Relevance Judges' Understanding of Topical Relevance Types: An Explication of an Enriched Concept of Topical Relevance

Xiaoli Huang

College of Information Studies, University of Maryland, College Park, MD 20742. Email: Xiaoli@wam.umd.edu.

Dagobert Soergel

College of Information Studies, University of Maryland, College Park, MD 20742. Email: dsoergel@umd.edu.

Despite the centrality of topical relevance in information retrieval system design and evaluation, understanding and implementation of it is usually limited to "direct overall topical matching" between the subject of the query and the subject of the document. The underlying assumption is that only a single type of topical relationship is involved. In related work, a relevance judgment instrument was developed for the Multilingual Access to Large Spoken ArCHives project (MALACH). It incorporates the five topical relevance types (direct relevance, indirect/ circumstantial relevance, context relevance, comparison relevance, and pointer relevance) and was applied by four judges to items in the MALACH test collection in Summer 2003. This paper reports on the experiences and perceptions of the judges making more nuanced judgments about topical relevance. The results suggest that more than only one variable/dimension, "whether it is on topic" as usually referred to, contributes to topical relevance, and more than a single topical relationship type, "direct matching" as generally assumed, play an important role in topical relevance.

Introduction

The purpose of this paper is to explore the complexities of topical relevance and to identify variables considered in judgments about topical relevance. The ultimate goals are to improve our understanding of the concept of topicality in the context of information retrieval (IR) and to improve IR system performance by incorporating a more complete notion of topicality.

Relevance is a fundamental concept in information science and serves as the most central concern of IR system design and evaluation. Topicality, in turn, lies at the heart of relevance. Despite recognizing topical relevance's importance, understanding of topical relevance is still limited and oversimplified. Topical relevance is generally referred to as direct topical matching or the direct matching between the overall topic of a relevant document and the overall topic of the user need. Although topicality is often mentioned in relevance studies, it is only treated as a self-explicable elementary concept and rarely receives more discussion beyond this general level. Relatively little is known about the ways in which the two topics "match" each other, how this topical matching relates to search situations or other user preference, whether there is only one type of topical relevance, what other factors responsible for topicality besides topic matching, and so on. Therefore, more thinking is urged on the issue of topical relevance.

Literature Review

Relevance

Up to the 80s, much of the discussion on relevance focused on topicality; for general reviews of the concept of relevance see Saracevic (1976) and Mizzaro (1997). In the 90s, a number of qualitative studies continuing the seminal work of Cuadra (Cuadra & Katter, 1967a, 1967b) identified numerous other variables, such as recency, novelty, and quality, to name but a few of over 80 variables (Barry, 1993; Cool, Belkin, Kantor, & Frieder, 1993; Park, 1992; Schamber, 1990, 1991, 1994; Froehlich, 1994; Wang, 1994). These variables vary from user to user and among different points of time.

Topical Relevance

Cooper's "logical relevance" (1971) and Wilson's "situational relevance" (1973) have been the most influential in-depth definitions of topic relevance and have built the foundation for this concept. As defined by Cooper (1971), a document is logically relevant to a user need if it is in a minimal premiss set of component statements answering that user need. In this sense, topical relevance is a logically inferential relationship rather than a direct matching relationship that is translated into most IR systems. Built upon Cooper's logical relevance, Wilson constructs a definition using probabilities of evidential relevance drawn from inductive logic that relates logical relevance to a particular individual's situation. Wilson suggests that situational

relevance captures the "essentials of the vague popular notion of practical relevance...that must bear on our actions". Unfortunately, IR systems rarely address this "situational" aspect of topical relevance.

Topical relevance has been criticized for its failure to prove a reliable indicator of relevance. This insufficiency of topical relevance lies in our inadequate understanding and implementation of the concept as well as not thinking about non-topicality variables of relevance. Green (1995) argues that "although topicality is a major factor in the establishment of the correspondence between a text segment and a user need, we have little real understanding of how the topics of text segments relate to the topics of user needs to which they are relevant". It is generally assumed that topical relevance is merely "on topic", i.e., a direct matching between the overall topic of documents and the overall topic of user needs, only considering a single relationship—matching, a single type—direct (matching), a single topic—the overall topic, and a single status—not varying with different user situations. Green (1995) points out that hardly any attention had been given to other possibilities. She further proves that topical relevance relationship is not a simple matching relationship, but also hierarchical relationships and structural (syntagmatic) relationships as identified in an empirical study. Other aspects of this basic unspoken assumption about topical relevance, treated as a single type, a single topicality, and a single status, would be further challenged by the results reported in this paper.

Study Context: Creation of the MALACH Test Collection

General Background

This study was done in the context of the development of the MALACH test collection. MALACH aims at improving access to oral history archives through automatic speech recognition (ASR) and subsequent information retrieval assisted by techniques from natural language processing (NLP) (Gustman, *et al.,* 2002; MALACH). MALACH works with the collection of 52,000 Holocaust survivor testimonies assembled by the Shoah Visual History Foundation (VHF). 4,000 of these testimonies have been carefully indexed: Trained indexers divided each testimony into topical segments, wrote a three-sentence segment summary, assigned 3-5 subject descriptors from the VHF Thesaurus and any persons mentioned, and wrote a half-page summary of the testimony as a whole.

We used this collection to develop an oral history speech retrieval test collection. Starting from 50 real topics – requests received by VHF – we configured a test collection of 400 testimonies yielding approx. 20,000 segments such that each topic would have a reasonable number of relevant segments for retrieval experiments. Four graduate students in history and information studies performed search-guided relevance judgments for 28 topics to produce a standard for IR system evaluation. Search-guided means that the judges did not judge every segment for every topic – a task that would be prohibitively expensive – but rather did thorough searches using a retrieval system based on the human indexing and then judged the relevance of the segments found (Cormack, 1998). This paper focuses on the nature of the relevance judgments.

Types of topical relevance

For a an in-depth analysis of the factors affecting retrieval performance we defined five categories of topical relevance with a view to the nature of the requests which mostly thought information about historical events and conditions and thus are concerned with evidence, which also led us to consider legal reasoning (Stong, *et al.,* 1999)(see Figure 1). Relevance was recorded on a five-point scale (0 – 4), which in a pilot the judges preferred over a three-point scale (see also Tang et al., 1999). Judges also recorded the fraction of a segments that pertained; the relevance score reflects the relevance of that piece, not matter how small. In addition, judges recorded a justification by linking a piece of summary text with a type of relevance. Additional information, such as the difficulty of the judgment and source of information used, were also recorded.

The search and relevance judgment interface

Judges worked with an interface that had a small query pane, a search result list pane, a pane for detailed information on one segment, and a pane for recording relevance judgments (see Figure 2). For each type of relevance there is a drop-down box to assign a scale value (default 0) and a slider to indicate the approximate percent of a segment that pertains.

Research Questions

To truly understand topical relevance, one must study interactions between the instrument and all the entities that are involved in the actual relevance judgments, including the system, the judge, and the collection. This paper focuses on the judge's perspective. For the MALACH judges, it analyzes their experiences with the instrument and their perceptions of different types of topical relevance. The specific research questions addressed in this paper are:

- How do judges perceive different types of topical relevance? What are the characteristics of each type of topical relevance?
- What are the roles of direct and non-direct topical relevance in judgments about overall topical relevance?
- Is there any overlap among different types of topical relevance? If there is, what is it and what makes it happen?
- What are the factors contributing to topical relevance?

Direct relevance
Direct evidence for what the user asks for. Directly on topic, direct aboutness. The information describes the events or circumstances asked for or otherwise speaks directly to what the user is looking for. First-hand accounts are preferred, but second-hand accounts (hearsay) are acceptable.

> *Direct relevance scale*
> 4 1st hand or verified 2nd hand accounts with lots of detail
> 3 4 with less detail
> 2 Even less detail
> 1 Mention of topic with little detail
> 0 Not relevant

Indirect or circumstantial relevance
Provides indirect evidence on the topic, indirect aboutness (data from which one could infer, with some probability, something about the topic, what in law is known as circumstantial evidence) Such evidence often deals with events or circumstances that could not have happened or would not normally have happened unless the event or circumstance of interest (to be proven) has happened This type takes precedence over context.

> *Indirect relevance scale*
> The scale values are as for direct relevance, considering also the strength of the connection between the testimony and the events or circumstances of interest

Context relevance
Provides background / context for topic, sheds additional light on a topic, facilitates understanding that some piece of information is directly on topic. So this type covers a variety of things. Things that influence, set the stage, or provide the environment for what the user asks for. Indirect takes precedence.

> *Context relevance scale*
> Considers the strength of the connection, including
> 4 "Immediate" precedent or consequence – very detailed and certain
> 3 4 with less detail and/or certainty
> 2 Not immediate, but very detailed and certain
> 1 2 with less detail and/or certainty
> 0 Not relevant

Relevance by comparison
Provides information on similar / parallel situation (at a different time or place) or on a contrasting situation for comparison. Identifies items that can aid understanding of the larger framework, perhaps contributing to identification of query terms or revision of search strategies.

> *Relevance by comparison scale*
> Similar to direct, considering also how much can be learned from the comparison

Relevance as pointer
Provides pointers to a source of more information. This could be a person, group, another segment, etc

> *Relevance as pointer scale*
> Considers the amount and importance of the information the pointer leads to and the quality of the pointer.
> 4 Very important
> 3
> 2 Unverifiable pointer
> 1
> 0 Not relevant

Overall relevance

Strictly from the point of view of finding out about the topic, how useful is this segment for the requester.? This judgment is made independently of whether another segment (or 25 other segments) give the same information

> *Overall relevance scale*
> Considers the amount and importance of the information the pointer leads to and the quality of the pointer.
> 4 Makes an important contribution to the topic, right on target
> 3 Makes an important contribution to the topic
> 2 Look at for exhaustive treatment of the topic
> 1 Look at to leave no stone unturned
> 0 No need to look at this at all

> **Considerations applicable to all relevance types**
> a Nature of the info.: level of detail, uniqueness
> b First vs. second hand account vs. rumor

Figure 1. Relevance types and their scales

Judgement for segment 89933

Pointers [0 ▼] Comparison [0 ▼] Context [0 ▼] Indirect [0 ▼] Direct [0 ▼]

Overall Relevance [0 ▼] Difficult ☐ Need audio ☐ Not enough information ☐

Sources Audio ☐ PIQ details ☐ PIQ Summary ☐ Other segments ☐

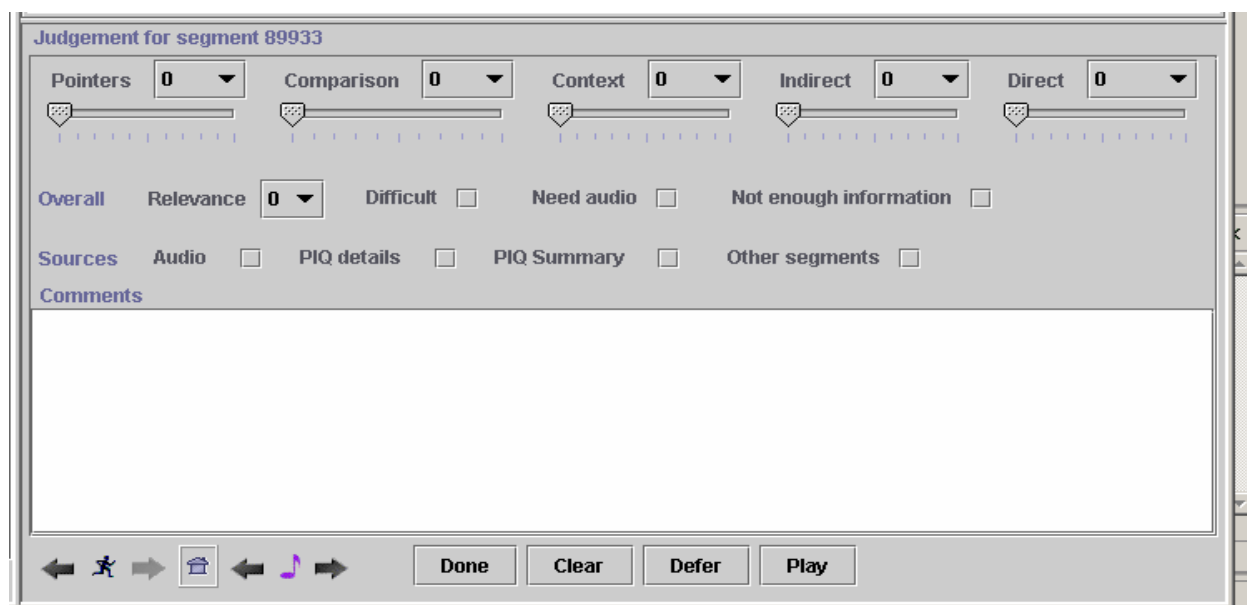Comments

Done Clear Defer Play

Figure 2. Search and judgment recording interface

Methodology

This paper reports on a qualitative study of the four judges who have used an instrument for rating topical relevance that incorporates different types of topical relevance and reports on their experiences and their perceptions of topical relevance after the task.

Participants

The four MALACH relevance judges.

Data Collection

Data were collected through interviews two to three months after the participants work as judges, not at the time of the actual relevance assessments. The interviews, therefore, do not reflect their perceptions as they did the judging but their remembrances of those perceptions and judgments based on the totality of the experience, as assessed after 2-3 months' lag between the judging and voicing those judgments. The interviews were semi-structured. All the interviews were audio-taped and transcribed to capture the rich original information from the participants.

Appendix A contains the interview questions used in this study. All the interview questions were designed to explore how the participants made meaning of their summer experience with the special instrument, without any theoretical assumptions. Most of the questions are open and address: 1) the participants' understanding of the different relevance types; 2) their perceptions of the associated scales; 3) their overall evaluation of the instrument; and 4) the influence on their following information behavior by using the instrument.

Qualitative Analysis

The study follows a grounded theory approach. Data were processed in three phases. In the first phase, the author read through the four transcripts to develop general perceptions of the content. Seventeen general theme groups emerged as a guide for the subsequent coding. In the second phase, each transcript was perused sentence by sentence. All occurrences of the specific themes/ theme elements were noted under appropriate general theme groups. In this phase, the theme identification was carried out in an exhaustive manner. Every statement that contained multiple themes was broken down into its individual components. In the third phase, the elemental themes were merged and organized into a coding scheme. After several rounds of grouping and merging, in the end, 17 major themes and 85 sub-themes were attained as shown in the attached coding scheme (See Appendix B).

Results and Discussions

Corresponding to the research questions, three major themes emerged from the four interviews. These themes are 1) the insufficiency of direct topical relevance for indicating subject relevance; 2) different topical relevance types as perceived by the participants; and 3) factors contributing to topical relevance.

Insufficiency of Direct Topical Relevance

As mentioned above, topical relevance is generally treated as a direct matching between the overall topic of a document and the overall topic of a user need. In the instrument this conception was considered *direct relevance.* Most content-based IR systems are based on this generally held but unspoken notion, which, in turn, affects human participants' perception of topical relevance and restricts their attention to a narrow focus of topical relevance. All the participants express their ignorance of non-direct type of topical relevance before participating in the project and feel excited about more options of topical relevance opened by the MALACH instrument. With these options, they have more flexibility in manipulating topical relevance judgments and emphasize the value of each non-direct relevance type under particular situations. As demonstrated by the following two sub-themes, as the single type of topical relevance that responds to diversity of documents and variety of user needs, direct relevance is clearly insufficient.

Greater Sensitivity to Non-direct Topical Relevance

The participants broadened their perception of topical relevance by increasing the sensitivity to non-direct topical relevance. The instrument forced them to consider additional forms of topic matching besides direct topical relevance, but they were able to make the adjustment to this broader perception and see the usefulness of doing so even later in their own research work. In other words, the participants expanded from the narrow focus on direct relevance to non-direct topically relevant information within documents. This improved sensitivity enables them to identify more useful materials that seem not relevant at first glance or that they would not have considered to be relevant before their work as MALACH relevance judges. This is an obvious influence of using the instrument on the participants' information behavior.

> I learned to look at other things beside the direct relevance. (P1)

> I think that [the instrument] allows you to look in more areas and I think when you do that you are bound to find more relevant research materials. That would certainly help any project. … I think that my relevance judgment before the summer might be more exclusively just direct, looking for something that specifically talked about my topic. And I don't think I was looking for things like pointers, that were not necessarily on topic at all but leading me in a direction something else was relevant. And certainly was comparison, you found something that is exactly opposite of something you are looking for but actually might be relevant for it provided a contrast. I think I would not look for that before the summer. (P2)

> This [making relevance judgments with the instrument] has sensitized me to issues of context, and things that are indirectly relevant. (P3)

> Try to find things in more categories, with more breadth. It [the summer experience] is kind of broadened my idea of what's relevant at all. … Well, I will be more able to zoom in

things that wouldn't appear relevant at first glance, would sharpen my idea of what I could possibly find that might actually be useful that I wouldn't have thought that before. … (P4)

Inadequacy of Overall Topical Relevance Score

In most cases, the direct relevance score figures most prominently in the overall relevance score while other relevance types only play a subsidiary role. The participants consider the overall topical relevance score misleading and insensitive to varied user needs and situations. They think that every non-direct topical relevance type can have a significant value or impact to a particular user under a particular situation, and that value or impact is not necessarily overridden by direct relevance all the time. For example, in some cases pointer relevance is of great interest to the user, in some cases the user pays more attention to comparison relevance, and in other cases the user particularly asks for contextual or circumstantial information. Therefore, direct relevance or an overall score based on direct relevance is insufficient to speak for the real topical relevance of a document. Users need to know more than that and make more informed decisions according to their needs. It is interesting to note that, once participants adjusted to seeing nuances in topical relevance as represented by the categories beyond direct topical relevance, they did not want to return to the older setting in which direct topical relevance was the only criterion.

> So after my experience in the summer, I don't know that I would look at the overall very much. I think I would look at all the five specific scales. … I would determine myself for a given topic which type of relevance was the most important or which area I was lacking in my research. It might be I already knew all of the direct, but I didn't have the background information, or I was writing specifically on comparison between camps, so I needed the context or comparison particularly. … You only need to look at the individual relevance scores, and you can make your own overall decision. (P1)

> …the overall score is kind of imposes my own values of what is important as far as relevance and it might be misleading to a researcher. The overall decision is up to the user. I think that depends on what the user is looking for, certain categories will be more important. So I think the overall score might be a little bit misleading in that way. (P2)

Different Types of Topical Relevance as Perceived by the Participants

Direct Relevance

Direct relevance is considered the major piece of topical relevance; a document that is directly relevant provides specific and rich information on topic in an explicit way. It is the most recognized and emphasized topical relevance type in the research and the practice of information retrieval. Direct relevance becomes the central meaning of topical relevance and most of the time it is even mistaken as the only meaning of topical relevance. The participants identified the following characteristics of direct topical relevance:

- *Right on topic:* direct relevance has high relatedness to the given topic; usually the overall topic of the directly relevant document exactly matches the overall topic of the user need.
- *Explicit statement of the topic:* within direct relevance, the topical relationship is not only an exact matching, but also a direct matching, with minimal inference involved.
- *Rich detail about the topic:* richness of information on topic is another important feature associated with directly relevant document. As pointed out by P2, a good direct has a wealth of details while a bad direct has much less information on the given topic.
- *Specific:* direct relevance is usually associated with high specificity.
- *Static:* this specific type of topical relevance is largely inherent within the document, bearing less influence from various user situations.
- *Obvious:* with explicit matching relationship with the target, it is easily identified by most users.
- *Great impact on overall relevance:* as the strongest sense of topical relevance, it has most significant impact on assessments of the overall topical relevance score.

> Direct was directly on the topic. Exactly, precisely, like what you are looking for. (P1)

> Direct is probably more obvious. It provides the testimony that deals exactly with what your topic is. It provides rich information specifically on the topic that you're dealing with. … I mean direct might not have that much detail but it is specifically on the topic. So within the direct, there is a scale within that; a good direct segment would have rich detail, a bad direct segment would be right on topic but would have less detail. So those are two different scales. … I think that for the most part, the direct scale really has most impact on the overall relevance. If there is a direct relevance and it is extremely detailed, I think the overall would be very high. (P2)

> Direct I found pretty much static. It's [direct is] relevant to the topic and it explicitly stated it. (P3)

> Direct relevance is something that is directly on topic, so it's pretty straightforward, you know, you can't get much more specific than direct. It really illuminates something about specific thing, the specific person, something…Right in the thick of what the topic describes, right on target. … Very useful, this is the thing that points you to what the client or whoever put the topic up cares about the most, deeply important. (P4)

Indirect /Circumstantial Relevance

Generally the participants found it difficult to understand and interpret indirect topical relevance. But all the participants agree that it is an implicit statement of the topic and it can provide specific information on the topic as direct relevance after making inference of the missing pieces.

- *Implicit statement of the topic:* inference is the central feature of indirect relevance; its relationship with the given topic is implicit that judges need to make their own inference to connect the topic of document with the topic of user need.
- *High relatedness to the topic:* after connecting the missing dots, indirect evidence is closely related to the given topic. "It was almost like direct" and "It has to have all the relevance of direct examples except it's just missing a specific piece of information". In other words, indirect relevance is an indirect topical matching relationship.
- *Specificity:* specific information is stressed for indirect relevance by most judges. Although just indirectly related to the given topic, it is still much more specific than context relevance.
- *Difficult to recognize:* due to its implicitness, indirect evidence is difficult to identify in the document and requires the judges to make their own inference from the document.
- *Infrequently used:* the difficult identification results in infrequent judgments of indirect relevance.
- *Great impact on overall relevance:* because of its high relatedness to the given topic, after connection, "most of the time it had the scale as the direct and contributed as much to the overall relevance".

> Indirect did not make sense to me in the beginning until we started talking about it as circumstantial evidence. And I started thinking about it in terms of US law and how court cases can use evidence that leads up to something and points to that but doesn't necessarily explicitly say it. So it's more implicit. … and most of the time, it [indirect] had the scale as the direct and contributed as much to the overall relevance. (P1)

> I think only one or two times I ever really did use it [indirect] and I think those times it was just very explicit information or very specific information that was just not telling us. It has to be much more specific [than context]. It has to have all the relevance of direct examples except it's just missing a specific piece of information. …it was almost like direct. (P3)

> For instances, medical care in the camps, an example where they complain about lice, pests and disease even if they don't say there was poor medical care, you can infer from their complaint that the medical care was poor. It's inferential measure. … So indirect was somewhat confusing and annoying to me, it was the one that I had most difficult means of interpreting it, and I'd say probably 20% of my results were indirect or maybe less. (P4)

Context Relevance

Context relevance provides background information about the topic. It is related to the topic only in a broad way, but it helps the user better understand the general picture how or where the given event or topic fits into it. It is defined in terms of setting or environmental factors, cause or consequence of the given event/topic, something that allows or

hinders the given event/topic to happen, and other supplemental information. Its features are:

- *Not a close match to the topic:* contextual information is "something just happening in the environment in the background but not directly about that topic", the overall topic of the document is not matching but only surrounding the given topic.
- *Remote relatedness to the topic:* the distance between the overall topic of document and the overall topic of user need is relatively bigger than other types of topical relevance. In other words, context "is just anything that might be remotely related". Toward an extreme extent, with the distance from the topic continuously increasing, any document can be somehow related to the topic in a broader and broader way. Thus, how good a context is is primarily determined by "how related it is".
- *Assisting a better understanding of the topic:* context relevance provides contextual information and usually take the following forms:
    - As setting of environmental factors
    - As precedent events
    - As consequent events
    - As background/supplemental information
    - As reasons behind the given event/topic
- *Not specific:* usually it does not provide specific information on a given topic.
- *Easy to recognize:* since this category makes intuitive sense to judges, it is easy to find contextual information in documents.
- *High occurrence:* because context relevance has relatively loose criteria on documents, it occurs much more often.

    Something just happening in the environment in the background but not directly about that topic but that allowed that topic to happen. … The context was more external factors, (P1)

    I also used context for follow-up information, if it discussed a man's life after he was liberated from a concentration camp. That might be context because it doesn't deal specifically with your topic but it gives you an idea of what was going on in the life of these people following or preceding the area your topic is dealing with. … It [context relevance] is not specifically on topic, but it's nearby. (P2)

    Context for me is just anything that might be remotely related. How related it is determines its level of relevance. … The environment, the situations, the policies, everything that the researcher would want to know about, to backup the argument, but not to base the argument about. (P3)

    Well, it [contextually relevant document] can be either from it supports or hinders the given event, the given topic, or it can also be something that happened before or after, the precedent or the consequence. Helps illuminate what's the process behind, the reasons for what's going on in the general topic. (P4)

Comparison Relevance

The participants interpret comparison relevance in two different ways: something not about the topic but very similar to the topic and something about the topic but with some difference. Both interpretations are correct and vary only in their emphasis on similarities or differences.

- *Not exactly match the topic but similar to it:* with comparison relevance, we cannot draw an exact matching relationship between a document and a given topic. The document topic partially overlaps with the given topic, with their intercept representing their similarities. Thus, the relatedness of comparison document to a given topic is primarily based on its similarity to the topic.
- *Comparison/contrast:* as suggested by the participants, contrast by its nature is also included under this relevance type. On the one hand, the event/topic is similar to the given event/topic; on the other hand, certain environmental factors or the subject are different. Thus, it can be the same event happening in a different place, at a different time, in a different situation, or on a different person.
- *Specific:* it provides specific information as direct, which makes it distinguished from the contextual information. To a large extent, "it was the same as what I developed for my direct relevance".
- *Sometimes particular important:* since it provides rich relevant information as direct but with one or two factors different, comparison relevance has special value, especially when little material is exactly on topic.
- *Infrequently used:* it occurs less often than contextual information.

    It was the same as what I developed for my direct relevance but there were one or two factors different. So either it was a different ship, or a different year, or a different place, different camp, different president, different country. … It [comparison relevance] is useful when there was little about the topic and so the comparison may shed light on the topic when you had nothing better. (P1)

    I think comparison has more general meaning: comparison and contrast. (P2)

    Comparison I found to have a lot of specific information, a lot of important information but the scenario, or place, or person that's different from what the topic asking for. … A high comparison would depend on how much information has been given and how related it is to the topic. … I think comparative information can be very important sometimes. (P3)

    The comparison showed up less often. … (P4)

<u>Pointer Relevance</u>

The participants perceive pointer relevance as a bridge leading to more information on the topic. If a document contains pointers that are useful, then they consider the document useful even if the document may not touch on the topic at all. They identified the following characteristics of pointer relevance:

- *Not necessarily on topic:* the document/segment that contains the pointer may not deal with topic at all, in other words, the overall topic of the document may have a weak relationship, if any, with the overall topic of the user need. The assumption that topical relevance should be the overall topical matching, exactly or partially, between the given document and the user is not held for this specific type of topical relevance.
- *Leading to more relevant information:* the central feature of pointer relevance is that it can point you to a wealth of information about the given topic.
- *Specific*: as mentioned by all the participants, pointers are usually very specific and have following manifestations: a specific person, a specific book, a specific organization, a specific place, a specific event, a specific law or court case, etc.
- *Small amount of relevant information*: pointers only show the direction and thus provide very limited information about the given topic. Usually, they are only brief descriptions without detail.
- *Easily identified:* pointers are described to be very obvious to recognize and easy to pick out in a document/segment as a type of topical relevance. Although it has nothing to do with the overall topical matching which is required by content-based IR systems, in the eyes of participants, it is quite helpful and handy for working on a given topic.
- *A dichotomous issue*: related to its extremely limited amount of information, a pointer is seen by participant as either yes it is a pointer or no it is not a pointer, without much middle ground. A dichotomous scale is more suitable for pointer relevance than a four-point scale.

It [pointer relevance] was very straight forward, in overall you could easily pick out a pointer. (P1)

This is like testimony itself might have nothing to do with the topic you're dealing with, but it has a name of a person who may have more information on this; it may have the name of a book that if you got that book, it would give good detailed information on your topic; it might discuss a law or a court case that deals with the topic that you're dealing with. So in itself, it is meaningless but the direction it leads you might help you with your research. (P2)

By its very nature it [pointer] is only a small amount of information that just point you to another topic. So by its very nature it's almost a yes or no for me, not a scale. (P3)

Factors Contributing to Topical Relevance

In the previous discussion, to capture the individual "meaning" of each type of topical relevance, each type is analyzed as a discrete category. The next step is to inspect the interrelationships among different topical relevance types to gain more insights into the nature of topical relevance as a whole. This issue is approached by analyzing the overlaps among different relevance categories and then the common dimensions/factors underlying different topical relevance.

Overlap among Different Relevance Categories

The overlap among different types of relevance manifests itself as reported difficulty of distinguishing between them.

<u>Direct Relevance and Indirect Relevance</u>

The overlap between direct relevance and indirect relevance indicates that topical relevance is not merely an issue of matching or not. It involves more factors than that, such as inference. For indirect relevance, after making certain inference, it can exactly match the topic as well as direct relevance. Both of them are "on topic", but one is directly without inference while the other is indirectly with more or less steps of inference. However, as the participants point out, sometimes "the implication was so obvious that it might be direct" (P2). The involving "inference" is definitely not a yes/no issue but a matter of degree, or a continuous scale. Instead of drawing a clear line between direct relevance and indirect relevance, we can only say that, towards on end of the inference scale is more of direct, while towards the other end is more of indirect.

Often it was difficult to differentiate indirect and direct because the implication was so obvious that it might be direct. It just didn't connect the dots entirely. So you had to make your own inferences but it was directly on topic after you made those inferences. (P2)

<u>Indirect Relevance and Context Relevance</u>

For indirect relevance, the overall topic of a document implies the overall topic of a request; for context, the overall topic of a document surrounds the overall topic of a request: "it's [the document] not specifically on topic, but it's nearby" (P2). To simplify, both indirect relevance and context relevance do not "match" the topic at the beginning: indirect relevance does not match the topic directly, while context relevance does not match it specifically. This similar aspect, that is, not matching, makes them sometimes seem to overlap with each other. Generally speaking, indirect relevance is much more specific than context and, after inference, it is more closely connected to the given topic. The scale for "specificity" or "relatedness" is also a continuous one, resulting in ambiguities in middle range.

As I said, indirect and context I think have a lot of overlaps. (P1)

It [indirect] has to be much more specific than context. … Context information was connected to the topic only in a very broad way. ... if it was very closely connected, it might fall into the indirect relevance. (P2)

Direct Relevance and Context Relevance

As the participant describes, while the context relevance is getting closer and closer from the peripheral area to the central area of the given topic, context relevance merges into direct relevance. Along with the increasing relatedness to the given topic, the information provided by a document becomes more and more specific. Both the "relatedness" and "specificity" are contributing to this merging process.

I thought that direct and context graded into each other at several points. Liberations in Buchenwald and Dachau was a particularly interesting topic where I think seeing overlapped into each other. The context would be before when there was the rumors of hearing the allies coming, hearing the gunfire and the bombing, and anticipate the rescuers. That would be context. As the liberations approaches, when the Nazi guards start fleeing, the liberation itself is direct. Context kind of grades into direct there…and then it becomes clearly context when they talk about their later life in Israel…It sees like a bell curve, rising context low to context high to direct high, and after liberation, context high, and context low, and bottoming off. (P4)

Direct Relevance and Pointer Relevance

Within pointer relevance, the overall topic of a document may have no intercept at all with the overall topic of a user request. Pointer itself provides nothing but a helpful direction, so that only minimal information within the document is relevant to the given topic of the information need. However, as noted by a participant, along with the increasing amount of relevant information within the document, the intercept between the overall topic of document and the overall topic of user request gets larger, and eventually reaches the same effect as direct relevance. In other words, the amount of relevant information is a very important factor here, which leads to the changes in degree of relatedness to a given topic.

At times a pointer would also be direct, say for the segment on CANDLES, they talked a lot about what the organization was and what it did, that would be ordinarily direct, or as opposed to merely mentioning it, which is what a pointer is. (P4) Note: CANDLES is an acronym for Children of Auschwitz Nazi Deadly Lab ExperimentS; for the topic of "medical experiment", CANDLES is a pointer leading to more information on the topic.

Direct Relevance and Comparison Relevance

As discussed above, comparison relevance has a relationship of comparing or contrasting to the given topic. For comparison/contrast relevance, with something similar and something different to a given topic concurrently, the overall topic of a document does not completely but only pa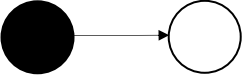rtially matches a given topic. When the topical intercept between them, that is their similarity, increases, the comparison/contrast relevance turns into direct relevance.

I think direct and comparison have a lot of overlaps in that often times you would have the same relevant information which just be like one specific piece is different, that's why it would be comparison and not direct. (P2)

Common Dimension/Factors Underlying Different Topical Relevance Types

Based on the findings in the previous sections, the different topical relevance types considered in this paper vary on at least five dimensions: on topic, overall topical relationship type, degree of relatedness to the given topic, specificity, and the amount of relevant information/richness (See Figure 3). These dimensions are related to and vary with one another. An increasing amount of relevant information/richness leads to increasing relatedness to the given topic. In turn, increasing relatedness to the given topic usually leads to increasing specificity of information provided by a document. The extent of matching between document topic and information request topic decides the overall topical relationship type and affects the relatedness to the given topic as well.

Different topical relevance is primarily decided by the first two dimensions: how on topic and the overall topical relationship type. Context relevance is a surrounding topical relationship; comparison/contrast relevance is a partially matching relationship; pointer relevance has no overall relationship with the given topic at all; direct relevance is direct matching; and indirect relevance is implicit matching. But different topical relevance types are also featured by typical values on other dimensions, for example, context relevance type is often associated with low specificity of information provided by the document and low relatedness to the given topic, and pointer relevance is usually characterized by small amount of relevant information. To some extent, different topical relevance is a specific combination of these factors, rather than a single-dimension issue.

Topical Relevance Type	On topic	Relationship between overall document topic & overall user need topic	Diagram Illustrating Relationship[1]	Relatedness to the given topic	Specificity of information	Richness of relevant information
Direct	Directly	Matching		High	High	High ~ medium
Indirect	After inference	Matching		High ~ medium	High	High ~ medium
Context	No	Surrounding		Low ~ medium	Low	Low ~ medium
Comparison	Partially	Comparing/ contrasting		Low ~ medium	High ~ medium	Medium ~ low
Pointer	No	Apart		Low	High	Low

1   In the diagram, the document topic is black; the user need topic is white.

Figure 3. Common Dimension or Factors Underlying Different Types of Topical Relevance

Conclusions

Topical relevance is a richer and more varied concept than previously considered. The analysis shows that:

- Direct topical relevance is not sufficient as a means of expressing the important concept of topical relevance.
- More than only one dimension/factor, "whether it is on topic" as usually referred to, is involved. Other factors/dimensions such as the overall topical relationship type, degree of relatedness to the given topic, specificity, and the amount of relevant information/richness should also be included.
- More than only one topical relationship type, "direct matching" as generally assumed, plays important role in topical relevance. Topical relevance can be decomposed into multiple types, such as direct relevance, indirect relevance, context relevance, comparison relevance, and pointer relevance. Each type has its own characteristics on different dimensions.
- Topical relevance is situational; it is a combination of factors with different weights according to different preferences or user situations.

Essentially, topicality is still the core of content-based IR systems and has the greatest influence on retrieval results. IR systems based on the single type of "direct overall topical matching" are quite limited. Just as relevance is not only topicality, topicality, in turn, is not just "direct overall topical matching". As demonstrated above, many other factors contribute to topical relevance and should be considered in the design of retrieval algorithms. In that lies the imperativeness to advance thinking on topical relevance.

Topical relevance still remains a black box. Little of the mystery of this complex concept has been solved. By taking advantage of the MALACH project, this paper explores an enriched topical relevance concept from the user's or judge's perspective. Discussion has, of necessity, focused on the five types of topical relevance identified and used in the MALACH relevance instrument. Developing an exhaustive list of topic relevance types or topical relevance factors is not an objective of the MALACH project nor of this paper. Thus, although the findings shed light on this complex concept, they are, by nature, limited in revealing all the variables of topical relevance. Future studies could extend and complement this framework by eliciting more variables and relationships directly from the user.

REFERENCES

Barry, C. (1993). The identification of user criteria of relevance and document characteristics: Beyond the topical approach to information retrieval. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.

Cool, C., Belkin, N., Kantor, P., & Frieder, O. (1993). Characteristics of texts affecting relevance judgments. In M.E. Williams (Ed.), *14th National online meeting: Proceedings* (pp. 77-84). Medford, NJ: Learned Information.

Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7,* 19-37.

Cormack, G. V., *et. al.*(1998). Efficient construction of large test collections. *SIGIR '98,* 282-289.

Cuadra, C. A. & Katter, R. V. (1967a). *Experimental studies of relevance judgments: Final report Vol. 1 Project summary, Vol. II Description of individual studies* (Report No. TM-3520; NSF C-424). Santa Monica, CA: System Development Co.

Cuadra, C. A. and Katter, R. V. (1967b). Opening the Black Box of 'Relevance'. *Journal of Documentation,* 23 (4): 291-303.

Froehlich, T, J. (1994). Relevance reconsidered—towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science, 45(3),* 124-134.

Green, R. (1995). Topical relevance relationships. I. Why topic matching fails. *Journal of the American Society for Information Science, 46(9),* 646-653.

Green, R. & Bean, C. A. (1995). Topical relevance relationships. II. An exploratory study and preliminary typology. *Journal of the American Society for Information Science, 46(9),* 654-662.

Gustman, S., Soergel, D., Oard, D., Byrne, W., Picheny, M., Ramabhadran, B., and Greenberg, D. (2002). *Supporting access to large digital oral history archives.* In Proceedings of the Joint Conference on Digital Libraries (JCDL), 18-27.

MALACH project. Retrieved from January 29, 2004 from: www.clsp.jhu.edu/research/malach/

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

Park, T. K. (1992). The nature of relevance in information retrieval: An empirical study. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Saracevic, T. (1976). Relevance: a review of the literature and a framework for thinking on the notion in information science. *Advances in Librarianship, 6,* 79-138.

Schamber, L., *et. al.* (1990). A re-examination of Relevance: Toward a dynamic, situational definition. *Information Processing & Management, 26(6),* 755-776.

Schamber, L. (1991). Users' criteria for evaluation in multimedia information seeking and use situations. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.

Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology (ARIST), 29,* 3-48.

Stong, J. W., *et. al*. (1999). McCormick on Evidence (5th ed., Sec. 185). St. Paul, Minn.: West Group.

Tang, R., *et. al.* (1999). Towards the identification of the optimal number of relevance categories. *Journal of the American Society for Information Science, 50 (3),* 254-264.

Wang, P. (1994). A cognitive model of document selection of real users of information retrieval systems. Unpublished doctoral dissertation, University of Maryland, College Park, MD.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval, 9,* 457-471.

VHF. Retrieved from January 29, 2004 from: www.vhf.org/

Appendix A. Interview Questions

1. Thinking back on your work in the summer, what impressed you most? What did you learn from this experience?
2. When searching in your own work, you often need to decide whether a document is useful for your purpose. What factors do you usually take into account for these decisions?
3. When we talk about "relevance", what comes to your mind?
4. How did you make relevance judgments before?
5. What is other influence on your consequent information searching or using?
6. Please simply describe the procedure that you followed in assessing relevance to a given topic?
7. Please express briefly in your own words what the following relevance categories mean. Please also tell me how you understand the function or role of each category.
   o Direct relevance
   o Indirect relevance
   o Context relevance
   o Comparison relevance
   o Pointer relevance
   o Overall relevance
8. Please describe the overlap among the categories.
9. When you are doing a search, what type(s) of relevance will you pay more attention to? Please rank the relevance categories according to their importance.
10. What scale do you think is most appropriate for relevance judgment? Why?
11. Between which of the following two points (0 and 1, 1 and 2, 2 and 3, 3 and 4)did you feel most difficult to differentiate?
12. When judging a segment, how did other segments you'd already judged influence your judgment for this particular segment?
13. How did you feel when you were using the instrument, both cognitively and emotionally?
14. Assuming you are going to design a relevance instrument on your own, how do you want to design it? Can you sketch your design for me?
15. What are the benefits and constraints of using such an instrument?
16. From your perspective, how useful are these categories and scales for academic info searching and daily info searching?

Appendix B. Coding Scheme

A sense of relevance
    Topical factors
    Non-topical factors
Direct relevance
    Primary user concern
    Right on target
    Static
    Explicit statement
    Richness of information on topic
    Relationship with detail
    Specificity
    Easy identification
    Great impact on overall relevance
    Agreement on terminology
Indirect relevance
    Confusing nature
    Different understandings
    Inference
    Relatedness to the topic
    Information on topic
    Specificity
    Infrequent use
    Level of usefulness
    Impact on the overall relevance
    Discussion of terminology
Context relevance
    As environmental factor
    As precedent
    As consequence
    As background/supplemental information
    As reason behind the event
    As public policy and laws
    Remote relatedness with the topic
    Specificity
    Level of importance
    Impact on overall relevance
    Frequency of use
    Discussion of terminology
Comparison relevance
    Amount of information
    Relatedness to the topic
    Contrasting
    Manifestations
        Different place/environment
        Different time
        Different situation/ scenario
        Different person
    Impact on overall relevance
    Level of usefulness
    Frequency of use
    Discussion of terminology

Pointer relevance
    Lack of information on topic
    Lead to more information
    Lead to more queries
    Manifestations
        Specific person
        Specific book
        Specific organization
        Specific place
        Specific event
        Specific law or court case
    Importance
    Impact on overall relevance
    Easy identification
    Frequency of use
    Agreement on terminology
Importance of specific relevance types
    Dependence on user need
    Dependence on collection
    Dependence on topic
    Ranking
Overall relevance
    As sum/combination of specific relevance types
    Overall score can be misleading
    Calculation formula
    Importance
    Relatedness to the topic
    Contain of additional judgment information
    Detail
    Accountability
    Reliability of the source
Overlap among types
    Between indirect and direct
    Between indirect and context
    Between direct and context
    Between pointer and direct
    Between direct and comparison
Preference in choosing the types
Scale for relevance judgment
Influence of previous judgments
Evaluation of the instrument
Application of the instrument
Suggestion to the instrument
Holocaust testimony collection

Learning from the summer work
    MALACH influence on relevance judgment
        Greater sensitivity to non-direct relevance types
        Ability to find more relevant materials
    Boolean searching
    Free-text searching
    Thesaurus use
    Query formulation
    Procedure of relevance judgment
    Holocaust knowledge
    Mechanism of database