RMIT
UNIVERSITY

How do healthcare professionals select the medical images they need?

Sedghi, Shahram; Sanderson, Mark; Clough, Paul

See this record in the RMIT Research Repository at:

http://researchbank.rmit.edu.au/view/rmit:20307

Version: Accepted Manuscript

Link to Published Version:

http://dx.doi.org/10.1108/00012531311297186

How do healthcare professionals select the medical images they need?

Shahram Sedghi
Department of Librarianship and Medical Information Sciences, Tehran University of Medical Sciences, Tehran, Iran
ssedghi@tums.ac.ir

*Mark Sanderson (*Corresponding author)
School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria, Australia
mark.sanderson@rmit.edu.au

Paul Clough
Information School, University of Sheffield, Sheffield, UK
p.d.clough@sheffield.ac.uk

Abstract

Purpose: This article investigates how healthcare professionals search for and select the medical images they need within medical settings.

Design/methodology/approach: In total 29 healthcare professionals participated in the study. Using a think-a-loud technique and face-to-face interviews, we asked participants to explain how they looked for medical images and how they judged the relevancy of retrieved images.

Findings: Fifteen criteria were applied by participants when determining the relevance of medical images in relation to their information needs. Topicality was found to be the primary and most important criterion used by participants.

Originality/value: This study compares the relevance criteria used for medical images with those identified in the literature and highlights six additional criteria which have not been suggested in previous work.

Keywords

Relevance criteria, Relevance judgment, Image retrieval, Medical imaging, Image seeking behaviour

1. Introduction

Different forms of images, x-rays and ultrasound/MRI/CRT scans, are widely used by healthcare professionals in health and biomedical departments (Enser et al., (2006). Healthcare professionals, especially those working in academic settings, use their specialised knowledge and expertise to identify relevant written materials for their information needs (Crystal and Greenberg, (2006). However, few attempts have studied the relevance criteria applied by healthcare professionals when searching for medical images. Given the considerable number of online medical image collections now available, it seems appropriate to begin to explore how these collections are being (or could be) used by healthcare professionals in medical departments.

There have been a number of studies that discuss how healthcare professionals currently search medical text-based resources and examine the medical retrieval systems currently used (Crystal and Greenberg, 2006, Covell et al., 1985, Gorman, 1995, Shelstad and Clevenger, 1996, Ely et al., 1999). However, despite of the importance of medical images, little is currently known about how healthcare professionals identify and select medical images according to their information needs in a work situation. In this paper we address this issue by analysing the relevance criteria used when searching for medical images by analysing the results of interviews held with 29 healthcare professionals who search for medical images on a regular basis as part of their job. The goals of our work were to: (1) determine the relevance criteria used by healthcare professionals to judge the relevance of medical images, (2) to determine whether the relevance criteria identified in this situation differ from the criteria suggested in the literature, and (3) to rank the importance of relevance criteria used for judging the relevance of medical images. This complements previous work we have reported on relevance criteria and sources of visual materials for medical image retrieval (PRL paper and ASLIB proceedings paper). Our study differs from previous work in a number of ways. Firstly, we did not study a particular collection of images or image retrieval system. Instead, we asked participants in our study to search for relevant medical images in any of their preferred sources. Secondly, we did not ask participants to search images for predefined image search task; rather we asked them to search for the images they would typically need for their daily work tasks. Thirdly, we did not recommend a list of criteria to the participants to judge; rather we asked them to explain the criteria and attributes that were important for them while assessing the relevancy of images.

The paper is structured as follows. Section 2 summarises past work that has studied relevance criteria used in various contexts including healthcare and when searching for visual materials; Section 3 describes the research methodology used for data collection and analysis; Section 4 presents the results of our study; Section 5 discusses the results in relation to past work; and Section 6 summarises the study and provides directions for further work.

2. Literature Review

Saracevic reports that there is no a widely agreed definition on what relevance means 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 (Saracevic, 1996: p.13).*" Relevance has been a central concept in the information retrieval field since the 1950s(Schamber, 1994); however, in the earlier studies relevance was regarded as an objective phenomenon and used as a measure for the evaluation of the information retrieval process for traditional textual documents, regardless of the person who made the judgment. Since the 1990s' relevance studies have become user-oriented as the subjective aspect of

relevance has been taken into account. The findings of early studies suggested that the degree of relevance is influenced by various factors. For example, (Cuadra and Katter, 1967) classified those factors into five groups: (1) Type of document being judged, including its subject matter and level of difficulty; (2) Query or topic which expressed the information needs; (3) Judge (subject knowledge of judge and his or her familiarity with the subject of document); (4) Judgment conditions such as time available, order of presentation, number of documents in the document set; and (5) Judgment mode.

In the most recent studies, the concept of relevance has been regarded as having moved beyond topicality in relation to information retrieval and the information seeking behaviour of the user. Barry (1994) suggested that the first step for researchers who want to study relevance criteria applied by different user groups within a range of situations would be to combine the results of previous studies, which have gathered relevance criteria and variables which affect the relevance judgment process, directly from users.

A limited number of studies have been conducted to explore relevance judgements for visual materials. Markkula and Sormunen (1998) studied the criteria used by journalists when searching for photographs in one collection[1] during their routine tasks. The authors grouped relevance criteria applied by the participants into four categories: topicality, technical, contextual attributes and visual attributes. Topicality referred to the situation in which journalists considered annotations to decide on topical relevancy of photos and obtain further information on the image. Technical and contextual properties were the most widely used criteria as most participants were interested in "images which were technically good, not recently published and current" (Markkula and Sormunen, 1998: p. 277). Journalists also regarded criteria such as cost, recency, and visual attributes, e.g. portrait images of a person. Additionally, participants would often consider the message they attempted to transfer by a photo as criteria criterion, e.g. dramatic, shocking, effective, funny or impressive.

Choi and Rasmussen (2002) carried out a research project entitled "User's relevance criteria in image retrieval in American history" to examine criteria applied by image users before and after search. Thirty-eight graduate students of American History and faculty staff from relevant departments were interviewed in the study. The participants used 'American Memory' photo collection and they were questioned to express how they assessed relevancy of photos. Since past research showed a strong overlapping of

---

[1] The collection discussed was the Aamulehti digital image archive.

criteria identified, the researchers suggested a list of nine criteria from identified in the past studies to interviewees and asked them to rank the weight of each criterion. The list included the following criteria: topicality, accuracy, time frame, suggestiveness, novelty, completeness, accessibility, appeal of information, and technical attributes of images. Subjects were then asked to find images and assess retrieved items using suggested criteria. Prior to the search, participants were asked to rate each criterion. Topicality, accuracy and completeness were ranked by participants as the top three criteria. After search and once interviewees had observed the images, they used criteria such as time frame and accessibility of images. Choi and Rasmussen reported that there had been a considerable difference in the rankings of each criterion before and after search.

Hung et al. (2005) examined the relevance criteria used by ten Journalism students. The purpose of this research was to draw out what criteria students applied to choose relevant images. The authors asked students to find images based on three image retrieval tasks: specific, general, and subjective image search tasks. Hung et al (2005) found several criteria employed by students when searching photos for three predefined tasks: typicality[1], emotion, action, aesthetic, text, familiarity, context, impression, preference, posture, facial feature and appearance. Typicality, emotion and aesthetic appearance were rated as the three most important criteria for all tasks; amongst them typicality was the most important criterion applied by participants for all tasks.

From the example studies mentioned above, we conclude that the process of judging the relevance of images in response to a query is multidimensional and the judgment, regardless of how people describe classes and types of relevance, is based on criteria that extends beyond topicality. Moreover, image retrieval poses difficulties compared to text document retrieval: because relevance judgments depend heavily on different levels of situational interpretation by users, there are no agreed features to support retrieval, interpretation and the relevance judgment process (Ingwersen and Järvelin, 2005: p.179). For example, textual documents have common features such as title, author and publication date that are considered by users during the relevance judgment process.

Image relevance studies address a range of relevance criteria that might be applied by image users in various disciplines, such as journalism and history. Nevertheless, as reported by Sedghi et al (2008), researchers are unable to generalise which image relevance criteria will likely be shared across users of information retrieval systems in various situations and settings. Additionally, researchers of previous relevance studies could not specify which criteria could be the most important criteria for the users and in

---

[1] Hung et al. (2005) defined the typicality as a criterion that refers to the universal representation of an object in an image.

which circumstance this significance could happen. We found that our understanding of the judgement process for images is incomplete; as yet there are no distinct explanations for the way that relevance judgment is made for images.

3. Methodology

3.1. Gathering the Data

Strauss and Corbin (1998: p.214) suggest that the site of study or the research population must be selected according to the research objectives. In light of our research objectives, we therefore studied healthcare professionals who were skilled and knowledgeable internet/computer users. In past research on relevance criteria, the number of interviewees varied between eighteen and forty. For example, Barry (1994) had interviewed eighteen participants. Similar qualitative studies of relevance judgment used thirty-eight (Choi and Rasmussen, 2002), thirty (Schamber, 1991), twenty-six (Yang, 2005), and twelve (Maglaughlin and Sonnenwald, 2002). Based on the results of her study, Barry (1994) believed that the full range of criteria of relevance can be collected through interviewing less than ten interviewees. Barry adds that in any possible ordering of participants, redundancy was achieved after the ninth participant had been interviewed. In other words, without regard for the order of interviewing the participants, no new criterion was identified once the ninth had been interviewed.

Taking into account the number of participants in previous studies, and in view of participants' time limitation and availability, it seemed logical to interview a minimum of twenty and maximum of forty healthcare professionals in this study. However, the current project is a qualitative study using grounded theory methodology and in grounded theory the size of research sample is based on theoretical sampling; thus we started the recruitment process and data collection as Strauss and Corbin (1998) suggest. According to Strauss, the researcher continues the cycle of alternation between data collection and data analysis until saturation is achieved. Using theoretical sampling and after interviewing 14 participants, we reached data saturation; however we continued the recruitment and data collection processes to ensure this was the case. The interviews with healthcare professionals who consented to take part in the study were carried out in mutually agreed venues, at mutually agreed dates and times. Most of the interviews were carried out in the offices of participants, and with participants' permission we recorded the interviews using a digital voice recorder. For two participants who did not have a suitable room, a room was booked in a different place. The interviews were carried out face-to-face by the first author and ranged between 28 and 92 minutes.

Ethics approval and permission to conduct the research was issued by the NHS[1] National Research Ethics Service and Sheffield Teaching Hospitals NHS Foundation Trust (referred hereafter as STH) in order to approach, recruit and interview participants. All participants knew the legal issues involved and they were aware of patients' privacy protection. The study was monitored by independent officers from The University of Sheffield and STH monitored to ensure that neither the researchers, nor the interviewees, breached the rules. The anonymity and confidentiality of all data was maintained and participants were informed about these issues.

A hallmark of the ground theory is the simultaneous data collection and data analysis as Strauss and Corbin (1998) suggest. Thus, we transcribed each interview immediately after the interview session. As we stated earlier, in grounded theory data collection and data analysis are based on theoretical sampling, meaning we must commence data analysis as soon as we collect the first bits of data. Theoretical sampling could not occur unless the researcher transcribes each interview immediately after the interview session.

As mentioned earlier, it was important to see the relevance criteria and the judgment situation explained in the healthcare professionals' own terms. Therefore, the types of data we collected from healthcare professionals were verbal. We selected two common ways of gathering verbal data: the semi-structured interview and the think-aloud approach as Schamber (1991); Park (1994) and Choi and Rasmussen (2002) suggested. We used these methods to extract and explain the relevance criteria used by the participants and their perceptions of the concept of relevance. In the first, and the main part of this study, we used semi-structured interviews.

We also asked the participants to conduct real medical image searches and explain exactly how they went about looking for images, and how they evaluated the retrieved images. This method is known as think-aloud protocol and using this method allowed us to verbalise participants' thoughts while searching and selecting images. Most of the time, participants conducted medical image searches during the interview to explain their image search behaviour, relevance judgment processes and other activities. In other words, interviews and think-aloud sessions were not separated. As participants responded to our questions and conducted searches for medical images, we were able to get additional insights and information from them regarding research aims and objectives. Therefore, we considered the data transcribed from the medical image search sessions as part the interview sessions, and did not analyse the data obtained during the two sessions separately. Before data collection, we conducted a preliminary study with 10 postgraduate students. This was crucial for several reasons: Firstly, we conducted a preliminary study to ensure the feasibility of

---

[1] National Health Service in the United Kingdom

the main study, to test the data collection instrument, and to assess the data analysis approach selected for our research. It was also an opportunity to develop interviewing skills and to prepare for the main study. Moreover, we believed that the findings of such a study would help us to clarify the research topic, and its aims and objectives.

3.2. Data analysis

A qualitative approach to analysis the data was adopted as suggested by Maglaughlin and Sonnenwald (2002); Ingwersen and Järvelin (2005); Hirsh (1999); Park (1994) and Myers(1997). We used Grounded Theory (GT) and its paradigms for data collection and data analysis, specifically we used the 'Straussian version' of GT detailed in Strauss and Corbin (1998). Using an approach based on GT, we characterised the criteria healthcare professionals used when making relevance judgments. There are some software packages available for qualitative data analysis. Using such software enables the researcher to search the assigned codes for patterns; and to establish categories of codes that resemble testable models of the conceptual structure of the underlying data as reported by Lewis (2004). Based on his experiments during the preliminary study, the main researcher chose the NVivo software package. In order to validate the data analysis process, the main researcher asked a grounded theory researcher to check the assigned codes to four interviews. This method is known as double-coding (Gilbert, 2001). Since both researchers were in agreement, the trustfulness and replicability of the data were further confirmed. Additionally, the data collection and analysis based on theoretical sampling allowed us to document any changes to the questions of interview protocol.

4. Results

4.1. Relevance criteria

This study primarily aimed to investigate the relevance criteria typically used by healthcare professionals when searching for medical images in their daily healthcare situations. Table 1 shows the fifteen relevance criteria identified by healthcare professionals[1]. Participants applied various criteria in their image selection processes; among them topicality, image quality and size of the image were the three most common relevance criteria. The order of relevance criteria is based on the number of participants who applied the criteria during their visual search tasks.

Table 1: list of criteria used by healthcare professionals
The number shows the number of interviewees who used each criterion (*N=15*).

Criteria	Frequency

---

[1] Please see appendix 1 for definition of each criterion.

Topicality	29
Image quality	27
Size (dimensions)	21
Age and gender	20
Modality	18
Orientation	15
Credibility	13
Targeted audiences	12
Technical information	11
Magnification	7
Colour	7
Copyright	6
Availability	4
Recency	3
Originality	2

4.1.1 Topicality

The most frequent and most important criterion used by all participants was topical relevancy. Although this finding is consistent with previous research (Schamber et al., 1990, Mizzaro, 1997) , we noted that participants used both textual description and the visual appearance of the image when making relevance judgments. Topicality reflects the 'aboutness' or subject matter of an image. All twenty-nine participants mentioned their requirements for textual information, such as medical background and diagnoses in the text related to a photo, to assess the topical relevancy of images. Our findings highlighted the need for textual information by healthcare professionals, especially a comprehensive description since they would usually want to use images for research, educational or clinical purposes. The healthcare professionals consistently read the text surrounding the images that were visually relevant since they wanted to know what was actually illustrated in the image, and what its original context had been. Although text was an important source of information for assessing the topical relevancy of images, we noted that the participants could not access relevant textual information when images were obtained from web-based resources or when they used departmental image collections. We also note that participants evaluated the topical relevancy of an image regarding its visual appearance. Our findings demonstrated that participants had a visual representation of the objects they sought in their mind, and that participants used this visual representation when assessing retrieved images.

Our findings show that to make relevance judgments, the user must visually inspect the image in order to know if the retrieved image is topically relevant, and whether the image contains the requested visual

information. We noted that, based on their knowledge and experience, participants had a 'mental image' of what they were looking for (i.e. a part of human anatomy, certain type of image or a medical instrument), and they were trying to find something similar. Further investigation of the data showed that the visual appearance of images, and the participants' visual memory, were the most important sources of information in assessing the topical relevancy of images. The interviews in this study also showed that after topical relevancy of an image was established, participants applied other criteria. For example, twenty-five participants said that if the image was topically relevant and illustrated what they wanted, they would then check the quality of the image to see whether they could use it or not. Similarly, thirteen participants stated that they would also consider the credibility of images if the images were topically relevant.

4.1.2   Image quality

The photographic quality of an image, including resolution, contrast, and brightness, was the second criterion applied by twenty-seven participants. The participants seemed to check the quality of images after topical relevance was established. Sometimes the participants who looked for images in web-based resources using Google image search complained about the low quality of images. Therefore, they preferred to find images in resources other than web-based resources. Some participants used images to illustrate their presentations, therefore they emphasized that they were interested in copyright-free and high quality images to show the desired visual details in their presentations.

4.1.3   Size (dimensions)

The selection criteria for medical images were closely connected to the participants' individual tasks. Often participants sought images to illustrate their presentations etc; therefore, they looked for images in an appropriate size. Given the participants' preferences for finding images in appropriate dimensions, they stated that they would amend the size of photos using image editing applications if necessary. Some of the participants mentioned that they might find a small size image, which seemed to be the most relevant. In those cases, they tended to reproduce larger versions of those images.

4.1.4   Age and gender

The age and gender of the case illustrated in the images (i.e. the patient) was mentioned by twenty participants. Information concerning the age and gender mainly was collected from the textual information attached to images. Age is an essential criterion for some diseases etc in medicine and, in some cases; this criterion is included implicitly in the query itself. For instance, in the query 'epiphyseal closure', age is included in the query.

One person commented further on this issue, and reported that age is important for some medical conditions that change with age:

> P3: If I am looking for images of the foramen of the skull when it is open or closed, age is important. The foramen of the skull is open in newborn babies but changes with age.

The gender was also used by some participants, when choosing a relevant image.

> P21: This [Ultrasound images of Endometrium] is something that relates to females. In some cases, all images are female related images, because I am looking for ultrasound images of ovaries.

As participants of the study mentioned, age and gender were critical factors for making clinical decisions. Medical databases such as PubMed allow healthcare professionals to limit their search by gender or age. Such search facilities are required to limit the image search to a certain age and gender. However, the diversity of medical images is the main problem for automatic detection of age and gender.

4.1.5 Modality

Different methods and techniques are used in medical imaging to create X-ray radiographs, Magnetic resonance imaging (MRI), microscopic images etc. We noted that participants wanted to retrieve certain type of medical images. Thus, they regarded the type of medical image (referred to as "modality" in this study) as a criterion. In total, eighteen participants applied this criterion and amongst them eight participants ranked the modality as the first criterion. Participants were interested in including the modality of images in the retrieval process, and wanted to retrieve certain types of images as they specified. However, they had some difficulties when they wanted to narrow down the image search to a particular type of image. For instance, when participants used Google image search to search for images, they used the modality of images as a search term. They added terms such as 'PET', 'MRI' and 'CT' to the query.

4.1.6 Orientation

Fifteen of the participants were concerned about the orientation of medical images. Participants described orientation as the view or overall representation of objects in an image. It depends on the location and the direction of imaging devices when producing an image. Participants emphasised that they would consider the orientation when comparing topically relevant images. In other words, participants applied this criterion to discard irrelevant images and select images with the desired orientation. Healthcare professionals used standardized anatomical terms of location such as Sagittal, Coronal, and Transverse to describe the

orientation of medical images they wanted. Accordingly, they distinguished easily between images considering this criterion.

4.1.7 Credibility

This was an important criterion since the interviewees preferred to obtain images from credible sources. The findings of this study show that participants always debated the credibility and trustworthiness of images retrieved from web-based resources using general image search tools such as Google image search tool. Therefore, sometimes participants were interested in using medical databases, such as PubMed[1], to retrieve papers and then find images used in the articles.

4.1.8 Target audiences

Twelve participants were concerned about their target audience when selecting images. Most of the time participants emphasized on the importance of images for educational purposes by saying "an image is worth a thousand words". They believed that some images convey the message more effectively to the target audience. Our findings showed that participants considered other criteria such as quality or size of image before they decided whether an image is suitable for the target audiences or not.

4.1.9 Technical information

This is a criterion used by eleven participants. This criterion typically refers to information associated with the retrieved images. Participants applied this criterion when they wanted comprehensive information on substances, techniques, the laboratory condition, which an image was produced etc. Participants stated that they needed information to compare their research (and their findings) with other researches and findings. Therefore, in order to access technical information relevant to the images for better comparison of their findings, participants preferred to find images in the published literature; in particular in academic papers.

4.1.10 Magnification

when participants looked for microscopic images of human tissues, cells or molecular structure of objects such as proteins, magnification played an important role during their image selection process. When the participants wanted to show the visual details of things such as cells etc they would pay attention to the magnification of the images they retrieved.

---

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

4.1.11 Colour

The colour of images was also regarded as a criterion by the participants when selecting images. Colour could help users to differentiate different parts of an image. The interviews in this study also showed that participants tended to select colour images. They believed that colour images could be understood easier than the single colour or black and white images. Although there are some colour imaging techniques in the medical domain (i.e. pathology images), not many medical images are colour.

4.1.12 Copyright

There are several legal concerns regarding the use of copyrighted visual materials including medical images. For instance, in the medical context; patients own their images as Tranberg et al (2003) stated. Copyright was raised by some participants, particularly if images were used to illustrate their publications etc. In particular, participants were concerned about copyright issues related to using images from web-based resources. Thus, participants preferred to find images in papers and books rather than searching for images in web-based resources.

4.1.13 Availability

It was important for healthcare professionals to access images they found on the internet. Sometimes participants who located images using search engines, could not access full-sized images. Google image search presents thumbnails of images stored in its cache, but participants wanted to see full size images in order to assess and select them. In some cases, web pages containing images, they retrieved, were eliminated or were replaced by other images. Therefore, healthcare professionals could not access those images. Although participants preferred to access images that were free of charge, they declared they would purchase images they wanted.

4.1.14 Recency

Findings of the current study show that participants would choose an image regarding its recency, requesting the newest images for a topic. Recency is another criterion that exactly matches the findings of the previous studies. For example, recency in Wang and White (1999); Maglaughlin and Sonnenwald (2002) and currency in Schamber (1991) refers to the extent to which participants assessed documents to be current or recent.

4.1.15 Originality

This is a criterion mostly used by participants from Medical Physics Department. They usually analysed original medical images by methods they had developed, therefore; they wanted original copies of images, i.e. without any kind of manipulation. The findings of the current study, and the research reviewed, support the assumption that topicality does not automatically result in relevant images and that users evaluate images with qualities that go beyond topicality. However, this does not mean that image retrieval based on topicality is necessarily false. As we discussed earlier, topicality plays a major role in the relevance judgment of retrieved documents including images. However, we believe that information retrieval systems can be designed to incorporate criteria other than the topicality. The results of this study, and the studies we discussed, indicate that users can determine whether those criteria are applicable for a retrieved document, or they can decide whether or not the document will provide the information they need. By identifying the criteria applied by the users, we may be able to incorporate the relevance criteria into the image indexing and retrieval process and consequently take image retrieval beyond the topical approach.

4.2.   Appearance of the criteria

In order to see whether the criteria we identified were different from those criteria suggested in the literature, we contrasted [1] our criteria against those found in the literature (see table 2). In order to access a comprehensive list of criteria, we used the lists of criteria collected by Schamber (1994) [2]; Mizzaro (1997) [3] and Maglaughlin and Sonnenwald (2002)[4]. We also used the relevance criteria for images identified by Hung et al. (2005); Markkula and Sormunen (2000) and Choi and Rasmussen (2002), and video relevance criteria documented by Yang (2005) [5].

Table 2: The appearance of criteria in the literature

Relevance criteria for medical images;	The appearance of criteria in

---

1 Unfortunately not all authors defined the relevance criteria. Therefore the contrast between the criteria we identified and those suggested in the literature was based on the meaning overlap of the criteria.

2 Schamber, (1994) reviewed the literature on relevance published between 1983 and 1994.

3 Mizzaro, (1997) reviewed the relevance studies published between 1959 and 1996.

4 Maglaughlin and Sonnenwald, (2002) compared the relevance criteria identified in their studies with the criteria suggested in ten major relevance studies conducted between 1991 and 1999.

5 As stated earlier, Choi and Rasmussen, (2002) did not study the relevance criteria applied by image users. They asked images users to use a list of ten widely used criteria in the literature (mostly those criteria were applied for relevance judgment of textual information) and assess the relevancy of photos to their information needs.

ordered according to frequency	the literature
Originality (2)	
Recency (3)	✓
Availability (4)	✓
Copyright (6)	✓
Magnification (7)	
Colour (7)	✓
Technical information (11)	
Targeted Audiences [1] (12)	✓
Credibility (13)	✓
Orientation (15)	
Modality (18)	
Age and Gender (20)	
Size (dimensional) (21)	✓
Image Quality (27)	✓
Topical Relevancy (29)	✓
*The number in parenthesis shows the frequency of each criterion	Total number of criteria appearing in the literature: 9

As table 2 shows nine out of fifteen criteria we identified are suggested in past literature. Table 2 also indicates that six out of fifteen criteria were applied by more than half of the participants (total=29 participants).

Overall, we identified six criteria which were not suggested in the literature, of which four criteria were applied by more than half of the participants. The most frequent and most important criterion we documented was topical relevancy. In the past publications on relevance, topicality is mostly described as the relationship between the users' query and content of retrieved items. Our findings show that participants judged the topical relevancy of images using the description and visual appearance of images related to the problem at hand. In response to the first research question of the study, and based on the findings of our experiment, we believe that the set of relevance criteria that healthcare professionals applied for the judgment of medical images, differed from the sets suggested in the literature.

4.3.   Core relevance criteria

Figure 1 gives a summary of all the criteria mentioned by participants during the study. Fifteen relevance criteria were summarized from the interviews and we calculated the percentage for each criterion we identified. The percentage give for each criterion in Figure 1 means the number of participants who

---

1 Sometimes researchers used terms such as 'difficulty level' to name this criterion.

mentioned each criterion. We believed that this figure show the weight of each relevance criterion mentioned.

Among the criteria used by all twenty-nine participants, the largest proportion went to visual (non-textual) criteria which related to visual attributes of medical images. The next was textual criteria, which related to the participants' domain knowledge, descriptions of images and textual attributes of an image including age and gender, copyright, recency and technical information. Finally participants relied on other criteria. Thus, it could be concluded that participants relied mostly on the visual attributes of medical images to make relevance judgments. Topical relevancy (100%) and image quality (93.1%) were the two most frequently mentioned visual criteria. However, age and gender (68.97%), grouped in textual criteria, was also an important criterion.
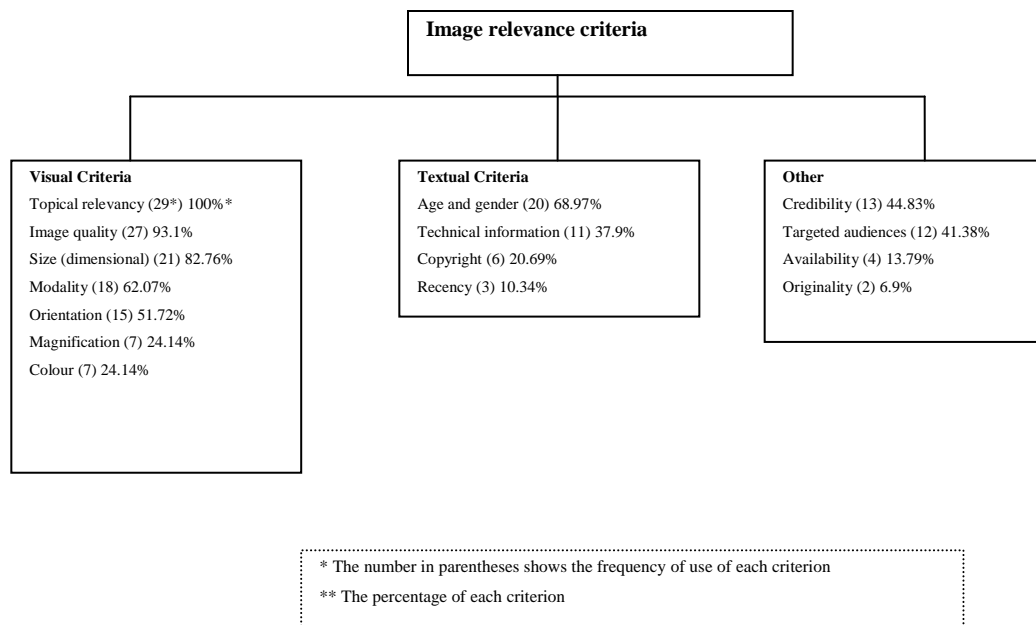
Image relevance criteria

Visual Criteria
Topical relevancy (29*) 100%*
Image quality (27) 93.1%
Size (dimensional) (21) 82.76%
Modality (18) 62.07%
Orientation (15) 51.72%
Magnification (7) 24.14%
Colour (7) 24.14%

Textual Criteria
Age and gender (20) 68.97%
Technical information (11) 37.9%
Copyright (6) 20.69%
Recency (3) 10.34%

Other
Credibility (13) 44.83%
Targeted audiences (12) 41.38%
Availability (4) 13.79%
Originality (2) 6.9%

* The number in parentheses shows the frequency of use of each criterion
** The percentage of each criterion

Figure 1: A summary of the criteria mentioned by all of the participants.

It can be seen from figure 2 that topical relevancy, image quality, size (dimensional), age and gender, and modality are the top five most frequently applied criteria by all the participants. These five criteria were mentioned by more than two-thirds of the participants.

To conclude, the current study shows the diversity of criteria that participants employ when judging the relevancy of medical images. We grouped the common criteria interviewees applied under the term '*visual*

criteria'. In general, participants used topicality as the first criterion to decide on the relevancy of images, before considering criteria such as image quality and dimensional size of the image. Our investigation shows that the participants judged the topicality of medical images using textual descriptions and the visual appearance of images. We believe that topicality is an independent criterion and can be used to judge the relevancy of all types of documents in general and medical images in particular.
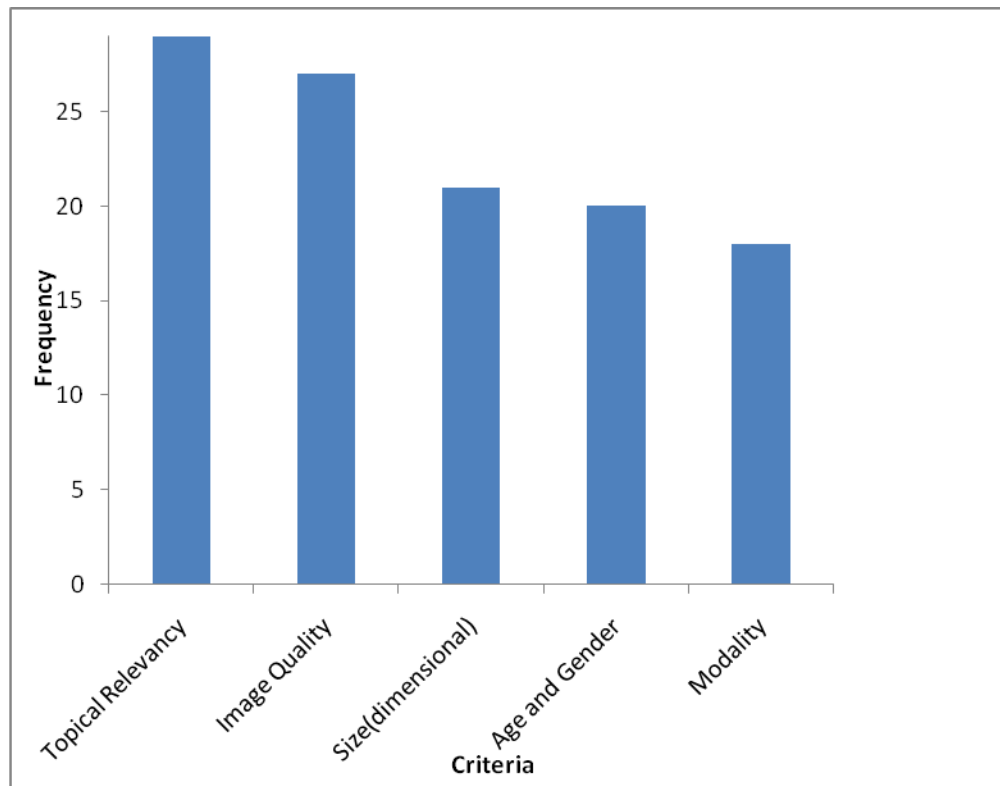

Figure 2: Top 5 most frequently mentioned criteria by participants.

5. Discussion

Using the Straussian version of grounded theory, fifteen relevance criteria were elicited from healthcare professionals who participated in this study. In particular, topical relevancy was a criterion applied by all of the participants, while criteria such as originality were mentioned by only two participants. Findings showed that healthcare professionals considered topical relevancy as the minimum requirement for beginning the judgment process. These findings highlighted the importance of topical relevancy for the judgment of images; however, other criteria such as quality of images were also applied. The critical criterion was

16

applied when the healthcare professionals wanted to select the best and most relevant images from among the candidate images. Healthcare professionals applied relevance criteria in different ways. Although they applied criteria such as topical relevancy or dimensional size of images according to the visual appearance of images or the textual information attached to images, they also applied criteria such as 'recency' that were not evident, or were even absent, in the images. Additionally participants used certain criteria to compare topically relevant images. For example, the criterion "the quality of an image" was used as an important criterion if an image was derived from a printed publication.

In order to see whether the relevance criteria we identified were different from those criteria studied in the literature, we compared the list of criteria used by healthcare professionals to those suggested in the literature. However, we believe that it is not easy to compare the medical image relevance criteria identified in the current study to those found in the past.

Park (1992); Cool et al (1993); Barry (1994); Wang (1994); Schamber and Bateman (1996); Tang and Solomon(1998); Bateman (1998); Spink et al (1998) and Maglaughlin and Sonnenwald (2002) investigated the relevance criteria used for the judgment of textual documents (mostly bibliographic information). Moreover, textual documents contain generally agreed textual features such as title, keywords and abstract. Using these textual features, users can make a decision on the relevancy of textual documents to their information needs; these textual features also facilitate the retrieval process. However, tangible and shared features such as these do not exist for all images. Thus comparison could not be meaningful unless we compare the criteria we found to those found for images; unfortunately, few studies investigated the relevance criteria for images. The second reason is the setting of the current study. We investigated the relevance criteria applied for real image information needs of our participants in their real-life context; however, studies such as Barry (1994); Park (1993); Maglaughlin and Sonnenwald (2002) and Hung et al. (2005) examined the relevance criteria used by users for predefined tasks, or studied the relevance criteria in artificial test situations. The criteria we identified were not obtained in the same manner as those found in the literature, and so cannot be compared. The third reason is the different names used for criteria found in the literature. We noticed that researchers used different names for the relevance criteria and categories of criteria they identified (see (Maglaughlin and Sonnenwald, 2002: p.330). For example, recency and recentness are two different names for the same criterion. The fourth reason is that healthcare professionals looked for medical images in different resources including Google image search, journal articles, personal collections, books and departmental collections (see Sedghi et al (2011) for detailed information on the image resources used by healthcare professionals). However in studies by Choi (2000) ; Barry and Schamber (1998); Park (1994) and Markkula and Sormunen (2000), the users made use of only certain databases to retrieve the documents they required.

17

Other findings of the study indicate that healthcare professionals relied more on visual criteria than on other groups of criteria. This category includes seven out of the fifteen criteria we identified. (Three out of seven criteria in the visual criteria group were common between this group and the textual group of criteria). The findings of the study showed that 'topical relevancy' and 'modality' have been clearly identified as the most influencing factors for the evaluation and selection of medical images. The 'topical relevancy' of medical images was not only the most frequent criterion applied by all participants of this study, but was also selected as the most important criterion by fifteen participants. Figure 1 shows that healthcare professionals mainly considered the visual attributes of images for the relevance judgment of medical images.

6. Conclusion and Further Research

This study is the first of its kind to investigate the relevance criteria applied by healthcare professionals when searching for medical images. Our research highlighted the importance of relevance studies in a particular domain for better understanding of relevance judgment and image seeking behaviour. To the best of our knowledge, there has not been a comprehensive study of relevance criteria applied by healthcare professionals. Thus, the study helps to contribute to our understanding of relevance criteria applied to medical images. The current study also considers the concept of relevance and the information seeking behaviour of healthcare professionals.

The criteria we identified were grouped in three main categories: visual criteria, textual criteria and other criteria. Overall topical relevancy was the most frequently applied criterion (used by all participants), which resembles the findings from image relevance studies by Markkula and Sormunen (2000); Choi and Rasmussen (2002) and Hung et al (2005).

Healthcare professionals main concern was the usefulness of the images found for their situational pictorial information needs. Topicality, by definition, refers to the relevancy of the query used by users to the theme of the retrieved documents. This implies the utilization of tangible and textual features of the document. Since such features do not exist for medical images, healthcare professionals judged the topical relevancy of these based on the visual appearance of images and related textual information. Medical images can be interpreted and used in different ways by different users. Though we may not know exactly which attributes of medical images, or which specific combination of attributes, determine the relevance of an image at a given point in time, we know that healthcare professionals assessed the relevancy of images considering the visual appearance of images and using what they described as 'visual memory'. This confirms the findings of Greisdorf and O'Connor (2002: pp.20-21)'s study that image users judge the relevancy of retrieved images using what they described as 'temporal prototypes'. Although we explained how healthcare

professionals compared and selected images they needed, further investigation is required to determine how image users use their visual memory to search for and select images they need.

The information needs and the medical image resources available seemed to be important factors influencing relevance judgment of images in the current study. This actually corresponds with the findings of previous research highlight the importance of the information needs of users in relevance judgment of retrieved documents in specific domain(Schamber et al., 1990, Ingwersen and Järvelin, 2005, Borlund, 2003, Saracevic, 2007, Hersh, 2003). For example, the healthcare professionals looked for medical images using Google image search for educational purposes, and mainly paid attention to the visual attributes of images when they evaluated the images retrieved. Healthcare professionals also did not consider criteria such as credibility when they used images from journal articles. There were also other factors such as the size of results set (when participants used Google image search), the work experience of healthcare professionals, and the quality of search results which affected the relevance judgment of medical images, and made the medical image relevance judgment a multidimensional and situational process.

Healthcare professionals obtained medical images from web-based resources (mostly using Google image search), journal articles, personal collections, books, image collections of colleagues and departmental collections. Although all participants began image relevance judgment by applying the criterion 'topical relevancy', they applied different criteria such as 'quality of images' or 'credibility' when they wanted to make their final decision concerning the relevancy of images to their information needs. The findings highlighted the fact that the medical image selection process would only begin after the topical relevancy of images was ensured by healthcare professionals.

As stated earlier, this has been the first study of its kind to investigate the relevance criteria for medical images used by healthcare professionals with a grounded theory approach. However, due to the limited number of participants interviewed, we expect that more similar image relevance studies will help to examine the credibility and trustfulness of the criteria we identified. In particular, we suggest that researchers conduct more medical image relevance studies, and investigate the relevance criteria employed by specific groups of healthcare professionals. One potential area for further research is the way that healthcare professionals seearch for images in the image resources they use, their search strategies, the way that they interact with image search tools such as Google image search, and the differences and driving factors of their image seeking behaviour. Further research is needed to focus on the image resources used by healthcare professionals, their interaction with image searching tools, and the factors which drive their image information needs. Another potential area for further research is to evaluate some of our identified criteria in relation to the future design of image retrieval systems and medical image retrieval. Inclusion of

criteria such as dimensional size of images and modality will improve the design of medical image retrieval systems and thus better satisfy healthcare professionals' pictorial information needs. Although we identified topical relevancy as the most important and most frequent relevance criterion for relevance judgment of medical images, we still need further research about the nature of topical relevancy as a relevance criterion for the relevance judgment of images.

7. Acknowledgements

8. Appendix 1

Definition of relevance criteria

Age and Gender	Whether the age and gender of patient was explained or not.
Availability	Whether the image was available or not.
Colour	Whether the image was a colour image, monochrome or black and white image.
Copyright	The copyright of an image. E.g., whether the image could be presented in public or the content of the image could be used for publication purposes.
Credibility	Whether the image was obtained from a reliable source or not. The participants wanted to choose images from resources or people whom they trusted.
Image quality	The photographic quality of image content; whether the image content was clear to see.
Magnification	What was the magnification scale of an image, and if it was a magnified image the magnification was shown in the image or text.
Modality	The type of an image such as MRI, X-Ray, PET, CT scan, Pathology image and Electron microscope images.
Orientation	How the objects were presented in the image and from what angles or direction images were taken.
Originality	Sometimes the participants wanted to find medical images in original format without any manipulation. Original medical images are large files and stored in electronic medical image archiving systems such as DICOM.
Recency	The date of production of an image. Participants wanted to find more recent images to use.
Size	The dimensional size of an image the participant was interested in.
Targeted audiences	Whether the image was targeted at certain audiences the participant wanted to focus on
Technical Information	Whether the methods and materials used to produce an image was explained or not.
Topicality	The aboutness of an image. Whether the image visually illustrates what the participants were interested in. The participants used visual appearance of images and textual information to assess the topicality of images.

9. References

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

BARRY, C. L. & SCHAMBER, L. 1998. Users'criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management: an International Journal,* 34**,** 219-236.

BATEMAN, J. A. 1998. *Modeling changes in end-user relevance criteria: an information seeking study.* University of North Texas.

BORLUND, P. 2003. The concept of relevance in IR. *Journal of the American Society for Information Science and Technology,* 54**,** 913-925.

CHOI, Y. 2000. *The characteristics of users' queries and users' relevance criteria in a image retrieval in American history.* PhD, University of Pittsburgh.

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

COOL, C., BELKIN, N. J. & KANTOR, P. B. Characteristics of texts affecting relevance judgments. Proceedings of the 14th National Online Meeting, 1993 Medford. Learned Information, Inc, 77–84.

COVELL, D. G., UMAN, G. C. & MANNING, P. R. 1985. Information needs in office practice: are they being met? *Ann Intern Med,* 103**,** 596-9.

CRYSTAL, A. & GREENBERG, J. 2006. Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology,* 57**,** 1368 - 1382.

CUADRA, C. A. & KATTER, R. V. 1967. Experimental studies of relevance judgments: Final report. Santa Monica: System Development Corporation.

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**,** 358–361.

ENSER, P. G. B., SANDOM, C. J. & LEWIS, P. H. 2006. Surveying the Reality of Semantic Image Retrieval. *Lecture Notes in Computer Science,* 3736**,** 177.

GILBERT, G. N. 2001. *Researching Social Life*, Sage.

GORMAN, P. N. 1995. Information needs of physicians. *Journal of the American Society for Information Science,* 46**,** 729-736.

GREISDORF, H. & O CONNOR, B. 2002. Modelling what users see when they look at images: a cognitive viewpoint. *JOURNAL OF DOCUMENTATION,* 58**,** 6-29.

HERSH, W. R. 2003. *Information Retrieval: A Health and Biomedical Perspective*, Springer.

HIRSH, S. G. 1999. Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science,* 50**,** 1265-1283.

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.

INGWERSEN, P. & JÄRVELIN, K. 2005. *The Turn: integration of information seeking and retrieval in context,* Dordrescht, Springer.

LEWIS, R. B. 2004. NVivo 2.0 and ATLAS. ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs. *Field Methods,* 16**,** 439-469.

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**,** 327-342.

MARKKULA, M. & SORMUNEN, E. Searching for photos-journalists' practices in pictorial IR. *In:* EAKINS, J. P., HARPER, D. J. & JOSE, J., eds. The Challenge of ImageRetrieval, A Workshop and Symposium on Image Retrieval, Electronic Workshops in Computing, 1998 Newcastle upon Tyne.

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

MYERS, M. D. 1997. Qualitative Research in Information Systems. *MIS Quarterly,* 21**,** 241-242.

PARK, T. K. 1992. *The nature of relevance in information retrieval: An empirical study. Unpublished doctoral dissertation,.* PhD, Indiana University.

PARK, T. K. 1993. The Nature of Relevance in Information Retrieval: An Empirical Study. *Library Quarterly,* 63**,** 318-51.

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**,** 135-141.

SARACEVIC, T. Relevance reconsidered. *In:* INGWERSEN, P. & PORS, N. O., eds. the 2nd Conference on Conceptions of Library and Information Science, 1996 Copenhagen. Royal School of Librarianship, 201–218.

SARACEVIC, T. 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *J. Am. Soc. Inf. Sci. Technol.,* 58**,** 2126-2144.

SCHAMBER, L. 1991. Users'Criteria for Evaluation in a Multimedia Environment. *Proceedings of the ASIS Annual Meeting. p126-33.*

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

SCHAMBER, L. & BATEMAN, J. 1996. User criteria in relevance evaluation: Toward development of a measurement scale. *Proceedings of the American Society for Information Science, Baltimore, MD***,** 218–225.

SCHAMBER, L., EISENBERG, M. & NILAN, M. S. 1990. A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management,* 26**,** 755-776.

SEDGHI, S., SANDERSON, M. & CLOGH, P. 2011. Medical image resources used by health care professionals. *Aslib Proceedings,* 63.

SEDGHI, S., SANDERSON, M. & CLOUGH, P. 2008. A study on the relevance criteria for medical images. *Pattern Recognition Letters*.

SHELSTAD, K. R. & CLEVENGER, F. W. 1996. Information retrieval patterns and needs among practicing general surgeons: a statewide experience. *Bull Med Libr Assoc,* 84**,** 490-7.

SPINK, A., GREISDORF, H. & BATEMAN, J. 1998. From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management,* 34**,** 599-621.

STRAUSS, A. L. & CORBIN, J. M. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Sage Publications Inc.

TANG, R. & SOLOMON, P. 1998. Toward an Understanding of the Dynamics of Relevance Judgment: An Analysis of One Person's Search Behavior. *Information Processing and Management,* 34**,** 237-256.

TRANBERG, H. A., ROUS, B. A. & RASHBASS, J. 2003. Legal and ethical issues in the use of anonymous images in pathology teaching and research. *Histopathology,* 42**,** 104-109.

WANG, P. & WHITE, M. D. 1999. A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *Journal of the American Society for Information Science,* 50**,** 98-114.

YANG, M. 2005. An exploration of users' video relevance criteria. PhD: THE UNIVERSITY OF NORTH CAROLINA.