for the placement of advertisement on web pages, but the use for medical viewing stations is to our knowledge relatively novel. Past work exists in this domain such as [10,11]. These descriptions usually used eye tracking for artificial tasks or in very specific diagnosis situation but not specifically in terms of image retrieval tasks.

The setup of the eye tracking for the Khresmoi project consisted of a stand-alone workstation at the University hospital of Vienna, which was not connected to the PACS network. There was one workstation PC connected to one 23" LCD-Monitor. The eye-tracking software and hardware were installed properly and patient studies were imported to the IMPAX-Software via CD.

Before performing the image viewing tasks, the eye-tracking system was calibrated for each candidate. When calibration was finished successfully, the candidates were performing their standard radiology tasks described in section 2.1

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 person worked on a knee-MRI. As the system was on a separate workstation these were chosen cases and not cases that the radiologists would have worked on in any case. The task for the candidates was to perform their 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 (as shown in Figure 1.2).


*Figure 2: 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-ray.*

Technical difficulties due to 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 fully capture the workflow of radiologists using their 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 and solve the administrative problems with the recordings. Further studies are foreseen with the Khresmoi prototypes to observe actual image search behaviour once the first prototypes will become 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. Even more than the simple heat maps shown in Figure 2, the videos produced during the recording are important.

Several administrative problems and limitations 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 could potentially be solved with a new version of the hardware although the main findings can also be obtained with a single screen;
- administrative problems regarding access to confidential data delayed the recording session and lead to a limited amount of data recorded; using a stand alone PC solved some of the problems;
- videos recorded had to be anonymized as well before they could be analysed, 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 better.

# 5 Constructing the Survey

Starting point for the survey was another user study that was previously performed in Portland, Oregon, USA , and then later in Geneva, Switzerland [4]. Based on the questions in this survey the form was adapted to comments from local radiologists and 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 in these situations:

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

Besides the search requirements some basic demographic data on the persons 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:

- the age;
- the gender;
- the medical specialization within radiology;
- the country of radiology specialization;
- the type of hospital;
- the years of experience in radiology;
- the activity distribution between teaching/research/clinical work.

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

- the tasks where images, other than the ones of the patient treated, need to be found;
- which sources are 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 failed;
- how much time it takes before stopping;
- how much time it usually takes when relevant images are found.

In the second part, the participants were asked to propose services and additional tools that would be useful to their search, and imagine a perfect image search system for their needs. The eventual system design should address the existing requirements of radiologists. Questioned were:

- what are useful additions for search systems;
- what would a perfect search system be like;
- how can visual information of images be exploited;
- whether any tools for an automatic annotation of images would be useful;
- whether any medical terminologies or ontologies are used.

These questions were expected to help find out the differences in search behaviour between activities and were also aimed at giving radiologists the possibility to provide ideas for future systems and desired functionalities.

# 6 Structured interviews

After a first survey form was constructed on paper we performed several structured interviews with the survey form to see whether the questions were understandable and whether the responses correspond to our requests. Such a detailed analysis did not have the goal to completely change the survey but rather find problematic parts and analyse whether the forms were understandable for radiologists or whether more information would need to be added.

Three detailed interviews were performed in Geneva with 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.

The main outcomes of these structured interviews were:

- radiologists are not very familiar with visual search and giving more examples of what we might be expecting as 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 all data acquisition even by clinicians has 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;
- many formulations were modified as clinicians referred to it as computer science jargon that could be hard to understand.

The results of the structure interviews were only used for modifying the forms but are not added in the results evaluated below.

# 7 Results from the Survey

This section describes the number of survey forms and the main results gathered from the acquired data. Part of the interpretation is also given in this results section and then the most important points are summarized in the conclusion section.

## 7.1 Analysis of the persons having answered the form

One of the main problems that became apparent when starting the survey was motivating people to take the time and fill in the complete survey. Particularly senior radiologists are often solicited to comment on new tools and applications and generally mention having very little time. At the same time the survey was not published largely but was only send to people known to guarantee a high quality of results. Both paper and electronic versions were available.
10 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.

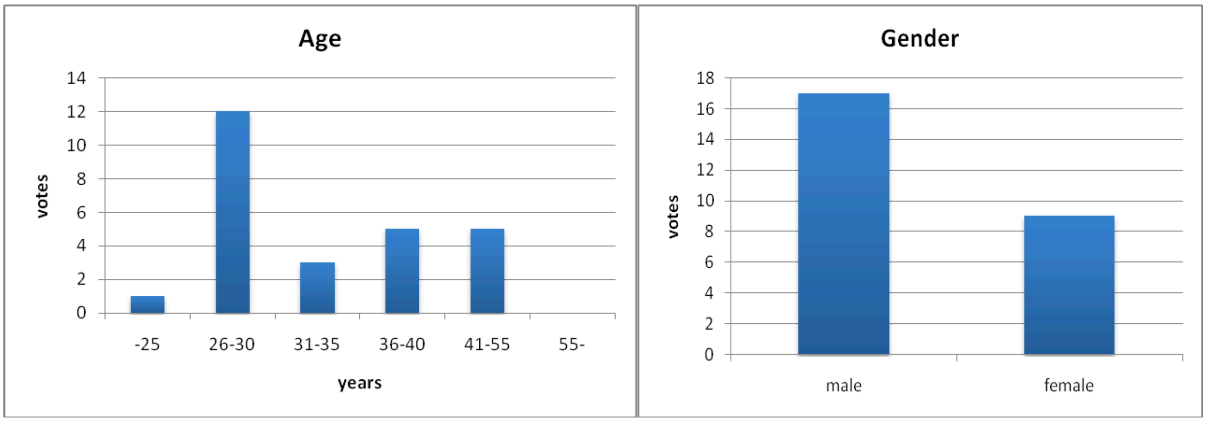


*Figure 3: Analysis of age and gender of the persons participating in the survey.*

In Figure 3 we can see the age and gender distribution in the survey. As the surveys were filled in seminars, the main respondents are between 26-30 years, which is about half of the population. 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.

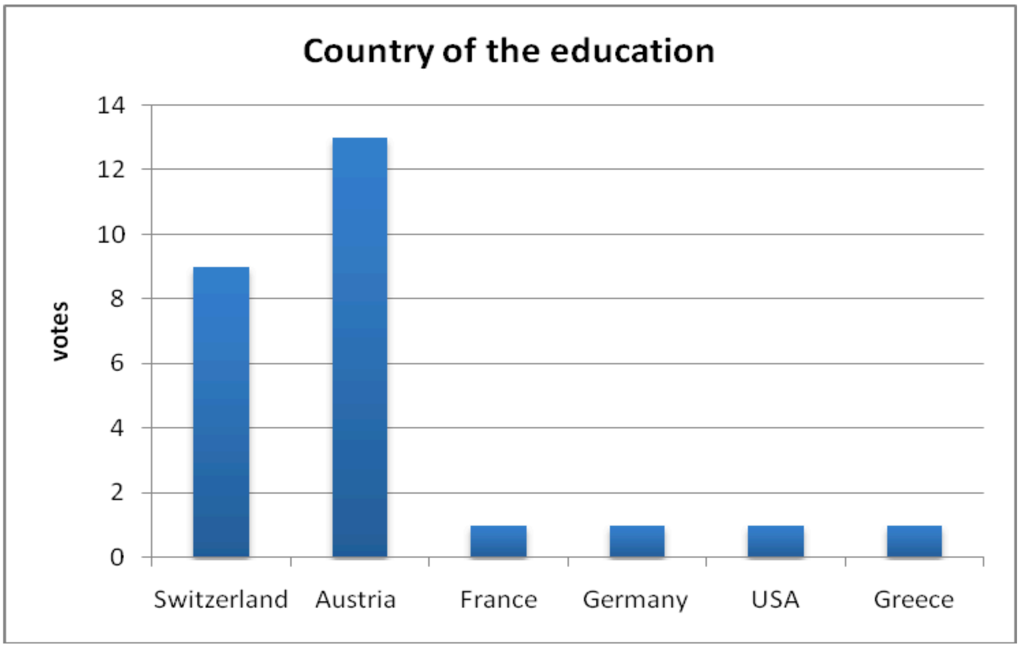


*Figure 4: country of the radiology education.*

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 (see Figure 4). All countries are form Western Europe or the US, meaning that conditions should be comparable. As we plan to keep the survey online we also hope to be able to present a larger spectrum in the long run.

In terms of radiology specializations there are also no surprises. 19 persons specialize in radiology. Two specialize in musculoskeletal radiology, and one each in neuroradiology, orthopaedic radiology and thoracic radiology. One person mentioned to specialise in CT and another one is still a student.
22 persons work in public hospitals and two in private clinics, with two persons mentioning to mainly work in research at the University.

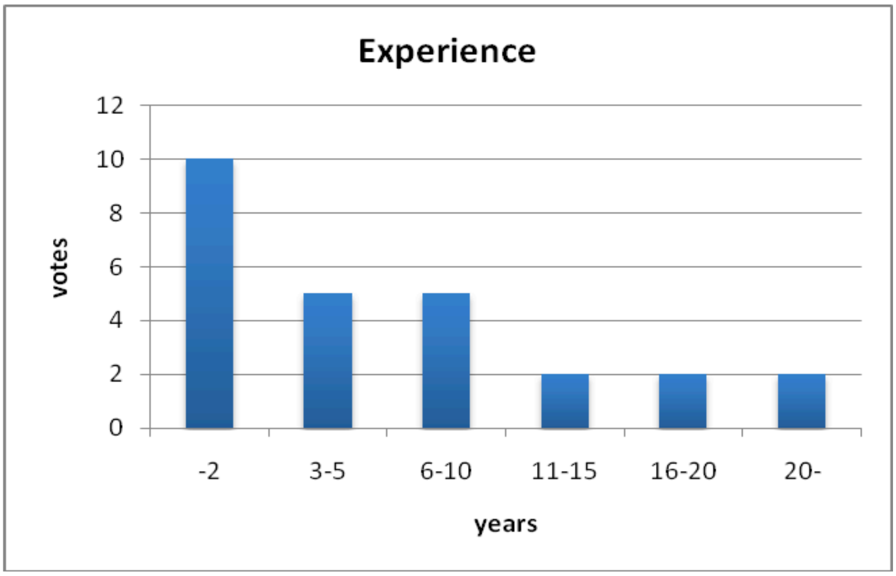


*Figure 5: Years of experience for the radiologists.*

With a rather junior sample, we can also see in the years of experience that ten persons have less than two years of radiology experience. Otherwise, the distribution is relatively even in terms of experience. Only two persons have more than 20 years of experience.
In terms of work time distribution it was possible to weight the amount of 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 and very few have no teaching or no research, with clinical work being done by all respondents. On average a person has 51% of clinical work, 24% of teaching and 25% on research in our sample.

**QUANTITATIVE RESULTS OF THE SURVEY FOR CLINICS/TEACHING/RESEARCH**

This section analyses the responses that were performed separately for clinical work, teaching and research. As clinical work was the first section, the participants generally filled in this section whereas the following sections were filled in less often completely, although almost all persons mentioned activity in the three sectors. Some persons mentioned that the other categories have similar characteristics and that they did not fill them in for this reason. As a results not all question can be analysed in full detail between the three areas and sometimes the three are combined whereas in other sections two or three are compared to highlight the differences between the activities.

**Question 1: Please, list the most important tasks where you need to find images other than the ones of the patient that you are treating (such as searching for reference cases/images in books or teaching files, or finding images on the Internet). For each of these tasks, give an example of the kind of images you would try to find.**

All 26 participants mention tasks where they search for images other than the person they are diagnosing.
There are different reasons for which the clinicians search images apart from the ones they are currently assessing for a specific patient. The main reasons to search for images are finding material for presentations (mentioned by 5 persons), differential diagnosis during a medical finding, for example for difficult cases or in case of an unclear pathology (mentioned by 10) or performing clinical

research (mentioned by 3). Some specific examples listed by the participants are lung fibrosis, brain- or bone tumours or lesions in brain, liver or other anatomical structures. Specific tasks mentioned where such images can be useful are also the grading of the disease.
For teaching, the main focus of the 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 and also includes differential diagnosis. Links with images of the scientific literature were also mentioned as being useful.
As many medical schools are now teaching based on a problem-based learning curriculum, the need to search for information regarding teaching might actually increase and information search skills become increasingly important. Students should thus already be familiar in linking a case in a problem they need to resolve in class with related cases in the literature and need to increasingly have such information search skills.

**Question 2: When performing each of these tasks, where exactly do you look for the images (Internet search engines, your personal files)? When looking for these images, how do you look for them (key word search, personal structure, patient name, asking a colleague…)?**

While performing the search the most frequently used source is the Internet using keywords (see also Figure 6) mentioned by 8 persons. The search engine Google (5) is used as well as public medical databases (PubMed, Goldminer, e-anatomy, mentioned each by 1-3 persons) and Wikipedia (mentioned by two). Most of the clinicians also have personal files stored locally on desktop PCs to search images (sometimes with keywords), as mentioned by 10 persons. Also the local hospital patient record system is queried using the patient name or ID, which are often stored on PDAs or local desktops regarding interesting or typical cases (mentioned by 3). Other options for finding information is looking at books (mentioned by 5) and asking colleagues (3). There is no significant difference between Clinical and Teaching activities. The clinicians focusing on teaching seem to have more organized and larger personal databases.
Generally keyword search is mentioned and no visual search systems.

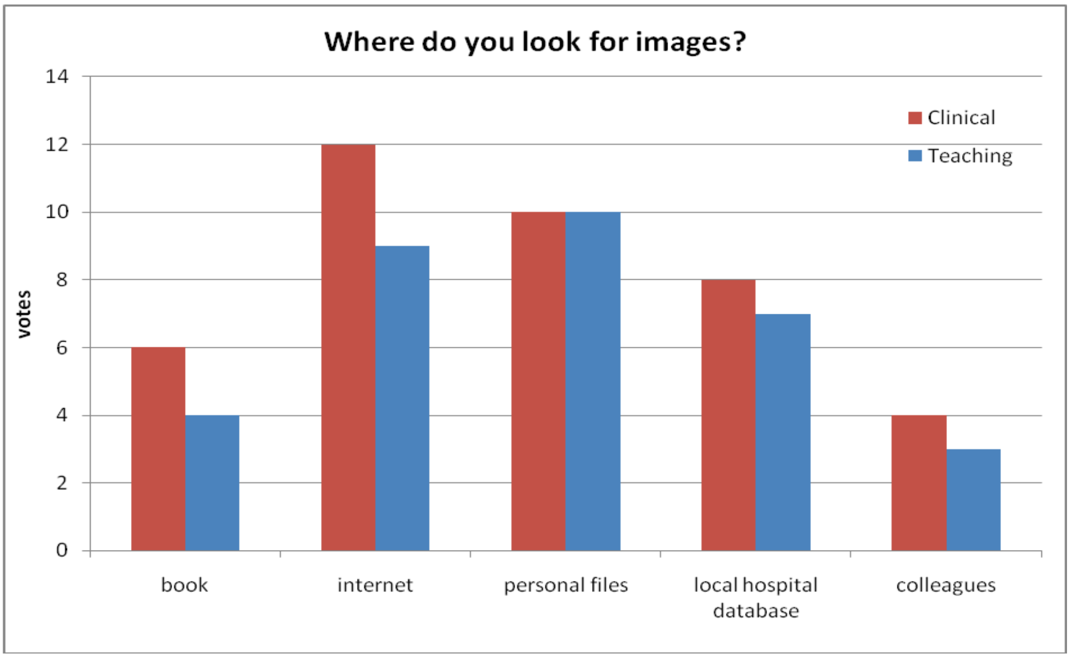


*Figure 6: Information sources for finding images.*

**Question 3: When you find images, how do you decide whether one (or more) is suitable for your needs?**

When an image is found, the clinicians have to decide whether or not it is useful for a particular need based on experience and comparison with their reference case. This corresponds to the notion of relevance in information retrieval tasks.
The correct image properties (e.g. modality, contrast, patient age / gender, record date, mentioned by 5) as well as the quality of the images and the reliability of the sources define the relevance or suitability of the found images. The availability of a detailed description or of comments on the image can also have an influence. Asking colleagues for their opinion is frequent, too, mentioned by four persons. Figure 7 notes the frequency of responses for clinical work and teaching. Personal experience is the most important criterion followed by image properties that can include for example a particular modality.
A notion that was mentioned by clinicians was also 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 or available.

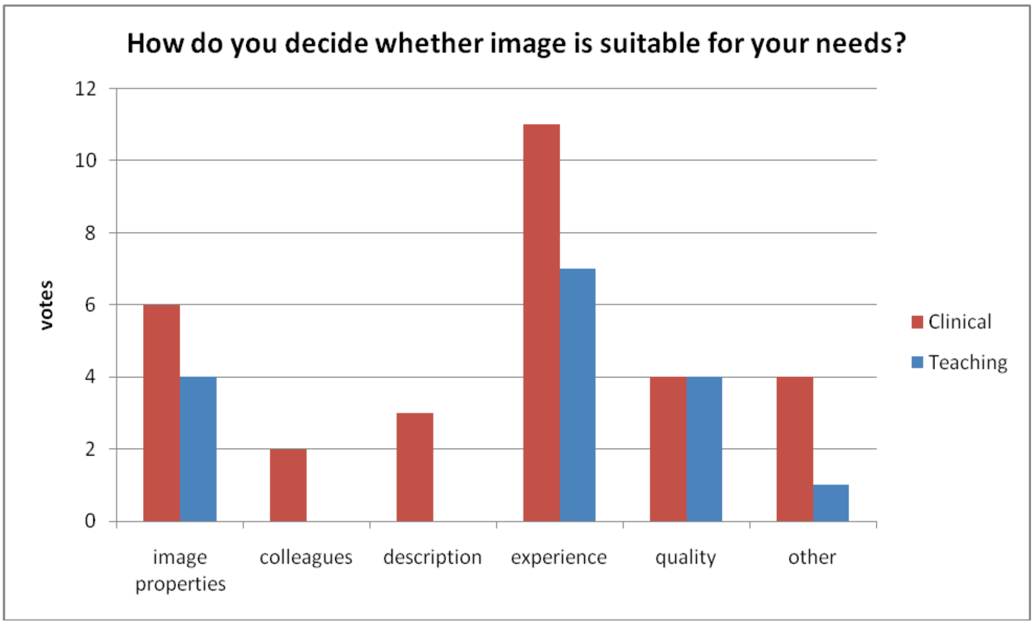


*Figure 7: Defining relevance or suitability of images.*

**Question 4: How often does your search for relevant images succeed?**

On average, the clinicians have roughly a 65% success rate when searching for relevant images (based on this self assessment). This could actually be an overestimation as people might not be aware of all available data and potentially more relevant items that were not found.
In Figure 8 we can see that overall the percentage of persons with a success rate below 40% is quite low. When comparing teaching and clinical work it becomes clear that clinical work has a higher risk of failed searches 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 has. Based on the question about the success rate of searches we could also see that in clinical work the time taken for searches is lower than for teaching and research, which can also explain the higher failure rate.

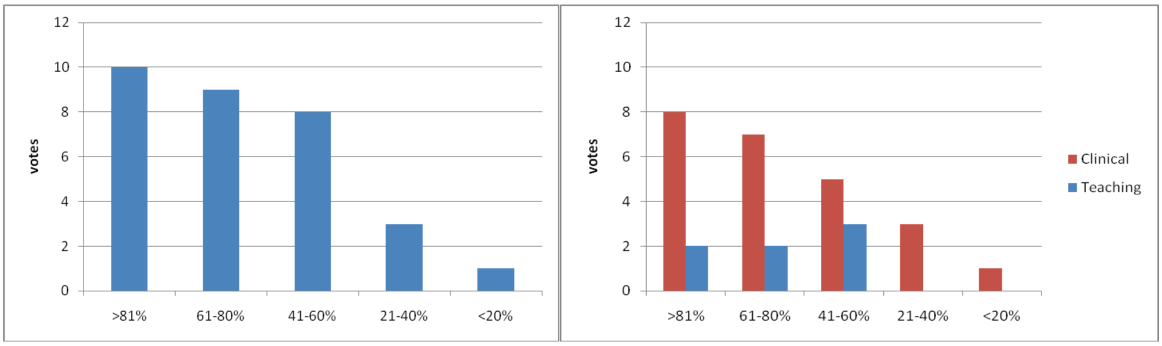


*Figure 8: Self-assessment of the success rate of image searches, on the left the overall absolute votes and on the right a comparison between clinical work and teaching.*

**Question 4.1: Why do you think you did not find any relevant images? In which situations do you not find the images you were searching for?**

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 too 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 9 compares the responses for clinical work and teaching but both categories lead to similar results.

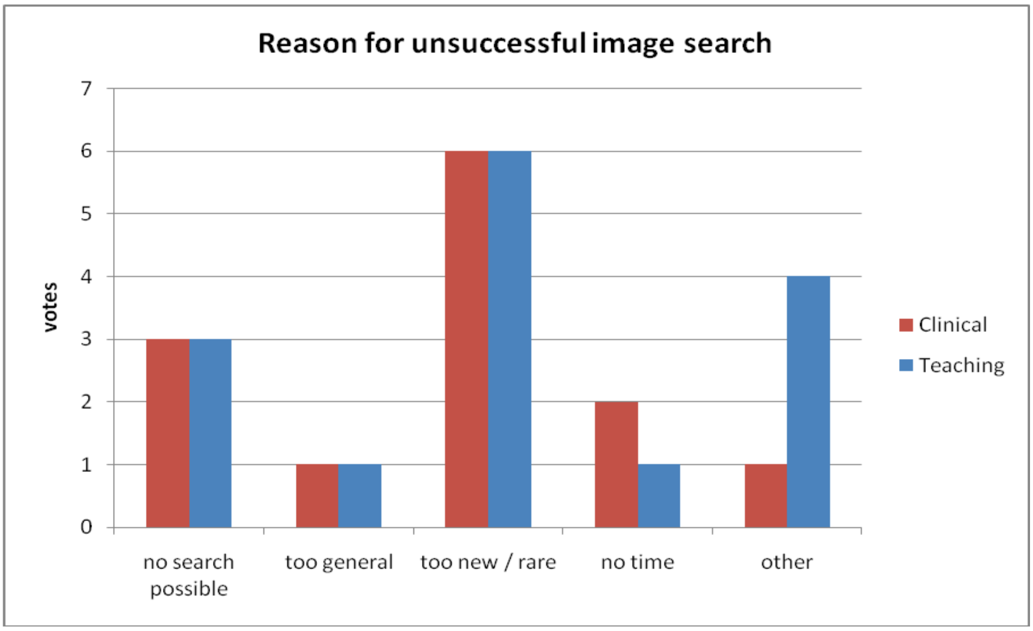


*Figure 9: Reasons for unsuccessful image search.*

**Question 5. When you find relevant images, how much time on average does it take you?**

When comparing the search times of successful image search it becomes overall clear that over 70% of successful searches finish after ten minutes or less (see also Figure 10). Only a few persons search longer to successfully find images. For unsuccessful image searches a few persons already abort after 5 minutes or less, but most often 10 minutes or even over 15 minutes are mentioned before abandonment. This highlights the important of image search in the workflow. It also highlights that there is room for improvement with optimized search tools to find relevant information quicker.

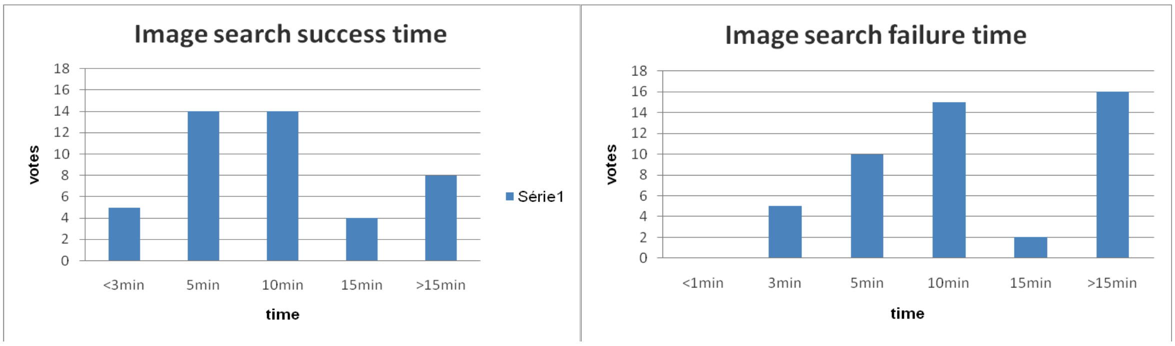


*Figure 10: time taken to find relevant images (left) and times before an unsuccessful search is abandoned (right).*

Figure 11 compares the search time for clinical work, teaching and research. It becomes clear that the length of successful searches 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 in general much larger and a significant part searches for 10 minutes or more and still finds relevant images.

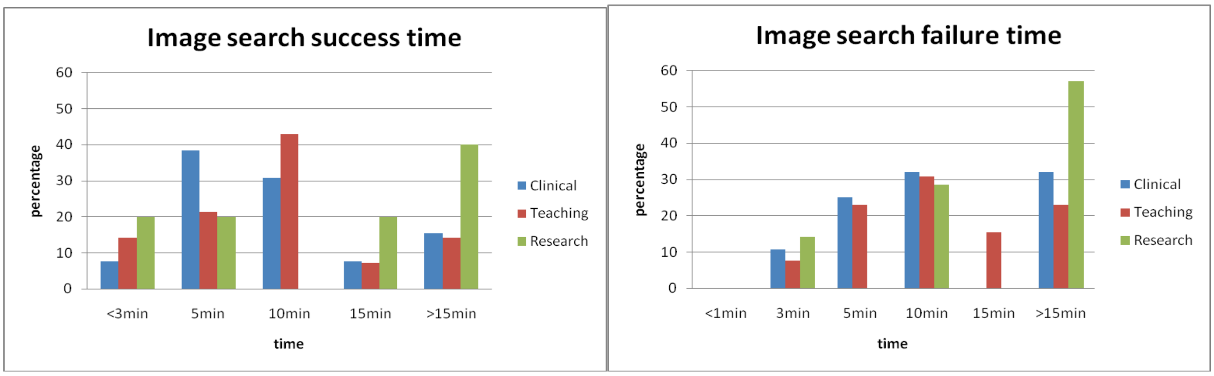


*Figure 11: time taken for successful (left) and failed (right) image searches compared for clinical work, teaching and research.*

For unsuccessful searches the times are similar for all three categories with research tasks most often taking more than 15 minutes. Compared to successful searches the searches generally take longer and a larger percentage of them take 15 minutes or more. As these unsuccessful search concern around one third of all searches it becomes 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 potentially lost.

**Question 6. What would be useful additions for an image search system (a sort of Google for medical images) and useful functionalities?**

In total, 15 of 26 persons responded to this questions, sometimes in several categories such as teaching, research and clinical work, with slight differences between the categories.
The functionalities that were most often suggested as useful additions for an image search system are search by pathology (13 times) and modality (10 times), followed by patient demography (6 times) and the option to search for similar images (8 times). Apart from the predefined options, the radiologists mentioned the need for multilingual retrieval and proposed several other additions, such as pathology classification (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 once as was the need to connect radiology images with histopathology or other criteria allowing to judge the confidence of a diagnosis.

**Question 7. What would a perfect search system for images look like?**

A total of 16 responses were obtained for desired input possibilities, and 20 for desired result formats. As there were no major differences between clinical work/teaching/research, the three were combined. The perfect search system for images should use images (with possible regions of interest) as well as keywords as input parameters. Keywords could include a wide variety of information, from describing the anatomical structure, the pathology and histology, up to more demographic information like the patient age (see Figure 12).

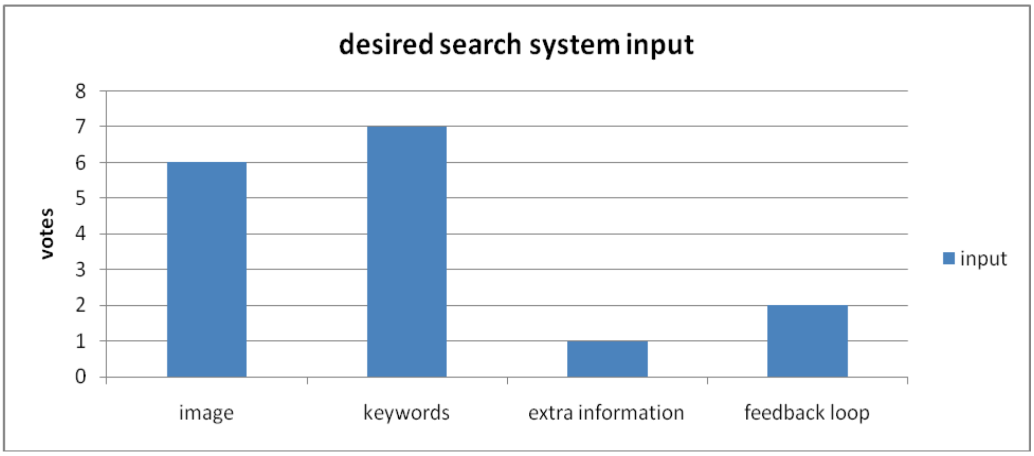


*Figure 12: Desired input data for a medical image search system.*

The output is desired to include several image examples and a corresponding description of the images. If available, differential diagnosis could be provided by the search engine. More detailed information including references would be helpful (see Figure 13).
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.

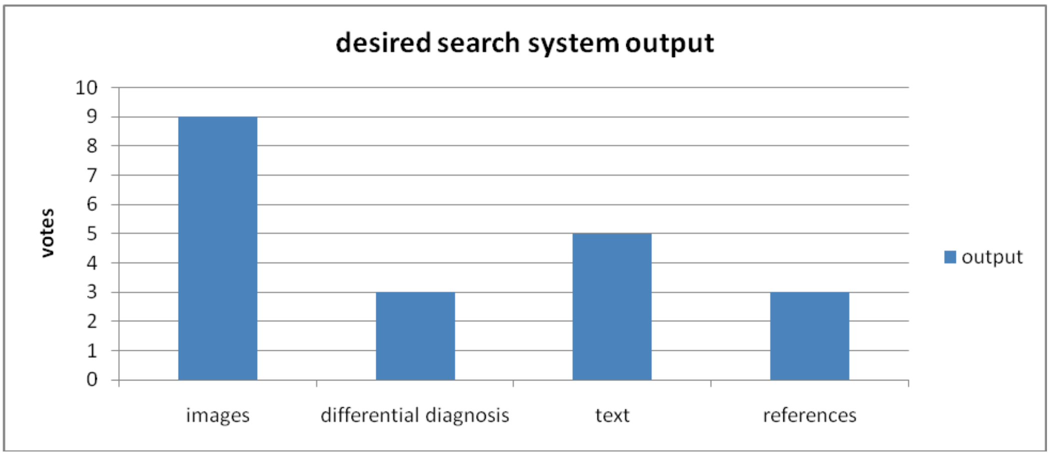


*Figure 13: desired output formats for a medical image search system.*

**Question 8. How could an automatic system exploit visual information for information search?**

There was a relatively small amount of feedback regarding the exploitation of visual information for information search, indicating that the radiologists are not very familiar with this concept, the state of the technology and research.
However, there were several interesting suggestions:

- the search for similar images and similar cases (mentioned by 4),

- search for similar regions of interest (mentioned by two persons),
- 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 also 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.

**Question 9. Tools for an automatic annotation of images using only the visual information in the images exist; what kind of annotation would be most useful?**

In total, 15 radiologists responded to goals of automatic annotation. 13 mention anatomic region as being important, 8 mention modality. 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.

**Question 10. Are there any radiology/medical terminologies that you use for searching or describing the images/cases?**

In general, only a small number of radiologists mentioned to use systematic terminologies for image search or image descriptions. Only 14 of 26 persons actually responded to the question and of these 7 mentioned to not use any terminology at all. MeSH (Medical Subject Headings) was mentioned most frequently with 7 persons using MeSH. Maybe the fact that MeSH is used by MedLine may be an important reason that this terminology is prevalent. There is no major difference in terminology use between clinical work, teaching and research. One participant mentioned to use RadLex for teaching in addition to MeSH. Two persons mentioned to use terminologies for diagnosis without detailing it, but most likely meant ICD (International Classification of Diseases).
There is a difference between countries in using terminologies. Austrian radiologists had a majority using MeSH whereas it is rare in the other countries.
The dominant PACS system mentioned among the radiologists is AFGA IMPAX (7 responses), with one participant each mentioning GE, Fuji Synapsis and Siemens Syngo.

## 7.2 The perfect search system

As question Q8 asking how visual information could be exploited is one of the most important for Khresmoi, we'll analyse the responses separately with a focus on the text given by the respondents.
It is clear that imagining the perfect search system can be hard when no example system is known to the persons surveyed. Still, besides the current image use and search behaviour several persons added comments about the perfect search system that they would like to see being implemented.
Most of the comments on perfect search are 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.

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 highlights the lack of knowledge about current research prototypes for visual search, but on the other hand also the importance of structured clinical data in the process that is important even for an image-based domain such as radiology. This means that even for visual search combinations with clinical data are important.

Some text search systems such as Goldminer and Yottalook are known to the clinicians and used by them. Similar to search engines such as Google that are 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.

# 8 Interpretation and conclusions

Based on past experiences it was clear that obtaining a very large number of responses from often very busy and overloaded radiologists would be difficult. In the end, 26 radiologists responded to the questions asked within the two months of the online survey. Most persons responding were rather junior in the age range of below 30 years and with less than 5 years of 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 themselves to be competent information searchers, which does not always fully correspond to the reality [12] - particularly in terms of how to use the information found.

There are several conclusions based on the responses that are relevant for the KHRESMOI project. They encompass insights into the role of image search in the clinical environment, the current search behaviour, and the limitations of the current search possibilities. In the following we list the most important aspects that together with the raw data in the preceding section should give a complete picture.

This analysis mainly takes into account the results from the survey. The analysis of the user observation including the eye tracking was mainly used for the construction of the survey form. The eye tracking allowed us to validate that very small regions of interest are what clinicians really focus their activity on and thus search by regions will be one of the important requirements for Khresmoi.

**Role of image search**

The search for images and comparable cases is an essential part in the radiologists' workflow. During the assessment of clinical data they typically use information from other images obtained from multiple sources: reference books, communication with more experienced 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 do not succeed in a substantial number of cases (around 35%).

**How and what to search for?**

Keyword search is currently the dominant search modality, including Internet search engines, and data access by patient ID in clinical records, or the oral communication with colleagues. During the

selection of results, experience plays a dominant role when analysing and choosing images. This indicates that substantial prior knowledge is necessary to perform efficient and successful search, and that the communication among colleagues is used to share this knowledge not only during training, but also in clinical practice. Past cases store experience of other colleagues and could make this experience available in a more systematic way. Trust in the information found and evidence for a particular diagnosis are mentioned as important points and this can most likely be more easily confirmed in direct communication with colleagues than when searching in other sources such as the Internet. The scientific literature has an advantage over general Internet sources in this respect.

Visual search is still little known although first prototypes do exist such as the IRMA (Image Retrieval in Medical Applications) and MedGIFT (Medical GNU Image Finding Tool) projects.

**Limits of current search systems**

There is clearly room for improvement considering the allocated time and the success rate of current image search. This would be consistent with the perception of the radiologists, who conclude that the obstacle for finding relevant images is not the availability of data but rather the limits of search technology or the novelty of the data searched for.

Keyword-based search is perceived to have limits due to the necessity of a rather accurate prior assessment of the present case before formulating a query, and a tedious selection of results based on individual inspection of every potentially ambiguous candidate image. This is especially 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 the reference database. Keywords are ambiguous and only the use of terminologies could help in this respect, which is currently rather rare. Among other strategies many radiologists resort to building their own personal reference data bases to compensate for search-ability in existing data bases, and search therein constitutes a significant portion of search queries. In many institutions the lack of institutional archives makes this sometimes necessary. Another way is to store patient IDs of interesting cases with short textual annotations in files that then allow finding cases in the future.

The dominant role of experience, the emergence of scattered personal reference databases and the culture of communication among colleagues suggests that facilitating the sharing of knowledge and removing requirements such as prior assessment and keyword identification would have significant impact on the radiologists' clinical work, teaching and research.

**Wishes for future systems**

Radiologists' suggestions for future search systems are consistent with the limits of current search systems. They name search for pathology as one of the top goals. They would value images equally to keywords as potential query inputs, and suggest the use of regions of interests in the query case to obtain more specific search results. Other simple functionalities include limiting the search to particular modalities and to include textual data into the visual search, which could be achieved by faceted filtering of the search results.

The idea of trust or confidence in a diagnosis of a case found was also mentioned to be important. Another important part is the linking search results and cases with the literature. The peer-reviewed literature does offer a certain level of trust.

One person also mentioned the use of 3D reconstructions for the search and also the visualization of results in 3D but generally the tomographic imaging data is only rarely mention in terms of a search target, most likely as the clinicians are not aware of the possibilities of this search.

**Next steps for Khresmoi**

The results of this survey are a first step towards a better understanding of the requirements for radiologists in handling images and searching for visual data that might help them in their daily tasks. The survey will remain online and new evaluations can be performed based on potentially more complete data. It also has to be mentioned that many radiologists are not familiar with visual search so being able to show them prototypes and having them work with the prototypes will most likely help them understand the problems and the potential and will also make it easier to formulate desires for a perfect search system. Such a test could be analysed with additional eye tracking ot see how actual systems are used in practice by radiologists.

After the first user tests in the Khresmoi project we thus plan to perform a similar survey among the participants. This might give us more concrete feedback regarding the existing system and for

planning the next steps.
All feedback obtained in this survey will be integrated into the prototypes as well giving us a precious source of information albeit on a small scale.

# 9 References

[1] A. V. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22(12), pages 1349-1380, 2000.
[2] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C. R. Shyu, A. Marchiori, Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis: System Description and Preliminary Assessment, *Radiology*, 228(1), pages 265-270, 2003.
[3] M. Markkula, E. Sormunen, Searching for photos - journalists' practices in pictorial IR, *The challenge of image retrieval*, Newcastle upon Tyne, UK, 1998.
[4] H. Müller, C. Despont-Gros, W. Hersh, J. Jensen, C. Lovis, A. Geissbuhler, Medical professionals' image search and use behaviour, pages 24-32, *Medical Informatics Europe, IOS press*, Maastricht, The Netherlands, 2006.
[5] H. Müller, J. Kalpathy-Cramer, W. Hersh, A. Geissbuhler, Using medline queries to generate image retrieval tasks for benchmarking, *Medical Informatics Europe 2008, IOS press*, pages 523-528, Gothemburg, Sweden.
[6] H. Müller, C. Boyer, A. Gaudinat, W. Hersh, 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, *Medinfo 2007, IOS press*, Brisbane, Australia, 2007.
[7] W. Hersh, H. Müller, P. Gorman, J. Jensen, Task Analysis for Evaluating Image Retrieval Systems in the ImageCLEF Biomedical Image Retrieval Task, *Slice of Life, conference on multimedia in medical education*, Portland, OR, USA, 2005.
[8] T. Beckers and N. Fuhr, User-oriented and Eye-Tracking-based Evaluation of an Interactive Search System, *4th Workshop on Human-Computer Interaction and Information Retrieval (HCIR 2010)*, 2010.
[9] G. D. Tourassi, M. A. Mazurowski, B. P. Harrawood, Exploring the potential of context-sensitive CADe in screening mammography, *Medical Physics*, 37:11, pages 5728-36, 2010.
[10] M. S. Atkins, A. Moise, R. Rohling, An application of eyegaze tracking for designing radiologists' workstations: Insights for comparative visual search tasks, ACM Transactions on Applied Perception (TAP), Volume 3 Issue 2, April 2006.
[11] H. L. Kundel, C. F. Nodine, E. A. Krupinski, C. Mello-Thoms, using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. Academic Radiology volume 15, pages 881-886, 2008.
[12] A. J. A. M. van Deursen, J. A. G. M. van Dijk, Internet Skills Performance Tests: Are People Ready for eHealth?, *Journal of Medical Internet Research*, 13(2), 2011.