A study on the relevance criteria for medical images

Shahram Sedghi [a, b], Mark Sanderson[a] and Paul Clough[a]
[a] Department of Information Studies, University of Sheffield, United Kingdom
[b] Department of Librarianship and Medical Information Sciences, Iran University of Medical Sciences, Tehran, Iran
{s.sedghi, m.sanderson, p.d.clough}@sheffield.ac.uk

Abstract

This paper reports the results of a study investigating the relevance criteria used by health care professionals when seeking medical images. Data was collected from 26 participants using a think-a-loud protocol and face-to-face interviews and analysed using the Straussian version of Grounded Theory (GT). Findings show that participants made use of 26 relevance criteria, although did not agree on the most important. Our findings suggest that users apply different criteria in different situations when evaluating the relevancy of medical images. In addition, we have investigated the coverage of relevance criteria to search statements from the medical track of ImageCLEF (ImageCLEFMed). Analysis indicates that some of the criteria identified by our participants could be included in new topics used for future incarnations of the medical image retrieval track.

Keywords

Relevance criteria, Relevance judgment, Image retrieval, Medical imaging

1. Introduction

The importance of digital images in domains such as medicine is great: digital imaging has become a vital component of a large number of applications within current clinical settings (Eakins and Graham, 1999; Glatard et al., 2004; Müller et al., 2004). According to Eakins and Graham (1999), medical images are utilized by a variety of users such as medical students, lecturers in medical departments, clinicians, etc each with different levels of subject knowledge. Access to images is commonly mediated through an electronic patient record system such as DICOM[1] or PACS[2]. Although research into the effectiveness of such systems is extensive, much of the research is based on a particular notion of relevance (i.e. topicality), and there is relatively little research into the criteria used by professional users who search for medical images as part of their daily work.

Researchers such as Armitage and Enser (1997), Markkula and Sormunen (2000), Greisdorf (2002), Cunningham and Masoodian (2006) and Tsai (2007) have investigated the needs and information seeking behaviour of users searching for images. However such studies have not addressed search of images in medicine. Similarly, although relevance studies have explored criteria elicited from the users of document retrieval systems (see, e.g. (Saracevic, 1996; Mizzaro, 1997)), the understanding of such criteria particularly for medical images, is limited. Previous work within the large body of user-oriented relevance has not addressed how health care professionals judge the relevancy of medical images for their information needs. Therefore, in this paper we describe a study, based on interviews with such professionals, to explore relevance criteria commonly employed for medical images. Specifically the research objectives of this study are the following:

- What criteria do health care professionals use to make relevance judgments when searching for medical images?
- What visual (pictorial) criteria do they apply, if any?
- What is the core relevance judgement criterion used to assess the relevance of medical images?
- Are these criteria different from those applied when searching text documents?
- What difficulties do health care professionals face when searching using medical image retrieval systems?

[1] Digital Imaging and Communications in Medicine

[2] Picture Archiving and Communications System

- What are the attitudes of health care professionals with respect to existing image retrieval tools and techniques?
- What is the coverage of the identified criteria in topics used in a large-scale evaluation of medical image retrieval systems (ImageCLEFMed)?

Section 2 of this paper describes past work on user-orientated studies to establish relevance criteria, Section 3 presents the research methodology employed by the study and Section 4 is devoted to our findings. In Section 5, we describe our analysis of ImageCLEF topics and Section 6 concludes the study.

2. Related Work

In the information seeking community, a great deal of research has been conducted on the notion of relevance and formulating suitable definitions for it. We discuss this work in Section 2.1, and in particular summarise relevance in the area of image retrieval in Section 2.2. Section 2.3 is devoted to the information needs of health professionals.

2.1 Relevance Criteria

According to Saracevic (1996) there is no common agreement between IR researchers on a single definition for the concept of relevance in IR, even though "*nobody has to explain to users of IR systems what relevance is, even as they struggle (sometimes in vain) to find relevant stuff. People understand relevance intuitively* (*page. 13).*" Despite disagreement over what criteria are actually used in judging relevance, researchers are in general agreement that two main categories of relevance exist and can be distinguished: topical relevance and situational relevance (Schamber et al., 1990). Topical relevance is perhaps the most widely used when performing system-orientated evaluation of document retrieval systems. In this view, a document is relevant to an information need (expressed as a query) if some or all of its theme or topic overlaps the query. The second category - situational relevance - incorporates the user's situation or context into judgements with the assumption that the situation in which information seeking is performed cannot be separated from judging relevance.

Schamber (1994) analysed related literature regarding relevance, dating from 1960 and produced a list of 80 criteria which likely influence relevance judgements. She classified these into six groups: attributes of the person making the relevance assessor (e.g. knowledge and experience), queries or topics, documents, the information retrieval system, judgment conditions and choice of scale. She also believed the process of judging relevance to be a dynamic phenomenon and based on several criteria such as informativeness of documents, personal knowledge of the end-users of information retrieval systems, credibility of the source of documents, etc. Schamber (1994) concluded by defining three classes of relevance: one system-oriented view and two user-oriented views. One relates to information and the other one relates to the users' situation. She named these three approaches as objective, subjective and situational views, respectively.

Mizzaro (1997) analysed 157 papers published since 1959 and classified them within three periods: "before 1958", "1959–1976" and "1977–present". He analysed papers within each time period regarding the following aspects: methodological foundation, kinds of relevance, beyond-topical criteria adopted by users, modes for expressing the relevance judgment, dynamic nature of relevance, type of document representation, and agreement among different judges. Mizzaro (1997) concluded that the focus of papers published in the "1959–1976" period had been on relevance inherent in the document and query (topical relevance). However in the "1977–present" period, researchers have attempted to understand, formalize, and measure a more subjective, dynamic and multidimensional relevance judgment (situational relevance).

2.2 Image Relevance Studies

There are few studies which investigate relevance judgements for visually-orientated documents. Markkula and Sormunen (2000) investigated the relevance criteria typically applied by journalists when selecting images for tasks in realistic work situations. The authors interviewed eight journalists who were given twenty illustration tasks based on searching for images in the Aamulehti digital image archive, a collection containing over 83,000 news agency photos. Markkula and Sormunen (2000) classified relevance criteria described by journalists into four groups: topicality, technical, contextual attributes and visual attributes. With topicality, journalists used captions to assess topical relevancy of images to obtain information about a relevant image and its background. Technical and contextual attributes were the most common criteria used, with most journalists preferring to find "*images which were technically good, not recently published and current"* (Markkula and Sormunen, 2000: 277). Further important factors included the financial cost of an image, recency or freshness, and certain visual attributes, e.g. journalists requiring images in a particular style such as a passport photo of a specific person. In addition, journalists would often use the message they wanted to convey through an image as relevance criteria, e.g. dramatic, surprising, effective, shocking, funny, expressive or threatening.

Choi and Rasmussen (2002) conducted a study to examine relevance criteria before and after searching for images. Thirty-eight graduate students of American History and faculty members participated in the study. All of them looked for images from the 'American Memory' online image collection and were asked to discuss how they evaluated relevant images. Since previous studies indicated a significant overlap between the criteria applied by end-users, Choi and Rasmussen (2002) offered a list of nine common criteria from those studies to participants and asked them to rate the criteria regarding the importance of each for their information needs: topicality, accuracy, time frame, suggestiveness, novelty, completeness, accessibility, appeal of information and technical attributes of images. Participants were then asked to search for images and evaluate retrieved images using these criteria. They were also asked to list other relevance criteria which they might apply. Before starting the search, participants were asked to rank the nine criteria: topicality, accuracy and completeness were rated as the top three. Once participants had seen the retrieved photos, they were found to apply criteria relating to aspects such as time frame and accessibility of the photos. The authors stated there was a significant difference in the ratings of each criterion before and after users saw the images.

In a preliminary study, Hung et al., (2005) investigated the relevance criteria applied by ten students of Journalism and Media Studies. The aim of this study was to elicit what criteria searchers employed to select relevant images. Hung et al., (2005) asked participants to look for images based on three pre-defined image search tasks, including those deemed as specific, general and subjective:

> ***Specific****: you are photo editing a story on Tiger Woods for a sports magazine. For this story, you need to find some photos of Tiger Woods as illustrations.*
>
> ***General****: you are photo editing a report on the crisis in the Middle East for a newspaper. For this report, you need to find some photos regarding this topic to be used as illustrations.*
>
> ***Subjective:*** *you are photo editing a special report on the topic of "Peace" and you need to find some photos to illustrate the meaning of "peace."*

Based on the findings of their study, Hung et al., (2005) identified several relevance criteria applied during the three image search tasks: typicality, emotion, action, aesthetic appearance, text, familiarity, context, impression, preference, posture, facial feature and appearance. They found that typicality, emotion and aesthetic appearance were the three most important criteria,

applied across all three tasks, where typicality was deemed the most important criterion for all three tasks (according to the authors, typicality is a criterion that can exhibit universal representation of an object in a photo).

2.3 Information needs of health care professionals

Lancaster (1979); Barry (1994); Mizzaro (1997); Barry and Schamber (1998); Tang and Solomon (1998) and Choi and Rasmussen (2002) reported that the end-users of IR systems decided the relevance of retrieved documents based upon their particular information needs.

In a number of articles the information needs of health care professionals has been investigated. For example, Shelstad and Clevenger (1996) examined the needs and information seeking patterns of ninety-nine general surgeons. Results showed that "patient care", "continuing medical education", and "casual curiosity" were the most common purposes for seeking information. In addition surgeons declared that "professional meetings", "the medical literature", "colleagues", and "continuing education courses" were the main sources for their information needs.

In an observational research study, Ely et al., (1999) studied the information needs and information seeking behaviour of 103 family doctors. Ely et al., (1999) stated doctors mostly required information to respond to questions such as "What is the cause of symptom X?", "What is the dose of drug X?" and "How should I manage disease or finding X?" The results of the study indicated that textbooks and colleagues were the primary source of answers to patient care questions; formal literature searches in medical databases and the Internet being rarely performed. The authors suggested that the information needs of doctors had a direct relation to their search task and effect on their relevance judgments. For example, Ely et al., (1999) reported that when doctors were faced with a clinical problem, often they tended to ask questions such as How should I manage disease or finding X? Authors added that doctors needed quick and "bottom line" answers to their questions.

To the best of our knowledge, there have been few previous studies specifically addressing the relevance criteria employed by health care professionals when searching for medical images, although existing studies have investigated their information needs. In a qualitative study, Hersh et al., (2005) examined the pictorial information needs of 13 biomedical professionals with various roles including clinician, researcher, educator, librarians and student. Results of study showed that medical image needs of biomedical professionals can be categorized into 4 groups: research-related, patient care-related, education-related and other.

Paling et al., (2006) investigated the image information needs of 34 dental faculty members and clinicians. They reported that participants looked for images in a variety of sources such as search engines, personal collections, digital textbooks, digital journal articles, database and CD/DVDs. The authors reported that a substantial number of the participants preferred to find and use digital images, and that none of the participants indicated an overall preference for physical slides. Paling et al., (2006) suggested that online dental image collections could be a good match for the participants' dental image information needs. Participants also preferred to access higher quality, manipulable images and metadata schemes for describing the content of images such as the name of a disease or injury.

As can be seen from the past work cited above, few studies have investigated the relevance criteria used in image retrieval compared with studies of document retrieval. In addition, there have been few studies concerning the medical image information needs of health care professionals. In particular, little work has investigated in detail the relevance criteria used by professionals in the field of medical image search, hence we were motivated to conduct such a study, which we now describe.

3. Methodology

Maglaughlin & Sonnenwald (2002) state that relevance criteria gathered from participants can be affected by the particular research method used; therefore we saw the selection of an appropriate research method as a vital issue for this study. Park (1994) maintained that qualitative research methods can capture an in-depth understanding of the complex and dynamic concept of relevance in IR, and therefore we also used a qualitative research method. Suitable tools and techniques for data collection and data analysis were based on using Grounded Theory (GT), specifically the "Straussian version" of GT.

According to Glaser & Strauss (1967), Corbin & Strauss (1990) and Glaser & Holton (2004), a grounded theorist must identify the following elements of theory in order to establish it: concepts, categories and hypotheses. Concepts (sometimes referred to as variables) are the basic units of data analysis in GT (Corbin & Strauss, 1990). In GT, any concept involved or discovered in the study is considered as provisional. The importance of variables in the induction of the theories relate to the repeated presence or absence of each variable in interviews, observations and documents. Categories are the second elements of GT. Corbin & Strauss (1990) state that *"categories are higher in level and more abstract than the concepts they represent"* (Corbin & Strauss 1990:9). The third element of GT, hypotheses, clarifies the relationship between a category and its concepts and between separate categories.

3.1 Data Analysis in the GT

The different phases of coding paradigm in Straussian GT can be described as follows:

Open coding: the process of identifying and naming categories from the written data, from field notes or transcripts of interviews. This is the first stage of data analysis in GT and the intention of this stage is to reveal fundamental characteristics of the phenomenon under investigation. In the open coding process, the data is broken down and conceptualized and variables such as events/actions/interactions related to the phenomenon are recognized.

Axial coding: Strauss & Corbin (1998) described axial coding as *"the process of relating categories to their subcategories, termed 'axial' because coding occurs around the axis of a category, linking categories at the level of properties and dimensions" (Strauss & Corbin, 1998: 123).* At this stage, researchers put similar codes together (categories and subcategories). The aim of this stage is to identify causal relations between categories and generate a paradigm model, which illustrate the relationship between categories.

Selective coding: In this stage, researchers integrate categories that have been developed in order to identify the core category which forms the hypothesis. In fact, this category is the central phenomenon that requires to be speculated about. The core category systematically relates to other categories and represents the main phenomenon around which all other categories are based. In selective coding researchers validate the relationship between categories and then refine the categories in case of need. In this stage researchers focus on coding data related to the core category and its subsidiary categories and they are required to establish a hypothesis based on the core category they have identified.

3.2 Data Collection

Using a qualitative study, GT was applied to identify and describe relevance criteria applied by health care professionals and their perceptions for medical image retrieval systems. Our aim was to capture the relevance judgment process and relevance criteria from participants' own words, therefore we utilised semi-structured interviews and think-aloud protocols for data collection. Twenty six health care professionals participated in our study, with the mean average of 13 years and 3 months work experience (range 2 to 35 years). Ethics approval to carry out the study was obtained from the NHS[1] National Research Ethics Service and Sheffield Teaching

[1] National Health Service in the United Kingdom

Hospitals NHS Foundation Trust in order to recruit and interview participants. Health care professionals were sought who used any kind of medical image in their daily work, were skilled and knowledgeable computer (and Internet) users, and held a degree in health or in bio-medical sciences. To recruit suitable participants, letters of invitation were distributed via email to subscribers of a Sheffield-based health and biomedical mailing list and by traditional postal services. Interested respondents were then selected based on their suitability for this study. A local research contact from Sheffield Teaching Hospitals NHS Foundation Trust facilitated our access to interviewees. The mean interview duration was approximately 43 minutes (range 28 to 92 minutes), and with their permission, we recorded the whole interview during the medical image search sessions.

P	Speciality	AD	IL	WE	Participants Roles
1	Dental Materials	PhD	50	4	Research and Teaching
2	Molecular Genetics (Genetics and Immunology)	PhD	57	6	Clinical research
3	Orthopaedic Surgeon	PhD	92	12	Clinical, Teaching and Research
4	Immunologist, Molecular Immunology	MSc	58	4	Research
5	General Surgeon	MD	47	5	Clinical, Research
6	Sport Medicine/Consultant of Orthopaedic Surgery	PhD	45	21	Clinical, Teaching, Publication and Research
7	Stem Cell	PhD	37	2	Clinical, Teaching and Research
8	Molecular Medicine and Female Infertility	PhD	51	7	Clinical, Research
9	Bone metabolism	PhD	53	17	Research and Teaching
10	Superintendent radiographer	PhD	38	16	Clinical, Teaching and Research
11	Virologist	PhD	43	3	Research
12	GP, Non-clinical lecturer	MD	33	22	Teaching and Research
13	Neurology	PhD	38	5	Clinical, Teaching and Research
14	Medical physicist	PhD	33	15	Clinical and Research
15	Radiologist	PhD	42	16	Clinical, Teaching and Research
16	Nuclear medicine	PhD	28	17	Clinical, Teaching and Research
17	Medical physicist	PhD	42	7	Teaching and Research
18	Nuclear medicine	PhD	38	11	Clinical, Teaching and Research
19	Medical physicist	PhD	35	35	Clinical, Teaching and Research
20	Consultant Haematologist	PhD	31	25	Clinical, Teaching and Research
21	Obstetrician Gynaecologist	FRCOG	34	10	Clinical and Research
22	Gynaecologist	MSc	39	7	Clinical and Research
23	Haematologist	PhD	28	18	Clinical, Teaching and Research
24	Obstetrician Gynaecologist	PhD	36	30	Clinical, Teaching and Research
25	Neurology	PhD	30	2	Clinical and Teaching
26	Human reproduction and development biology	PhD	29	19	Teaching and Research

Table 1. Profiles of participants (P: participants, AD: academic degree, IL: interview length and WE: work experience years, FRCOG: Fellow of the Royal College of Obstetricians and Gynaecologists)

Unlike previous studies (i.e. Markkula and Sormunen (2000), Choi and Rasmussen (2002) and Hung et al., (2005)), we did not focus on a particular image collection or image retrieval system; we did not ask participants to look for images for any predefined image retrieval task; and we did not provide a list of criteria for our participants. By contrast, participants were asked to specify (and conduct) medical image searches as typically carried out in their day-to-day activities. Example topics included the following: a pathologic image of a biopsy of bone marrow, MRI images of the meniscus tear in the knee, Microscopic images of cartilage injuries in children, X-Ray of Monteggia fracture, Interleukin-4 Electrophoresis diagram, microscopic photo of Interleukin-4 proteins derived by mast cell and images of an anencephalic embryo. Two participants did not have access to the Internet at the time of the interview, and so were simply asked to describe their searches and the relevance criteria they had applied.

3.3 Example coding

Using Straussian GT, we analysed the interview transcripts to allocate every line or paragraph a concept label. The following example illustrates the open coding process:

> *"I prefer to use images of my previous patients who gave me their consent. I mean I prefer real images. I tend to look at those images more than searching for images in books. If I could not find them I go through Medline articles or medical organizations' websites like Sport Medicine Association, Cartilage Association and Joint Association websites. For example, if I am looking for images about meniscus tear in the knee I go to a relevant site or a journal site and I will definitely find what I need." (P3)*

The previous paragraph (and associated part of the interview) illustrates a range of emerging phenomena. By referring to *"I prefer to use images of my previous patients who gave me their consent"* the participant states that he uses his personal image collection. The participant then explained he prefers *"to look at those images more than searching for images in books."* In addition he stated that he would use image sources other than books to find images: *"If I could not find them I go through Medline articles or medical organizations' websites like Sport Medicine Association, Cartilage Association and Joint Association websites."* Finally the statement *"For example, if I am looking for images about meniscus tear in the knee I go to a relevant site or journal site and I will definitely find what I need"* implies that the participant looks for medical images in sources such as specialized journals or websites. Therefore, the following concept labels were derived:

- The participant has a personal image collection containing images from his previous patients, and he prefers to use images from his own collection.
- The participant uses images from published articles.
- The participant looks for images in related websites.

Following the coding paradigm of Strauss and Corbin (1990), we compared concept labels to establish any relationship between them. At the next stage, as the following example illustrates, we put related concepts labels together to form an open-coded category for sources of images:

- P3: *"I prefer to use images of my previous patients who gave me their consent."*
    - Concept label: Image sources/ Personal collections
- P2: *"I also use books to find images for example if you want to know and see the mechanisms that already have been discovered, you can use the books."*
    - Concept label: Image sources/ Books
- P1: *"But for the special images I go to find them in papers."*
    - Concept label: Image sources/ Papers

These example concept labels indicate that participants looked for images in personal collections, books and papers and we therefore classified them under the category of "image resources" during axial coding. Then we continued with the analyses to identify the properties of each concept label classified under categories like "image resources". The following examples illustrate this process:

Axial Code: Image resources

Concept label: Web

Properties of the concept level:

Property 1- Using the Internet helps participants to find images quickly

"Using the Internet saves time as you can find images quickly." (P2)

Property 2- Participants may find many irrelevant images in the Internet.

"The Internet is available everywhere, but you have to change or modify the query term and see many pages of images to find the ones that you need. As you know, there are lots of irrelevant images that do not fit your needs." (P3)

Property 3- Participants believe that they may find required images on the Internet.

"I also believe that through the Internet and image search engines, it is more likely that I find the images I need." (P4)

As stated earlier, the focus of the work reported here was to identify relevance criteria for medical images applied by health care professionals. In total, 26 different criteria were mentioned by participants and using GT and its coding paradigm, we could identify the core relevance category to address our initial aims and objectives. In this study, the core category was identified as "visual relevancy" and we concluded the selective coding stage of the GT approach and we made a statement as follows: health care professionals used a variety of relevance criteria to judge the relevancy of images to their information needs and there was no agreement between them on the most important criterion. However, all participants regarded the visual relevancy of medical images as the most frequently used criterion, and amongst them 12 participants stated that they considered the visual relevancy of images as the *most important* criterion.

4. Relevance criteria for medical images

In this paper we listed relevance criteria regardless of the role of the participant, and the activities carried out in their associated departments. All participants were aware of legal issues and the protection of patient privacy. Independent research monitoring officers from The University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust monitored the study to ensure that neither researchers nor participants breached the rules. We noticed that participants could access anonymous images through the health information systems they used, and in addition used images in different ways for a variety of reasons (e.g. viewing images from medical websites or electronic journals for clinical purposes). Some participants used patient images though only if they had written consent from the patient.

Criteria	Freq.	Criteria	Freq.	Criteria	Freq.
Visual relevancy	**26**	Appropriateness	**13**	Simplicity	**7**
Image quality	**22**	Clarity	**13**	Colour	**6**
Background information	**21**	Credibility	**12**	Informativeness	**6**
Age and Gender	**19**	Image understanding	**11**	Copyright	**4**
Size(dimensional)	**18**	Technical information	**11**	Availability	**3**
Modality	**17**	Didactic value	**8**	Recency	**3**
Components of the image	**15**	Targeted audiences	**8**	Suggestiveness	**3**
Diagnosis	**15**	Anatomic region	**7**	Originality	**2**
Orientation	**14**	Magnification	**7**		

Table 2. Relevance criteria employed by participants (Total=26); the number indicates the number of participants who applied each criterion.

In total, 26 relevance criteria were elicited from participants, as shown in Table 2, together with the number of participants who specified each criterion. Visual relevancy, background information and image quality were the three most frequently used relevance criteria.

The criteria were classified into four main groups: Medical, Visual, Textual and Other. Medical criteria relate to medical information conveyed in an image or its associated text, such as caption or annotation. Visual criteria are associated with visual/photographic attributes of an image, such as orientation, image quality, magnification and size (dimensional). Textual criteria relate to text attached to an image conveying non-medical information. We also established the 'Other' category to group five criteria that were not easily attributable to the previous three categories.

It was observed that certain criteria could be classified under more than one category. For example, anatomic region was considered as a medical criterion, however most participants identified the anatomic region illustrated in medical images based on textual or visual attributes of images. Therefore, based on the overlaps, we established three intersecting criteria: Textual-Visual, Textual-Medical and Medical-Visual- Textual which are detailed in Figure 1.

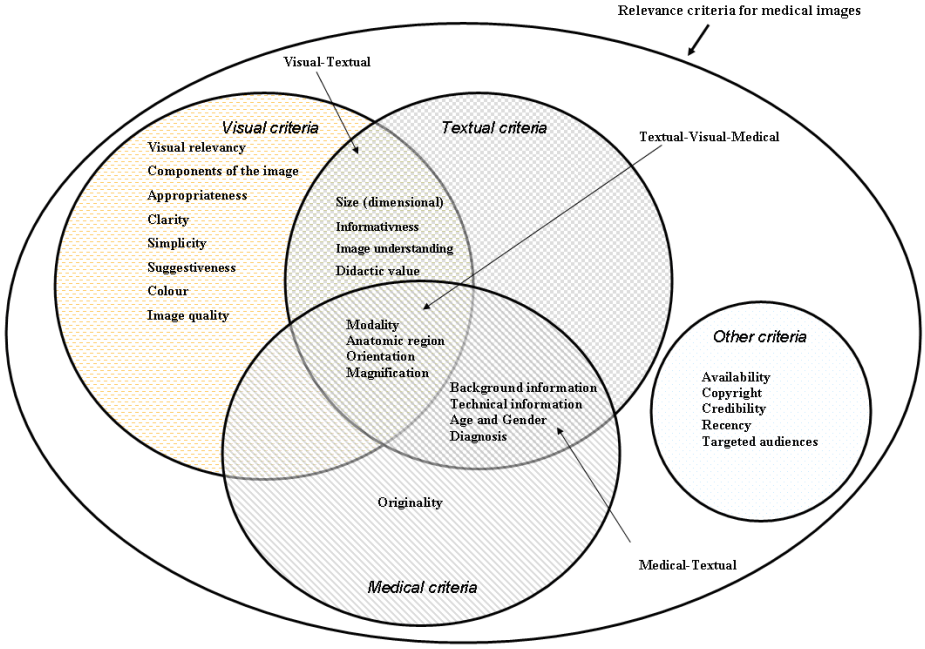

Figure1. Groups and subgroups of relevance criteria we identified.

Visual-Textual is the intersection of textual and visual criteria; Medical-Textual includes those criteria could be classified in both Textual and Medical groups; Textual-Visual-Medical includes criteria from textual, visual and medical categories.

Based upon the findings, participants would appear to apply certain criteria to compare images which were visually relevant to their information needs, e.g. the quality of an image is regarded as important if derived from a printed publication. Similarly, participants applied certain criteria regarding the sources used, e.g. credibility was considered an important attribute when looking at images online. Section 4.1 describes the criteria, arranged into the four main groups and ordered by the number of participants who used them.

4.1 Visual criteria

As stated earlier, visual criteria arose when the participants considered the visual/photographic attributes of an image in making relevance judgments. They are visual relevancy, image quality, size (dimensional), modality, components of an image, orientation, appropriateness, clarity, image understanding, didactic value, magnification, simplicity, colour, informativeness, and suggestiveness.

Visual relevancy

Visual relevancy refers to whether an image is relevant or not based on its visual content (and appearance). This criterion is in part dependent upon the ability of being able to recognise images visually similar to those which were previously relevant and recognise the meaning of visual content, e.g. the visual appearance of human anatomy, cells and organs:

> *"It visually illustrates what I want and that's the most important thing. There is something in my mind and I want something very similar to it. Yes, I knew what I was looking for so I knew what I would want to see."(P16)*

> *"Mostly what I want is something that I have seen before or I know what it is. …I know how most things work … that is all you need to be a haematologist."(P20)*

> *"Something that gives a visual representation of things you are trying to show. Yes, there is some sort of visual memory. …I mean visual memory of something that illustrates what I am after as much as possible."(p23)*

This criterion was identified by all participants in the study.

Image quality

The quality of an image (e.g. image resolution, quality of printed images, contrast and brightness, etc) was applied by 22 participants. However, they would check the quality of images only if they were visually relevant. Some participants emphasized that if the quality of medical images was poor, they were unable to use them:

> *"Because it has to be seen in the presentation, so I think the image quality, the actual resolution of the image, would be important." (P10)*

Size (dimensional)

Often participants looked for medical images to illustrate their publications and presentation; therefore they wanted to make sure images were an appropriate size. Participants stated that they might change the size of images using graphics editing software, such as Adobe Photoshop. However, they preferred to find images in the desired dimensions or size and believed that changing the dimensions of an image affects its quality:

> *"I usually select 3 or 4 images and then compare them to see which one will fit. When I compare them, I will consider the size and how clear the image is if I want to use it for a presentation or in printed material. Some of the images are very big in size, so I will reduce the size of them to fit. If you reduce the size sometimes the quality will be degraded." (P11)*

Modality

There are different types of medical images such as X-Ray radiograph, MRI, microscopic images, diagrams or general photographs. Many of the participants wanted to search for specific types of medical image and regarded the type of medical images as a criterion therefore we used the word 'modality' to describe this criterion. Seventeen participants used this criterion when

searching for medical images, and amongst them 8 considered modality as the most important criterion. For example, P16 looked for T2 weighted MRI scans and he regarded the modality as a key criterion to distinguish between relevant and irrelevant images.

Components of an image

When participants searched for images, they often required specific visual details in images. Sometimes participants used this criterion to distinguish between relevant and irrelevant images:

> *"Why are the other images irrelevant? That image is a nice image. It shows the structure of a molecule and I'm not interested in the structure for this context - the paper I'm writing is not about the structure." (P9)*

Orientation

Participants described orientation as the view or overall visual appearance of objects in an image. It depends on the location and the direction of imaging devices when producing an image. Participants stated that if an image was visually relevant to their information needs, they would be interested in the orientation and they would consider that image as a candidate image for their information needs. Moreover, participants reported that there are some standard and predefined orientations, such as Sagittal and Coronal sections, for taking medical images. Therefore, they know which orientation is best for their needs:

> *"What do you want to be seen? If the tumour is on the edge of the bone, you consider the angle." (P9)*

> *"… in MRI images there are only trans, sagittal or coronal views and it depends on what you are looking for and what you want. I mean we never look at MRI images of lung from a coronal view and sagittal because it will be too small."(P17) [This participant showed an example and explained that sagittal MRI images of the lung do not present an appropriate view for the user].*

Appropriateness

Appropriateness was a criterion mentioned by participants to verify that selected images were appropriate to the context of use and illustrated their topic properly:

> *"I want the image to state clearly the point of the argument or whatever I'm trying to get over. If it is a seminar or if it is a paper, I would put in an illustration." (P9)*

> *"We do a lot of lectures and presentations and representative images are always very useful." (P15)*

> *"Obviously the most important thing is what the image actually shows - is it a good example of what I want? But after that, if you have a choice of a few images, you might pick the one which is the best one for the particular context, e.g. a presentation would require an exemplar." (P16)*

Clarity

In response to "what do you mean by clarity?" two participants stated:

> *"That the topic of the image is not hazy and the object of interest is not small …the object is in the middle of the screen and showing all the characteristics which I am looking for."(P21)*

"If I want to select one of these images as relevant, I will probably choose this one because I think it gives me a general picture, and is visually clear... there is not too much detail in it."(P9)

Image understanding

Mostly, participants preferred to look at and select images they understood. They reported that without understanding what an image illustrates, they were often unable to make accurate relevance judgements. Sometimes they used the text attached to images to help understand and interpret what an image showed.

"You need some information attached to images to describe what it is. In some of the images we produce as the results of our findings, you will not understand it unless you access detailed contextual information. There are some images that you can recognize and understand, but we mostly need background information to understand and interpret images … for instance each colour shows certain molecules or molecular structure and components." (P26)

"If you saw the photo and you didn't know what it showed, you should read the text …the photo and the text caption should be self explanatory … if you saw the photo and you didn't know what it showed you should be able to read the text and understand the image." (P6)

Didactic value

Sometimes participants sought medical images for educational purposes, thus they considered whether the image was educationally useful or had didactic value. Although some participants emphasized that an image can be more influential than text for educational purposes, they believed that some images effectively conveyed educational topics better than other images.

"If I am going to show a specific injury in shoulder to my students, obviously the image should illustrate that clearly."(P6)

"Last time, I gave a talk on magnetic nerve stimulation and you can see there is a web version of this talk [he showed it to us]. I use images because a picture is worth a thousand words. It is much easier to communicate with images than with words and you can show students a slide that words are on it. They will get bored quickly. If you show them pictures, they absorb the information much more quickly." (P19)

Magnification

For those participants who looked for images at a microscopic level, magnification played an important role:

"Size and magnification are very important especially for images at a microscopic level. In fact it is a major requirement. You can't show a microscopic image without giving information on the magnification." (P8)

Simplicity

Sometimes participants used a certain criterion to compare a relevant image to another relevant image. For example, some participants preferred to select "simple" and uncluttered images. They believed that an image should not be busy or ornate:

"It shouldn't be complex with a lot of information in it. It should be simple and not overcrowded." (P5)

"If I want to select one of these relevant images I will probably choose this one because it gives me a nice general picture and is visually clear… you can understand what it shows...there is not too much detail in it."(P9)

Colour

The colour of the images, including colouring techniques and composition, was also a criterion identified by the participants. They thought that colour could help show differences of different parts of an image. For example, one of the participants, P3, stated that he could show certain type of proteins using green staining (colouring technique). He added that he could differentiate between protein X and Y using the colouring technique utilized in the images.

"Sometimes colour is important … the colour of an image can effect the selection of an image … sometimes the colour of an image can help you to see the differences." (P8)

"It is a representative image for what I am looking for …it is coloured …so I can easily understand and distinguish between the different parts." (P22)

Informativeness

When participants wanted to select images, they wanted images to provide information on their topic:

"If there were some similar images … the information in the image is the most important criterion … what an image shows and what I gain through looking at the image ... I will obtain the information via looking at the image … the image will serve my information needs." (P8)

Sometimes they emphasised that an image can convey more information and more quickly than textual documents:

"I want the image to state very clearly the point of the argument or whatever I'm trying to get over. …I would be able to put in an illustration that says a lot." (P9)

Suggestiveness

Sometimes participants expressed that an image was selected because it would suggest to them new ideas. Therefore, three participants stated that when they look at images they would consider the suggestiveness of images as a criterion related to the visual attributes of images:

"I will use an image if it provides me with new information or gives me new ideas." (P8)

4.2 Medical criteria

Medical criteria are related to a combination of attributes such as the participants' domain knowledge, information needs, and visual/textual attributes of an image, e.g. modality, anatomic region, age and gender. We ordered medical criteria based on how many participants applied each criterion when assessing images.

Age and Gender

The age and gender of the case (i.e. the patient) was a common criterion and participants mainly obtained such information from the textual information attached to images. For instance, one of the participants stated *"diseases of the children are different from the adult."(P5).* Age is an

important criterion for some medical conditions and sometimes is included in the query text itself, such as "epiphyseal closure", as participant 3 explained:

> *"Sometimes I want to look for things which change with age ... the age range is also important … if I want an image of epiphyseal closure I will put the age 18 [as part of the query] because it is common in the age 18." (P3)*

The gender of the patient in the case was also important for some interviewees when selecting an image. Some of them reported that gender, together with age, could be the most important criterion. In reply to the question "which one is the most important criterion: Age or Gender?" A participant said:

> *"… some injuries affect adolescents and some the older ones … if you have a specific injury it doesn't matter if you say male or female but you have different injuries in a 12 year-old compared to someone older. So gender is not so important in that sense … also Anterior Cruciate Ligament of the knee was common in men [Football and Rugby players] in 1960s, but now it is common in women because they started playing football and rugby."(P6)*

Diagnosis

Some interviewees highlighted the importance of medical diagnosis whilst selecting an image, generally found within the text associated with an image. Some participants declared that they preferred to access the comments and diagnoses for similar medical conditions from other health care professionals. Therefore, diagnostic information was regarded as an important criterion for clinical purposes:

> *"Sometimes you need to know the comments from other health professionals on a case [medical condition] which is similar to your case [a patient's medical condition] … so the information on clinical diagnostic of the case will be more important than criteria like gender or age." (P2)*

Another participant highlighted the importance of diagnosis for rare medical conditions:

> *"The text will be helpful if it was added by health professionals. Let's say we have a rare [medical] case and you have not seen it in the books or x-rays before. You can go to a website or articles and find an image of that rare case and compare it with the image that you have." (P3)*

Anatomic region

Sometimes participants looked for medical images where human tissues and cells, or a specific anatomic region, were illustrated:

> *"Definitely the atomic region is important. I mean in this case all images are supposed to be ultrasound images of ovaries." (P21)*

> *"This image doesn't show what I want. The other images are also irrelevant. Because one of them illustrates endometrial sections and the second one shows the stromal part." (P22)*

Originality

This is a criterion mostly used by Participants 14 and 15, who worked in a Medical Physics Department. Since they wanted to analyze the content of actual medical images using techniques they had developed, they wanted original versions of the medical images, i.e. without manipulation:

4.3 Textual criteria

Participants also considered a range of non-medical textual criteria when judging the relevancy of medical images.

Background Information

Some participants declared that they required background information such as medical history, origin of images, etc in the text associated with an image in order to understand it. They also declared that they needed background information for an image to clarify the problem or object(s) represented in the image. Therefore, they assessed the relevancy of images regarding the background information they could obtain from the textual information such as captions, annotations, etc. Although we established background information as a criterion under textual criteria category, we noted that there is an overlap between this criterion and criteria such as diagnosis and technical information.

Technical Information

This criterion typically relates to information included in the text associated with the images, such as annotations. Participants applied this criterion in situations where they needed information on materials, methods, the laboratory situation in which an image was taken, phases and stages of a test or a mechanism, and stages prior or subsequent to the stage presented in the image. Participants stated that the criterion was important and required to obtain supplementary information, including technical information, for a better comparison of their work (and research findings) with prior cases. This criterion was used by four participants.

4.4 Other criteria

Participants also considered other criteria that we could not group under Medical, Visual and Textual, therefore, we established a fourth group 'Other'.

Credibility

This criterion was considered important because participants often wanted to make sure that images were from reliable and valid sources. Sometimes participants used image search engines to locate medical images from online repositories; however general-purpose search engines (e.g. Google) do not distinguish between medically credible and non-credible websites. Some participants believed that *"the source of the image is important and I choose images from highly specialised sources because you cannot get them everywhere."(P4).* Therefore, sometimes participants such as Participant 3 preferred using medical databases such as PubMed[1] to locate relevant articles and then look for images in the articles.

Target Audiences

Some participants cared about their target audience when selecting images. For example, if they wanted to show the images to their students, they would want to make sure that the images used were suitable for that particular group of individuals:

> *"It depends on your target audiences … deficiency, for example … if I want to teach first year medical students you want something absolutely typical … If you are teaching for junior doctors because you will need something a bit more settled. You may try to demonstrate some complexity of it so will end up having multiple images. So it depends on your audience."(P20)*

[1] http://www.ncbi.nlm.nih.gov/pubmed/ (site accessed: 18/06/2008).

Copyright

There were many legal issues concerning the use of copyrighted medical images. For example patients are the owner of their images. Copyright was an issue raised by some participants, especially if they wanted to use images in their publications and presentations.

> *"Due to copyright rules and regulations when I put my lectures up on the web … all lectures have to be published electronically …none of the graphics go with them … I can't breach copyright rules." (P20)*

Availability

Sometimes participants who used search engines, electronic journals and medical databases to find images, could not access full-sized versions of images retrieved from a search. Sometimes web pages containing images they requested had been removed or had been replaced by different images. Sometimes participants said they would pay to access images they needed:

> *"Unfortunately the cost of journals is now very high for individual subscribers and the NHS no longer has an organizational subscription … while there are a lot of online journals, the ones I am more interested in are readily accessible. You have to go and pay the subscription, so it is basically we'll do through the Internet generically. We do a lot of lectures and presentations and representative images are always very useful." (P15)*

Recency

Sometimes participants stated that they would select images based on recency, e.g. the latest medical images for a particular topic:

> *"I want to find the latest images since everything changes in medicine."(P5)*

> *"You can say if you want to do research on a topic you prefer to find the last and the best images from the most reliable resources and well-known authors then you can compare your image which represents your findings with that image." (P8)*

4.5 Importance of criteria

Participants considered different criteria as the most important. As Table 3 shows, visual relevancy of images was the most important criterion for 12 of the participants, with the modality of medical images the second most important criterion for 8 of the participants. One participant said that she would consider the anatomic region illustrated in images, and 5 participants did not specify a particular criterion.

Most important criterion	Number of participants who used
Visual relevance	12 (P1, 2, 4, 6, 8, 9, 11, 16, 20, 21, 26)
Modality	8 (P3, 10, 13, 14, 15, 17, 18, 24)
Not specified; varies	5 (P7, 12, 19, 23, 25)
Anatomic region	1 (P22)

Table 3. Most important relevance criteria.

4.6 Discussion

The objective of this study was to elicit the relevance criteria health care professionals apply when searching for medical images. Our findings show the diversity of the selection criteria

which health care professionals apply when judging the relevancy of images to their medical image information needs. The common criterion health care professionals used was a suite of criteria grouped under the term, visual relevancy. Our interviews show that the participants generally use visual information as the first criterion to judge the relevancy of images, before applying other criteria such as diagnosis, credibility, informativeness, etc. for their final selection. In the research literature on relevance, topicality is mainly defined as the relationship between the query terms and content of text documents. Our findings show a visual analogue: namely participants judged the relevancy of images based on the visual relevancy of images to their image information needs. However, more classic topical relevancy elements such as examining the title or abstract of the textual annotations of an image were also used.

With this set of criteria established, we next examined the topics of a well know medical image test collection.

5. An experiment with ImageCLEFMed topics

ImageCLEFMed is an international evaluation campaign for medical image retrieval. It is the medical track of ImageCLEF which was established as a part of the Cross Language Evaluation Forum (CLEF[1]) in 2003. CLEF itself is an offspring of the Text Retrieval Conference (TREC). The topics of the track were created after conducting surveys and examining the search logs of a number of medical search systems (see research described in Section 2.2). The type of relevance judgments applied in ImageCLEF is generally referred to as topical relevance (Müller et al., 2006). Our findings showed users apply an apparently wider range of criteria to evaluate the relevancy of medical images to their situational medical image needs. To understand this difference more clearly, we analysed 85 topics used in ImageCLEFMed 2005, 2006 and 2007 to investigate how they cover the relevance criteria we identified. Example topics used in ImageCLEFMed included "Show me CT or x-ray images showing the heart", "Gastrointestinal endoscopy with polyp", "Fetal MRI", "Show me images of osteoarthritis in the hand", "Show me images of right middle lobe pneumonia", "Show me images showing peptic ulcers or part of it", and "Show me images of findings with Alzheimer's disease".

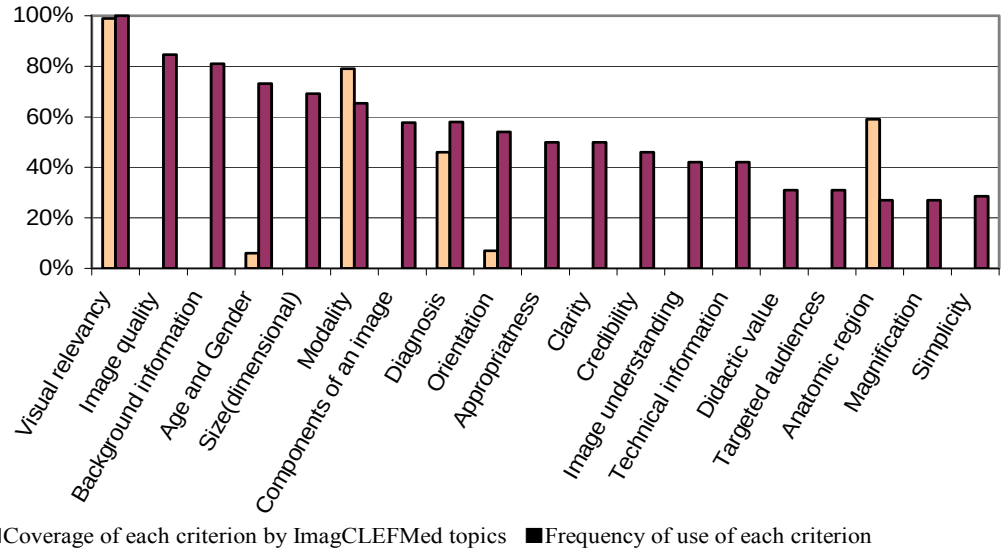

Figure 2. Coverage of relevance criteria by the topics of ImageCLEFMed.

1 http://www.clef-campaign.org (site accessed: 16/08/2008).

The assessment of the coverage of criteria was based on the text of each topic. For example, if in the text the modality of an image was mentioned we recorded that the criterion 'modality' was covered. There were also topics where particular criteria were implied. For example, a query about cancer of the ovary implied a particular gender and for this topic, gender was recorded. We selected 19 out of 26 criteria that health care professionals applied (the seven criteria we withdrew as they were specified by less than 25% of the participants). The results of our analysis are shown in Figure 2.

ImageCLEFMed topics were found to cover explicitly six of our identified criteria. They were visual relevancy (covered by 99% of topics), modality (79%), anatomic region (59%), diagnosis (46%), orientation (7%) and age and gender (6%).

The organizers of ImageCLEFMed reported in 2007 and 2006 they had selected topics with the aim of covering at least two of the following criteria (they referred to as axes): modality, anatomic region, pathology, and visual observation (Müller et al., 2008; Müller et al., 2006). Accordingly, the criteria such as diagnosis, modality and anatomic region were covered significantly by the topics. However, the following criteria were by design not covered by ImageCLEFMed topics: didactic value, credibility, technical information, targeted audiences, size (dimensional), simplicity, magnification, image understanding, image quality, components of images, background information, appropriateness, and clarity of images.

There may be potential for ImageCLEFMed to address some of these criteria in the future. First, creating topics that cover a wider set of criteria (the existing topic, "Fetal MRI", could be adapted to "24 weeks male fetal MRI T2 weighted coronal scan"). Second, one could adapt the existing ImageCLEFMed Image Annotation Task to become a task that tests automated annotation of criteria such as age, gender, modality, and anatomic region from the semi-structured or unstructured text of ImageCLEFMed test collection annotations. In its three years of running, an image's modality, the body region shown in the image and the orientation of the body are key parts of the task (Muller et al., 2007). (Note, these three classes correspond strongly with the three visual medical criteria that our study identified, confirming the importance of them in the annotation task.) It might be possible to extend future annotation tasks to include medical annotation from text.

According to Müller et al., (2007) the majority of images in the ImageCLEFMed collection contain annotations. As we can see in Figure 3, age and gender of the patient has been mentioned in the annotation of the image. In addition, in the abstract, we found some background information and the history of the problem that patient has had. Moreover, the anatomic region presented in the image has been stated as well as the type of image, which is a conventional radiograph. Therefore, it might be possible to create annotation tasks that address at least some of these criteria.

We also noticed in our study that users regarded as important the credibility of images due to the diverse range of sources being searched. However, ImageCLEFMed ensures that its image collection is drawn from teaching files which are high quality credible sources. Therefore, this is not a criterion that could be tested in ImageCLEFMed currently. However, it is clear that searchers often retrieve in collections where this is an important issue: for example, the searchers in our study often searched on Google Image, where credibility is critically important. In such situations, building a medical image search engine that ensures only highly credible images are retrieved could be an important research challenge. If ImageCLEFMed was looking to expand its successful campaign to new fields of research, a study of the credibility of an image source might add a challenging new line of exploration. This could be achieved by building a collection of medical images sampled from a wide range of web sites where the task for system developers is to locate both credible and relevant images.

Title: PNEUMONIA RLL ---- DOES THE CT HELP GIVE US SOMETHING OTHER CAUSE THAN ROUTINE BACTERIAL PNEUMONIA

Abstract: 15-year-old white female who presented to the ER October 30, 2004 with chest pain and shortness of breath. She was feeling well until Thursday, October 28, 2004, when she developed a sore throat at a football game. On October 29, 2004, she had decreased appetite and began developing some shortness of breath as well as low back pain and cough. The cough was noted to be productive with green yellowish sputum. It was thought at that time that she had some viral illness. She has a known past history of asthma. She also noted increasing fatigue.

There was some questionable history of calf pain in both of her calves while running last week; however, she denies any recent travel or trauma

She does have an old meniscal tear on her right knee.Also of note, she was started on oral contraceptive pill approximately five days ago which was started for dysmenorrhea and menorrhagia symptoms. Due to her continued chest pain and shortness of breath, she was seen in the ER. At that time, chest x-ray was normal and chest CT showed a right pulmonary artery density thought to be a clot in a right lower lobe, questionable infarct. On admit, White count of 6.4, hemoglobin 12.4, hematocrit 36, platelets 192. PT of 15.8, PTT is 30. Sodium 141, potassium 2.3, chloride 101, CO2 26, BUN 10, creatinine 0.7, glucose 117, calcium 9.3, unconjugated 0.3, conjugated 0.0. Alkaline phosphatase 71, albumin 4.4, protein 7.3, AST 22, ALT 18. UA clean catch was normal. During hospitalization, she has been on Rocephin and Azithromycin but still has been spiking temperatures. Her CRP was 8.9 on 10/31 and was 13.9 on 11/2.

Findings: Even though there are less air bronchograms than usual, the appearance is more suggestive of bronchogenic than hematogenous origin.

Discussion: No specific signs are seen pointing to the organism. Unfortunately, many atypical organisms that confound the clinicians have nonspecific appearances. The patient has no adenopathy, effusion, cavitation that delay the response. Unfortunately, gray images don't make up for gram stains.

Pathology: Infection

Anatomy: Chest

Modality: CT, Conventional Radiograph

Figure 3. A sample image from an ImageCLEFMed image collection. The original size of image was 10.16x13.55 cm and we extracted annotation from the xml file for this image. Image and annotations adopted from the ImageCLEFMed 2007 test collection (used with permission).

6. Conclusion

In this study, twenty six health care professionals were interviewed about the relevance criteria they used for finding images required in their work setting. This is an area, until now, not investigated within past work on user-orientated relevance. Using a Grounded Theory approach to studying relevance, 26 criteria were identified concerning how participants made relevance judgments. Amongst them visual relevancy, background information and image quality were the most frequent criteria used by participants.

Health care professionals interviewed in this study searched images for different reasons with regards to their medical image information needs. They considered their image information needs as the most influential factor affecting their image relevance judgments. Participants mentioned that the importance of each criterion was dependent upon their information needs.

Participants also chose different relevance criteria as the most important criterion in different situations. Overall in our study, twelve participants selected visual relevancy of images and 8 of them selected modality as the most important criterion. Meanwhile, visual relevancy was a universally applied and used criterion.

Our investigation on the potential coverage of relevance criteria covered by the information needs (topics) specified within the ImageCLEFMed benchmark revealed that it might be possible to include some of our identified criteria in this kind of evaluation setting, thereby testing more realistic criteria which affect searching in a real setting.

Results of our study revealed that different groups of users use different relevance criteria regarding their speciality and their information needs. Through further studies investigating the

relevance judgment process by different users in different situations, it will be possible to have an overview of the relevance judgment process.

7. Acknowledgements

This project is funded by the Iranian Ministry of Health and Medical Education and Iran University of Medical Sciences. We also would like to thank all who facilitated our study and all of the participants who provided us with the opportunity to interview them.

8. References

ARMITAGE, L. H., ENSER, P. G. B., 1997. Analysis of user need in image archives. *Journal of Information Science,* **23** (4), 287-299.

BARRY, C. L., 1994. User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science,* **45** (3)**,** 149-159.

BARRY, C. L., SCHAMBER, L., 1998. Users' criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management,* **34** (2-3)**,** 219-236.

CHOI, Y., RASMUSSEN, E. M., 2002. User's relevance criteria in image retrieval in American history. *Information Processing and Management: an International Journal,* **38** (5)**,** 695-726.

CORBIN, J. M., STRAUSS, A., 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology,* **13** (1)**,** 3-21.

CUNNINGHAM, S. J., MASOODIAN, M., 2006. Looking for a picture: an analysis of everyday image information searching. Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries**,** 198-199.

EAKINS, J.P., Graham, M.E., 1999. Content-based Image Retrieval: A report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle.

ELY, J. W., OSHEROFF, J. A., EBELL, M. H., BERGUS, G. R., LEVY, B. T., CHAMBLISS, M. L., EVANS, E. R., 1999. Analysis of questions asked by family doctors regarding patient care. *British Medical Journal,* **319**(7206), 358–361

GLASER, B. G., STRAUSS, A. L., 1967. Discovery of Grounded Theory: strategies for qualitative research. Chicago, Aldine.

GLASER, B.G., Holton, J., 2004. Remodelling grounded theory. *Qualitative Sociology research,* 5 (2), Art. 4.

GLATARD, T., MONTAGNAT, J., MAGNIN, I. E., 2004. Texture based medical image indexing and retrieval: application to cardiac imaging. Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval. New York, 135-142.

GREISDORF, H., 2002. Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation,* **58** (1), 6-29.

HERSH, W., JENSEN, J., MÜLLER, H., GORMAN, P., RUCH, P., 2005. A qualitative task analysis of biomedical image use and retrieval. In: ImageCLEF/MUSCLE workshop on image retrieval evaluation, Vienna, Austria, 11-16.

HUNG, T. Y., ZOELLER, C., LYON, S., 2005. Relevance judgments for image retrieval in the field of journalism: A pilot study. *Lecture Notes in Computer Science,* **3815,** 72-80.

LANCASTER, F.W., 1979. Information Retrieval Systems: Characteristics, Testing, and Evaluation. Wiley, New York.

MAGLAUGHLIN, K. L., SONNENWALD, D. H., 2002. User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and Technology,* **53** (5)**,** 327-342.

MARKKULA, M., SORMUNEN, E., 2000. End-user searching challenges indexing practices in the digital newspaper photo archive. *Journal Information Retrieval,* **1** (4), 259-285

MIZZARO, S., 1997. Relevance: The whole history. *Journal of the American Society for Information Science,* **48** (9), 810-832.

MÜLLER, H., MICHOUX, N., BANDON, D., GEISSBUHLER, A., 2004. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International journal of medical informatics,* **73** (1)**,** 1-23.

MÜLLER, H., DESELAERS, T., DESERNO, T., CLOUGH, P., KIM, E., HERSH, W. (2007) Overview of the ImageCLEFMed 2006 medical retrieval and medical annotation tasks. *Lecture Notes in Computer Science,* **4730**, 595-608.

MÜLLER, H., KALPATHY-CRAMER, J., HERSH, W., GEISSBUHLER, A., 2008. Using medline queries to generate image retrieval tasks for benchmarking. *Studies in health technology and informatics,* **136,** 523-528.

PALING, S., MISZKIEWICZ, M., ABBAS, J., 2006. Digital images in United States and Canadian dental education: the role of delivery medium and metadata. In Grove, A., Ed. Proceedings 69th Annual Meeting of the American Society for Information Science and Technology (ASIST), Austin, USA, 1-21.

PARK, T. K., 1994. Toward a theory of user-based relevance: A call for a new paradigm of inquiry. *Journal of the American Society for Information Science,* **45** (3)**,** 135-141.

SARACEVIC, T., 1996. Relevance reconsidered. Proceedings of the 2nd Conference on Conceptions of Library and Information Science, Copenhagen, Royal School of Librarianship, 201–218.

SCHAMBER, L., 1994. Relevance and information behaviour. *Annual Review of Information Science and Technology,* **29,** 3-48.

SCHAMBER, L., Eisenberg, M., Nilan, M.S., 1990. A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing and Management.* 26 (6), 755–776.

SHELSTAD, K. R., CLEVENGER, F. W., 1996. Information retrieval patterns and needs among practicing general surgeons: a statewide experience. *Bulletin Medical Library Association,* **84** (4), 490-497.

STRAUSS, A. L., CORBIN, J. M., 1990. Basics of qualitative research: grounded theory procedures and techniques. Newbury Park, Sage Publications.

TANG, R., SOLOMON, P., 1998. Toward an understanding of the dynamics of relevance judgment: an analysis of one person's search behaviour. *Information Processing and Management,* **34,** (2-3)**,** 237-256.

TSAI, C. F., 2007. A review of image retrieval methods for digital cultural heritage resources. *Online Information Review,* **31** (2), 185-198.