

CHARACTERISTICS OF TEXTS AFFECTING RELEVANCE JUDGMENTS

C. Cool*, N.J. Belkin*, O. Frieder#, P. Kantor*
*Rutgers University and #George Mason University

Keywords: Relevance judgments, usefulness, information retrieval

Abstract: We report on preliminary results from two studies, investigating the factors which underlie people's judgments of the relevance or usefulness of documents to particular information problems. The studies aimed to see whether factors other than topical relevance are significant in such decisions, to identify any such factors, and to relate such factors both to the users' situations and goals, and to representation and retrieval strategies for IR systems, which would take account of such factors. Six facets of the judgment of document usefulness or relevance were identified, only two of which seem directly related to topical relevance. These facets are interpreted in the framework of user goals and situation, and some example strategies for their use in IR systems are presented.

1. INTRODUCTION

Although the *relevance* of a document to a query has long been the primary criterion according to which information retrieval (IR) systems have been designed and judged, there is an equally long history of dissatisfaction, or disquiet, with this criterion. See, e.g., Doyle(1) Cooper (2); Su (3), for discussions of this issue. The problem appears to be that relevance, as most traditionally construed, deals with *topical* relationships between document and query, while the judgments that people make about the *usefulness* of a document to their information problems seem often to be based upon other criteria (see Saracevic (4) for a review of the concept of relevance, and its problematic aspects). Additionally, topical relevance itself may be a complex phenomenon, with judgments being based upon a variety of aspects of the *aboutness* of the document in relationship to the information problem. There has been a fairly recent resurgence of interest in the concept of relevance and how one might address it in the design and evaluation of information retrieval systems, e.g. (5). In this paper, we report on preliminary results of two empirical research projects which aim to investigate the dimensions of relevance, or more broadly, usefulness, of documents.

The projects that we are engaged in address the following questions:

- What are the relationships between a person's goals (or information problems) and the documents used in responding to those goals (problems)? That is, what are the uses that people will make of the documents, and how do they judge (evaluate) documents with respect to those uses?
- Are there characteristics other than topical relevance which affect a person's evaluation of a document's usefulness?
- Are there ways to decompose the concept of topical relevance into constituent dimensions?

- Are these characteristics/dimensions (potentially) representable and usable in the support of information interaction/retrieval?

The overall aim is to try to make IR systems more directly responsive to users' concerns and interests, by understanding the conditions of importance to users in IR systems, and making characteristics of those conditions available to the representation, retrieval and processes in the IR system.

2. THE STUDIES

We report on two studies of quite different sorts, whose results complement one another rather nicely. One is concerned with the judgments made by students of documents which they might use in writing an assigned essay; the other is concerned with the interactions between scholars and texts in support of all of their scholarly activities. Below, we describe these two projects in detail, and discuss their relationships in terms of the overall questions listed above.

The first study (the *GMU* study) aimed explicitly to identify aspects of documents and people's situations that lead them to decide whether or not to use a document in response to a given goal; and, to relate these results to the place of the user in the problem solving process. This project is a collaboration between Rutgers University and George Mason University, and is seen as the first step toward the design of IR systems which attempt to represent and use characteristics of texts other than topicality in the retrieval process. The study was organized as follows.

Approximately 300 freshmen, taking the introductory computer science course in the engineering program at George Mason University, were given an assignment to write an essay on a topic of general computer science interest. They were required to cite at least five sources in this essay. In addition to researching and writing the essay, the students were asked to fill out a questionnaire for each document that they considered using for the essay, at the time that they looked at that document. The instructions to students, and the questionnaire, are included as Appendix A. Briefly, the students were asked to: identify the document; indicate whether they thought that they would use it for their essay; say why they made this decision; and, indicate where they were in the process of writing the paper when they made this decision. These questionnaires were submitted together with the papers. A content analysis of the narrative reasons that were given for the decisions was done, in order to identify factors, or *facets* relevant to the judgment process. The preliminary results of this analysis are presented in section 3. Further analysis will relate these facets to the actual decision made, and to the state of the person in the problem solving process.

In contrast to the GMU study, which considers relevance judgments in relation to a predefined goal, the purpose of the second investigation is to understand something about the prototypical tasks and activities engaged in by humanities scholars, the goals associated with each, and the information interactions characteristically associated with these objectives. A particular focus of the study was on the nature of information seeking interactions between scholars and the texts they use and the ways in which texts are evaluated.

A qualitative research approach was taken in this investigation. Eleven in-depth interviews were conducted with senior level scholars at Rutgers University in the departments of history, English and philosophy. The interview was constructed around the following topic areas: typical tasks engaged in by the scholar; information resources used in the support of these tasks; typical information interactions engaged in by the scholar; and, specific information interactions in relation to the goals identified above.

This study was especially concerned with the investigation of the relationship between the humanities scholars' goals or information problems and the documents used in responding to those goals. During the course of the interview, an effort was made to understand the specific uses these scholars make of texts, and the judgment processes they engage in when they evaluate documents intended for such uses. These interviews were carried out within the framework of a

joint investigation between the School of Communication, Information & Library Studies and the Center for Electronic Texts in the Humanities, both at Rutgers University, and Bell Communications Research.

These two studies clearly investigate quite different situations, with quite different methods. The former is concerned with students, facing a predefined task, and is focused on the specific characteristics which they use to judge documents with respect to that task. It addresses these issues by studying a large number of people, and a few categories of data, looking for some regularities in their behaviors, without necessarily addressing the reasons for those behaviors. The latter is concerned with experienced scholars, in their entire scholarly life and associated problems and goals, and is focused on their interactions with texts in general. It addresses these issues by studying in depth a relatively small number of people, in their activities in general, and attempts to understand the processes which influence their uses of documents. We report here on these two studies together, because their results complement one another, as well as indicating points of similarity between the two different populations.

3. RESULTS

Data from both studies are currently being analyzed. Here, we present preliminary results.

3.1 The GMU Study

From the study of college freshmen, we identified six facets of the judgment process, shown in Table 1. The entries in each facet are examples of the types of specific factors that were mentioned by the students, and are not exhaustive lists. Our method in arriving at these facets is as

TOPIC (How a document relates to a person's interest)

Defines the topic itself; On/not on the topic; Focus (directly on topic, or not); Part of topic; Treatment (deep/superficial); Important (+/-)

CONTENT/INFORMATION (Characterization of what is 'in' the document itself)

Basic concepts; Facts/factual; Explanation; Examples; Definitions; Connections; Description; Reasons; Ideas; Tips; Guidelines; Technical knowledge; Interview; (About) people; Variety; Point of view; Survey; History; Level of detail

FORMAT (Formal characteristics of the document)

Lists; Diagrams; Statistics; Pictures; Class text; Book review; Title; Introduction; Division into topics

PRESENTATION (How a document is written/presented)

Organization; Matter-of-factness; Precision; Writing style; Understandability; Technicality; Scientificness; Simplicity/complexity

VALUES (Dimensions of judgment - these are modifiers of other facets)

Interest (+/-); Amount (lot/little); Specificity (specific/general); Goodness (+/-); Usefulness (+/-); Age (of document); Entertainment value (+/-); Precision (precise/vague); Bias (+/-); Authority (+/-)

ONESELF (Relationship between person's situation and the other facets)

Need; Utility; Desire ('want'); Like; Teaches; Informs; Supports understanding; Use to which document will be put (e.g. references, quotes)

(+ / -) indicates the positive and negative aspects of the dimension

Table 1. Facets of judgment of document usefulness

follows. We first noted all of the reasons given by the respondents for their judgments of documents, abstracting slightly. This gave us the opportunity for some preliminary grouping of reasons, which was based on the words the respondents used. Thus, a preliminary 'Topic'

category was composed of all of the phrases that contained the word 'topic', and a preliminary 'Adjectives' category was composed of all phrases containing evaluative adjectives. These data in these preliminary categories were then reduced again by grouping similar phrases, and by breaking up compound phrases (adjective plus topic, for instance) into their component categories. The categories themselves were then modified (phrases containing the word 'information' and those having to do with description of content being conflated into one category, for instance; or, the list of adjectives being reduced to a set of 'Values'), leading to the set of facets displayed in Table 1.

Several aspects of this set of facets are of interest. First, we note that of the six, only two, Topic and Content/Information, deal explicitly with what the document is about, the traditional basis for relevance decisions. The other facets, all of which were cited by people as the basis for deciding whether to use a document or not, deal with non-topical dimensions of the documents, or of the user's situation. It is also of interest to note that the 'aboutness'-related facets seem to be multi-dimensional. That is, for instance, not only is whether a document "on a topic" important, but also, the treatment of the topic, whether the topic is a main or subsidiary focus of the document, and so on, are significant characteristics for the potential users of the documents. The facet, Oneself, is also of interest. It indicates directly that aspects other than topical relevance, such as use, the person's needs or desires, and other relationships to the person's situation, are significant in the judgment process. Note that these have been provided spontaneously by the users of the documents.

3.2 The Humanities Scholars Study

The analysis of the data from the study of humanities scholars, while preliminary, suggests a complex relationship between the humanities scholar's tasks, goals and his/her judgments of texts. To begin with, the tasks and goals that are mentioned by these scholars in their descriptions of their professional lives are multiple and overlapping. In contrast to the GMU students who were assigned one task - that of completing a term paper - the humanities scholars in this study talked about the many tasks that define their professional lives. Tasks involved in teaching, research, writing, service activities, and other things as well. For each of these activity or task areas, the scholars were able to associate specific goals, and then to discuss the nature of their information interactions in relation to these goals and activities.

From these interview data, several interesting patterns seem to emerge, which reflect the nature of the relationship between a scholar's goals and her/his information behavior, and why particular aspects of texts are relied upon in judgments about a document's usefulness. In their evaluation of the usefulness of a text, the scholars typically rely on multiple aspects of texts, in relation to more than one goal.

For example, in order to keep up to date in their field, and to identify new problem areas, the humanities scholars typically scan a wide periodical literature. In order to select useful articles to read from this wide body of literature, many of the scholars look first at characteristics of the text such as title or author, to judge topical relevance of a specific text. These document characteristics are most useful in identifying texts which are *not* interesting to the scholar. If the text appears interesting or relevant based on title, other criteria may be employed. The title of an article may appear to be on a topic of interest to the scholar, but the author may be unknown; or the author may be a well-known scholar, but the specific topic of the article may be unspecified. In these ambiguous cases, the scholars frequently examine an additional attribute of the text, such as the footnotes, or the acknowledgments, to identify the author's specific point of view, or reference group. From this aspect of the text, the scholar can tell something about the way in which the topic is treated in the text, and about who the author is. In this example behavior, the of the text involves more than an assessment of topical relevance; it involves establishing the author's 'point of view', determining the credibility or authority of the author; and, determining the scholarly community the text is intended to address.

Other examples illustrate this same point, that why, or how, a person uses a particular attribute or characteristic of a text in judging its usefulness is related to broader questions about the person's problematic situation, and specific information-related goals or objectives.

Throughout the interviews, the humanities scholars mentioned aspects of content in their descriptions of reasons for judging a text useful or not. However, these references to content attributes were tied closely into discussions of goals and objectives. For example, a scholar who is in the early stages of a new project told us that his goal is to identify experts in the topic area. This person uses the following method. After selecting several top journals in the topic area (mental health policy), he looks at every article in every journal that seems to be topic related, according to the title or abstract. The usefulness of the text is determined by the extent to which it contains references to subject matter experts; in this case, other aspects of the text such as the nature of the topic treatment, or even quality of text, are less important than references to who is participating in the scholarly discourse on the topic. The goal of this preliminary stage of research is to develop a list of experts, with the subsequent goal of identifying manuscripts written by these experts. The longer-range research goal is to locate and analyze these manuscripts themselves, not the journal literature. For this person, specific format characteristics do not determine a text's usefulness, but there is clearly a sense of the content or information criteria that can be used.

The role of topic relevance in judgments of document usefulness among humanities scholars is quite interesting. For many of those interviewed, the determination of 'topic relevance' was only a first step, as in the previous example. In other cases, determination of topic relevance was more complex. In some interviews, scholars spoke of using manuscripts (which are not topically specified) to "guide them to other relevant sources". For example, one scholar told of how he routinely looks in texts across various disciplines, for 'evidence' that might be useful in further developing his position. These pieces of evidence might be statistical data, examples from legal cases, or current event happenings that appear in the news media. In these examples, the scholar has some general sense of the usefulness of these discrete bits of texts, which may not be topically related to his research subject at all.

An interesting finding in this study is that the humanities scholars did not mention presentation characteristics as being important in their of research-related texts. Although some of the interviewees mentioned such characteristics as writing style, or format, in discussing their selection of texts for classroom use, when it came to talking about their own judgments of texts for research use, these characteristics were not mentioned. In fact, many of the scholars we talked with told of their ability to adapt to different presentation modes, different styles of writing, different languages, and so on, in the texts they interacted with. It seems as if that what we see here is a normative sense of the range of presentation modes that a humanities scholar might be expected to interact with, and as such, this is not an important judgment criterion.

3.3 Summary of Results

We point out that the results of these two studies complement one another in interesting ways. First, for both groups of people, factors other than topical relevance were clearly important in decisions about whether or not to use a document. Indeed, there is a significant overlap between these two groups in the kinds of characteristics that are important. Second, there are also significant differences between the two groups in the kinds of decisions that are made, and the categories that are used to support them. This suggests that the nature of the user's situation (knowledge, goals, etc.) is significant in determining categories for relevance or usefulness judgments. And third, the data from the humanities study strongly support the facet: Oneself, in the results from the GMU study, and indicate how this facet may be related to the goals of the user, and the problem that the user is trying to address.

4. DISCUSSION

The data from our two studies converge on the same issue from two different yet related perspectives. The GMU study elicits brief descriptions of aspects of usefulness, in a narrowly controlled situation, from a homogeneous group of students. The study of humanities scholars elicits extended narratives ranging over a complex variety of information situations. To integrate the results further we might analyze the scholars' narratives, with the key facets revealed in the GMU study as points of reference. This will establish the extent to which the facets revealed in this narrow situation are, nonetheless, indicative of a broader reality structure underlying the evaluation of texts, for relevance, by scholars in complex research situations. Conceivably the results of this investigation could lead to a structured interview instrument which might guide scholars to address the several facets explicitly, and to assess their relative importance in a situationally and contextually dependent fashion with particular attention to the information retrieval problem. Developments of this type will help us to enlarge and confirm the conceptual scheme outlined here, and to make it more amenable to adoption by other scholars and researchers.

Reviewing the stated goals of this research, we see that we have demonstrated that scholars makes use of documents in a rich diversity of ways, which directly affects their judgement of usefulness, in terms that may reflect much more (or less) than mere topicality. We have elicited, from students, explicit constructs and terms indicating aspects other than topical relevance which affect the assessment of a document's usefulness. Our preliminary content analysis of those terms suggests that the terms define a small number of constituent dimensions, which serve to organize the notion of "non-topical" characteristics for further research.

From both of these studies, we have identified characteristics of the user's situation, and of documents other than topical relevance, which are important in their decisions about whether a document will be useful or not (or, whether to use a document for a particular purpose). In addition, we have identified aspects or dimensions of topical relevance, which indicates that this is not a unitary construct associated solely with some measure of the aboutness relationship between query and text. The significance of the user's goals, and of the use to which a text will be put, in the decision-making process, has also been demonstrated in both of the studies. The results of the humanities study are, in addition, at least tentatively explanatory of the uses of the various text characteristics, and of the relationships of these characteristics to other aspects of the user's situation. So, these studies have provided at least partial answers to the first three questions which motivated this paper.

The fourth question, are the new characteristics/dimensions which we have identified (potentially) representable and usable in the support of information interaction/retrieval, remains to be addressed. It seems clear that some aspects could be automatically used in the support of IR directly. For instance, characteristics associated with the Format facet can clearly be directly assigned to a document representation and used for supporting query formulation and retrieval. Indeed, traditional descriptive cataloging already does this sort of indexing, although it is rarely used for retrieval. Aspects of the Presentation facet could also be automatically assigned with some minimal forms of text, or natural language processing. There exist, for instance, well-known measures of simplicity or complexity of presentation, and many word-processing programs make suggestions about the organization of a text. These, non-topical characteristics could be fairly easily incorporated as indexing and retrieval devices. Some of the aspects from the Topic facet are also amenable to textual analysis, such as whether only a part of the user's topic is treated in the text, or whether the topic is a focus of the text, or peripheral to it. Other facets will clearly be more difficult to represent and use directly for IR. But, it seems that some of them could be straightforwardly used to support the user's interaction with the text (and thus, the decision-making process), once a text, or texts, have been retrieved. The use of acknowledgments or references to indicate point of view or authority is a good example of such possibilities. Many other of the individual characteristics associated with the facets we have identified seem amenable to this sort of use; what remains to be done is to develop realistic techniques for their representation, and, evaluation of their utility in online, interactive IR.

5. CONCLUSIONS

We have presented some preliminary and still tentative results, which indicate that there are a wide variety of factors associated with usefulness or relevance judgments, which go far beyond the traditional concept of topical relevance. We have also suggested how these various factors might be used by people in making their relevance decisions, and further, how some of these factors might be used to support better online information retrieval. Clearly, the suggestions made here must be considered as tentative. The data on which they are based are still not completely analyzed, and the generalizability of the data to other situations is still an open question. Nevertheless, these are promising enough to lead us to suggest a program of further research.

The next steps in our general research program will be to enumerate more precisely the individual aspects of the facets we have already identified as being important to the process; to attempt to implement some of representational strategies and retrieval and interaction techniques implied by these results in prototype systems; and to evaluate the effectiveness of the use of these characteristics in online, interactive information retrieval. We hope that next year we will be reporting to you on the results of these studies.

6. ACKNOWLEDGEMENTS

We are indebted to the students and scholars who took part in our study, and to Lynn Mullins of Rutgers University, who conducted many of the interviews with the humanities scholars, and contributed significantly to these results.

7. REFERENCES

1. Doyle, L.B. "Is relevance an adequate criterion for retrieval system evaluation?" In: H.P. Luhn, ed. *Automation and scientific communication, short papers, part 2*. Washington, D.C., American Documentation Institute, 1963: 199-200.
2. Cooper, W.S. "On selecting a measure of retrieval effectiveness, part 1." *Journal of the American Society for Information Science*, Vol 24, 87-100, 1973.
3. Su, L.T. "Evaluation measures for interactive information retrieval." *Information Processing and Management*, Vol. 28, 513, 1992.
4. Saracevic, T. "Relevance: A review of and a framework for the thinking on the notion in information science." *Journal of the American Society for Information Science*, Vol 26, 321-343, 1975.
5. SIG/FIS-RELEVANCE: A dialogue about fundamentals. Moderator T.J. Froelich. In: *Proceedings of the Annual Meeting of the American Society for Information Science*, Vol. 29, 330-332, 1992.

Appendix A

Instructions to Students

To Improve undergraduate education, we are participating in a study of how undergraduate engineering students choose the books and articles they use to prepare writing assignments.

We are therefore asking you to fill out one copy of the form that you're getting today for each book or article that you consider (even those you think you might not use) for this paper assignment. Make as many copies of this form as you need, and fill them out when you look at each book or article, not at some later time.

These forms should be all turned in together with your paper.

Research Questionnaire

1. Class user account:
2. Today's date and time:
3. Book or article reference:
First author:
Title:
Journal title:
Year, volume number, pages:
4. Do you think you'll use this book or article for your paper?
(circle one) Yes No Don't Know
5. What is it about this book or article that makes you think this way? Put down ALL the reasons that you can think of.
6. How far along are you in doing your assignment? (check one)
Picking a specific topic
Learning about the topic
Formulating your thoughts
Writing the paper
Something else (please specify)
7. Right now, how likely are you to get an 'A' on this assignment? (circle the appropriate number on the line below)
1-----2-----3-----4-----5
Not a chance Certain
8. How comfortable are you with reading and writing the English language? (circle the appropriate number on the line below)
1-----2-----3-----4-----5
It's a struggle Like a native