

# Relevance: The Whole History

**Stefano Mizzaro**

*Dipartimento di Matematica e Informatica, Università di Udine, Via delle Scienze, 206—Loc. Rizzi, 33100—Udine, Italy. E-mail: mizzaro@dimi.uniud.it*

**Relevance is a fundamental, though not completely understood, concept for documentation, information science, and information retrieval. This article presents the history of relevance through an exhaustive review of the literature. Such history being very complex (about 160 papers are discussed), it is not simple to describe it in a comprehensible way. Thus, first of all a framework for establishing a common ground is defined, and then the history itself is illustrated via the presentation in chronological order of the papers on relevance. The history is divided into three periods (“Before 1958,” “1959–1976,” and “1977–present”) and, inside each period, the papers on relevance are analyzed under seven different aspects (methodological foundations, different kinds of relevance, beyond-topical criteria adopted by users, modes for expression of the relevance judgment, dynamic nature of relevance, types of document representation, and agreement among different judges).**

## Introduction

Why has information science emerged on its own and not as a part of librarianship or documentation, which would be most logical? It has to do with relevance . . . to be effective, scientific communication . . . has to deal not with any old kind of information but with relevant information. (Saracevic, 1975, pp. 323–324)

Since information science first began to coalesce into a distinct discipline in the forties and early fifties, relevance has been identified as its fundamental and central concept . . . an enormous body of information science literature is based on work that *uses* relevance, without thoroughly understanding what it *means* . . . without an understanding of what relevance means *to* users, it seems difficult to imagine how a system can retrieve relevant information *for* users. (Schamber, Eisenberg, & Nilan, 1990, pp. 755–756)

. . . the topic of relevance, acknowledged as the most fundamental and much debated concern for information science. . . . Early on, information scientists recognized that the concept of relevance was integral to information

system design, development and evaluation. However, there was little agreement as to the exact nature of relevance and even less that it could be operationalized in systems or for the evaluation of systems. . . . this lack of agreement continues to an extent at the present. (Froehlich, 1994, p. 124)

This is an article on the history of *relevance* in the fields of documentation, information science and information retrieval. *Why* to write an article on the history of relevance? *How* to write it? The first question is answered by the following points:

- The above three citations witness that relevance is one of the central concepts, if not *the* central concept, for documentation, information science, and *information retrieval* (IR in the following): In the first citation, Saracevic maintains that relevance is the reason for the birth of information science and emphasizes its importance for the field of documentation; in the second one, Schamber, Eisenberg, and Nilan remark that relevance is the “fundamental and central” concept for information science; in the last one, Froehlich reminds us of the importance of relevance for building and evaluating *Information Retrieval Systems* (IRS).
- Relevance is not a well understood concept (as emphasized in the last two citations). Its history, if presented in an opportune way, is very useful for understanding what relevance is.
- There is no recent paper that describes in a complete way the history of relevance. Actually, some surveys exist (Saracevic, 1970a, 1970c, 1975, 1976; Schamber et al., 1990; Schamber, 1994), but the first four are now more than 20 years old, and the last two are not as complete, schematic, and methodical as Saracevic’s. Moreover, none of the surveys reviews in an exhaustive manner the literature on relevance, while in this article I try to take into account *all* the work done in the last 40 years including *any* paper that seems to me “relevant for the relevance topic.”
- This work can be situated at a higher level than the above mentioned surveys; it can be seen as a sort of index to, or annotated bibliography of, the relevance literature; and it can be used as a first step in approaching the study of relevance.

For the second question (how to write the article):

- Relevance is a widely studied concept in many fields: Philosophy, psychology, artificial intelligence, natural language understanding, and so on. In this article, I will not cross the frontiers of documentation, information science, and information retrieval.
- I will try to be as objective as possible throughout the article. An objective way of analyzing the history of a concept is to rely on all the published (or widely known) papers about that concept. Obviously, there are some problems with this approach: One piece of work may be described in many papers, or many pieces of work in one, or some pieces may not be described at all, and so on; I may miss some paper; I had to subjectively choose which papers are “relevant to relevance”; and I need to subjectively interpret the work done by other researchers in order to synthetically describe it. Anyway, I believe that it is a good, if not the best, objective approximation obtainable.
- Another problem to face is the complexity: A lot of “relevant to relevance” papers have been published (about 160, as we will see later), from different points of view and with interrelations among them. A simple list of them in chronological order would be completely incomprehensible. If one wants to present the whole history in an understandable way, a schematic style and some preliminary work for preparing a common ground are needed. Thus, automatically, the aim of this article becomes not only to present the history of relevance, but also to give a framework for understanding the history and the concept.

Summarizing, having read this article, the reader should: Know the history of relevance, know better what relevance is, and know how to proceed in studying relevance himself.

The article is structured as follows. First of all, the next section describes a framework that takes into account the existence of various kinds of relevance. This framework is needed for two reasons: To introduce the terminology used in the following, and to sketch a common ground for presenting the issues of the next sections. In the subsequent section, I introduce the three periods into which the history of relevance is divided (“Before 1958,” “1959–1976,” and “1977–present”) and the seven aspects (methodological foundations, different kinds of relevance, beyond-topical criteria adopted by users, mode of expression of the relevance judgment, dynamic nature of relevance, type of documents representation adopted, and agreement among different judges) under which the papers on relevance are analyzed. Then, each of the following three sections presents one period. Finally, the last two sections analyze the work done and conclude the article.

## A Framework for Various Kinds of Relevance

There are *many* kinds of relevance, not just *one*. This statement is justified in this section, where a framework

for classifying the various existing relevances is presented, and will be confirmed in the next ones, where the work of other authors will be analyzed. For the sake of brevity, the framework is described only in an intuitive manner: Its purpose is to allow to say without ambiguities which relevance we are talking about; see Mizzaro (1995, 1996b) for a more formal approach.

It is commonly accepted that relevance is a relation between two entities of two groups. In the first group, we have one of the following three entities: (i) *Document*, the physical entity that the user of an IRS will obtain after his seeking of information; (ii) *Surrogate*, a representation of a document. It may assume different forms and may be made up by one or more of the following: Title, list of keywords, author(s) name(s), bibliographic data (date and place of publication, publisher, pages, and so on), abstract, extract (sentences from the document), and so on; and (iii) *Information*, what the user receives when reading a document.<sup>1</sup>

In the second group, we have one of the following four entities: (i) *Problem*, that which a human being is facing and that requires information for being solved; (ii) *Information need*, a representation of the problem in the mind of the user. It differs from the problem because the user might not perceive in the correct way his problem<sup>2</sup>; (iii) *Request*, a representation of the information need of the user in a “human” language, usually in natural language; and (iv) *Query*, a representation of the information need in a “system” language, for instance Boolean.

Now, a relevance can be seen as a relation between two entities, one from each group: The relevance of a surrogate to a query, or the relevance of a document to a request, or the relevance of the information received by the user to the information need, and so on.

These are not all the possible relevances. The above mentioned entities can be decomposed in the following three components (Brajnik, Mizzaro, & Tasso, 1995, 1996; Mizzaro, 1995): (i) *Topic*, that which refers to the subject area to which the user is interested. For example, “the concept of relevance in information science”; (ii) *Task*, that which refers to the activity that the user will execute with the retrieved documents. For example, “to write a survey paper on . . .”; (iii) *Context*, that which includes everything not pertaining to topic and task, but however affecting the way the search takes place and the evaluation of results. For example, documents already known by the user (and thus not worth being retrieved), time and/or money available for the search, and so on.

<sup>1</sup> I know that the definition of “information” is hard work. Furthermore, probably information is not the same kind of entity as surrogate and document, and it should not be put together. Anyway, I am not interested here in such issues: I am just supposing that information exists. See Mizzaro (1996a) for a definition of information.

<sup>2</sup> Actually, it is possible to think of at least two kinds of information need: An implicit one and an explicit one; see, for instance, Taylor (1968). Here I assume that the information need is implicit in the mind of the user.

Usually, IR research concentrates on the topic component, but the user is not interested in obtaining information not useful for the task he has to execute, or already known documents. In other words, a surrogate (a document, some information) is relevant to a query (request, information need, problem) with respect to one or more of the components: It is possible to speak of “the relevance of a surrogate to a query for what concerns the topic (task, context) component,” or “the relevance of a document to a request for what concerns the topic and task,” or “the relevance of the information received by the user to the information need for all the three components,” and so on.

The scenario presented so far is static. But the information seeking situation takes place on a time interval: The user has a problem, perceives it, interacts with an IRS (and maybe a human intermediary), expresses his information need in a request, formalizes it into a query, examines the retrieved documents, reformulates the query, re-expresses his information need, perceives the problem in a different way, and so on. Also the time has to be taken into account: A surrogate (a document, some information) may be not relevant to a query (request, information need, problem) at a certain point of time, and be relevant later, or vice versa. This happens, for instance, if the user learns something that permits him to understand a document, or if the user problem changes, and so on.

Therefore, each relevance can be seen as a point in a four-dimensional space, the values of each of the four dimensions being: (i) Surrogate, document, information; (ii) query, request, information need, problem; (iii) topic, task, context, and each combination of them; and (iv) the various time instants from the arising of problem until its solution.

The situation is (partially) depicted in Figure 1. On the left hand side, there are the elements of the first dimension, and on the right hand side there are the elements of the second one. Each line linking two of these objects is a relevance (graphically emphasized by a circle on the line). The three components (third dimension) are represented by the three gray levels used. For simplifying the figure, the time dimension is not represented. Finally, the gray arrows among the relevances represent how much a relevance is near to the relevance of the information received to the problem for all three components, the one in which the user is interested, and how difficult it is to measure it. This analysis shows how it is short-sighted to speak merely of “system relevance” (the relevance as seen by an IRS) as opposed to “user relevance” (the relevance in which the user is interested), and how “topicality” (a relevance for what concerns the topic component) is conceptually different from “system relevance.”

Until now, I have illustrated different kinds of relevance. Now, let us come to relevance *judgment*. A relevance judgment is an assignment of a value of relevance (now, we know that it is more correct to say “a value of a relevance”) by a judge at a certain point of time. There-

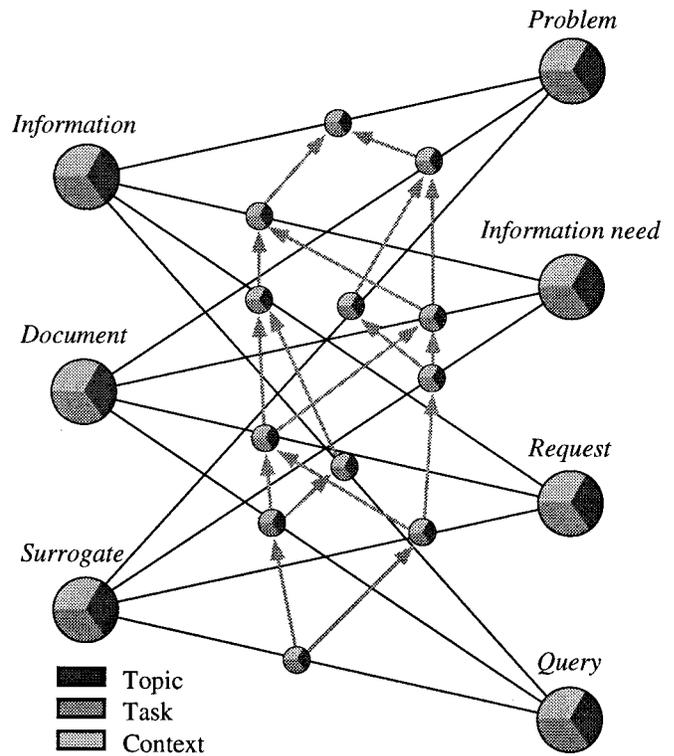


FIG. 1. The partial order of relevances.

fore, similarly to what has been done above, it is possible to say that there are many kinds of relevance judgment that can be classified along five dimensions: (i) The kind of relevance judged; (ii) the kind of judge (in the following I will distinguish between user and non-user); (iii) what the judge can use (surrogate, document, or information) for expressing his relevance judgment. It is the same dimension used for relevance, but it is needed, since, for instance, the judge can judge the relevance of a document on the basis of a surrogate; (iv) what the judge can use (query, request, information need, or problem) for expressing his relevance judgment. It is needed for the same reason as the previous point; (v) the time at which the judgment is expressed. It is needed because at a certain time point, it is obviously possible to judge the relevance in another time point.

In the following, I will refer to the above classifications in order to avoid ambiguities about which relevance or relevance judgment we are talking about.

### History of Relevance

Now let us come to the history of relevance. In Table 1, all the publications that I have found on this subject are presented, in chronological order. The first column contains the year, the second one the bibliographic citation, and the third one summarizes the type of the research, and can take one or more of the following values: “C” (Conceptual), indicates a paper discussing method-

TABLE 1. Papers on relevance—Part I.

| Year | Paper                           | Type | Year | Paper                      | Type |
|------|---------------------------------|------|------|----------------------------|------|
| 1959 | (Vickery, 1959a)                | C    |      | (Rees & Schulz, 1967)      | E    |
|      | (Vickery, 1959b)                | C    |      | (Shirey & Kurfeerst, 1967) | E    |
| 1960 | (Bar-Hillel, 1960)              | C    |      | (Weis & Katter, 1967)      | E    |
|      | (Maron & Kuhns, 1960)           | CTE  | 1968 | (Katter, 1968)             | E    |
| 1961 | (Rath, Resnick, & Savage, 1961) | E    |      | (Lesk & Salton, 1968)      | E    |
|      | (Resnick, 1961)                 | E    |      | (O'Connor, 1968)           | CE   |
| 1963 | (Doyle, 1963)                   | C    |      | (Paisley, 1968)            | C    |
|      | (Fairthorne, 1963)              | C    |      | (Wilson, 1968)             | C    |
|      | (Rees & Saracevic, 1963)        | C    | 1969 | (Gifford & Baumanis, 1969) | E    |
| 1964 | (Barhydt, 1964)                 | E    |      | (O'Connor, 1969)           | E    |
|      | (Goffman, 1964)                 | T    |      | (Saracevic, 1969)          | E    |
|      | (Hillman, 1964)                 | EC   | 1970 | (Foskett, 1970)            | C    |
|      | (Resnick & Savage, 1964)        | E    |      | (Goffman, 1970)            | TC   |
| 1965 | (Hoffman, 1965)                 | E    |      | (Saracevic, 1970a)         | SC   |
|      | (Taube, 1965)                   | C    |      | (Saracevic, 1970b)         | EC   |
| 1966 | (Goffman & Newill, 1966)        | C    |      | (Saracevic, 1970c)         | S    |
|      | (Rees, 1966)                    | E    | 1971 | (Cooper, 1971)             | T    |
|      | (Rees & Saracevic, 1966)        | C    |      | (Foskett, 1972)            | C    |
| 1967 | (Barhydt, 1967)                 | E    | 1973 | (Belzer, 1973)             | EC   |
|      | (Cuadra & Katter, 1967a)        | E    |      | (Cooper, 1973a)            | C    |
|      | (Cuadra & Katter, 1967b)        | E    |      | (Cooper, 1973b)            | C    |
|      | (Cuadra & Katter, 1967c)        | E    |      | (Thompson, 1973)           | E    |
|      | (Dym, 1967)                     | E    |      | (Wilson, 1973)             | C    |
|      | (Goffman & Newill, 1967)        | TC   | 1974 | (Kemp, 1974)               | C    |
|      | (Hagerty, 1967)                 | E    |      | (Kochen, 1974)             | T    |
|      | (Katter, 1967)                  | E    | 1975 | (Saracevic, 1975)          | S    |
|      | (O'Connor, 1967)                | C    | 1976 | (Saracevic, 1976)          | S    |

ological aspects; "E" (Experimental), indicates a work describing an experiment; "S" (Survey), labels a paper that reviews previous work; and "T" (Theoretical), indicates a theoretical or mathematical paper.

For the sake of illustration, I have divided the history of relevance into three conventional periods: "Before 1958," "1959–1976," and "1977–present." Each of the next three sections is devoted to the presentation of one period. Furthermore, the research on relevance can be divided into subtopics; thus (with the exception of the brief section on the "Before 1958" period) the sections are divided into seven subsections, each one presenting one particular aspect of the research about relevance. Within each subsection, the various works (a work may be described in more than one paper) on relevance are presented in chronological order (and in alphabetical order if they are published in the same year). The subsections are the same for both the "1959–1976" and "1977–present" periods, and are listed below, together with a brief description of the particular aspect faced.

### Foundations

Relevance can be defined from different standpoints, using different mathematical instruments and conceptual approaches. A line of research is devoted to such foundational issues.

### Kinds

As seen in the section on the framework, there exist many kinds of relevance, and each one presents its strengths and weaknesses. This is obviously a very substantial point: It is important to know which relevance we are talking about.

### Surrogates

The type of surrogate used can affect relevance judgments (and, as seen above, also the relevance itself). As most of today's IRSs are not full-text, it is important to understand this aspect. The *quality* of a surrogate is a measure of how much the relevance judgment expressed on the basis of a surrogate is similar to the relevance judgment expressed on the basis of the whole document.

### Criteria

Relevance and topicality are different, as seen above. A line of research is devoted to elicit (from experts or users) which criteria beyond the topical one are adopted by the users when expressing their relevance judgments.

### Dynamics

Relevance is a dynamic phenomena: For the same judge, a document may be relevant at a certain point of

TABLE 1. Papers on relevance—Part II.

| Year | Paper                                | Type | Year | Paper   | Type |
|------|--------------------------------------|------|------|---|------|
| 1977 | (Bookstein, 1977)                    | T    |      | (Foster, 1986)                                  | S    |
|      | (Davidson, 1977)                     | E    |      | (Meadow, 1986)                                  | C    |
|      | (Maron, 1977)                        | C    |      | (Swanson, 1986)                                 | C    |
|      | (Robertson, 1977)                    | SC   |      | (Taylor, 1986)                                  | C    |
|      | (Swanson, 1977)                      | C    |      | (van Rijsbergen, 1986a)                         | T    |
|      | (Tessier, Crouch, & Atherton, 1977)  | C    |      | (van Rijsbergen, 1986b)                         | T    |
| 1978 | (Cooper & Maron, 1978)               | T    | 1987 | (Eisenberg & Hu, 1987)                          | E    |
|      | (Figueiredo, 1978)                   | E    |      | (Rorvig, 1987)                                  | E    |
|      | (Marcus, Kugel, & Benenfeld, 1978)   | E    | 1988 | (Eisenberg, 1988)                               | E    |
|      | (Wilson, 1978)                       | C    |      | (Eisenberg & Barry, 1988)                       | E    |
| 1979 | (Bookstein, 1979)                    | CT   |      | (Halpern & Nilan, 1988)                         | EC   |
|      | (Kazhdan, 1979)                      | E    |      | (Nie, 1988)                                     | T    |
|      | (Koll, 1979)                         | E    |      | (Nilan, Peek, & Snyder, 1988)                   | EC   |
|      | (Lancaster, 1979)                    | CS   |      | (Regazzi, 1988)                                 | E    |
| 1980 | (Brookes, 1980)                      | C    |      | (Rorvig, 1988)                                  | S    |
| 1981 | (Koll, 1981)                         | E    |      | (Saracevic & Kantor, 1988a)                     | E    |
|      | (Tessier, 1981)                      | CE   |      | (Saracevic & Kantor, 1988b)                     | E    |
| 1982 | (Boyce, 1982)                        | C    |      | (Saracevic, Kantor, Chamis, & Trivison, 1988)   | E    |
| 1983 | (Bookstein, 1983)                    | T    |      | (Swanson, 1988)                                 | C    |
| 1984 | (Ellis, 1984)                        | C    |      | (Tamiyu & Ajiferuke, 1988)                      | T    |
| 1985 | (Meadow, 1985)                       | C    | 1989 | (Janes, 1989)                                   | T    |
|      | (Rorvig, 1985)                       | E    |      | (Nie, 1989)                                     | T    |
| 1986 | (Eisenberg, 1986)                    | E    |      | (van Rijsbergen, 1989)                          | T    |
|      | (Eisenberg & Barry, 1986)            | E    |      |   |      |
| 1990 | (Katzner & Snyder, 1990)             | C    |      | (Meghini, Sebastiani, Straccia, & Thanos, 1993) | T    |
|      | (O'Brien, 1990)                      | C    |      | (Park, 1993)                                    | E    |
|      | (Purgailis, Parker, & Johnson, 1990) | E    |      | (Thomas, 1993)                                  | E    |
|      | (Rorvig, 1990)                       | E    |      | (Wilson, 1993)                                  | C    |
|      | (Sandore, 1990)                      | E    | 1994 | (Barry, 1994)                                   | E    |
|      | (Schamber, Eisenberg, & Nilan, 1990) | SC   |      | (Bruce, 1994)                                   | EC   |
| 1991 | (Bruza & van der Weide, 1991)        | T    |      | (Froehlich, 1994)                               | SC   |
|      | (Froehlich, 1991)                    | C    |      | (Hersh, 1994)                                   | C    |
|      | (Gordon & Lenk, 1991)                | T    |      | (Howard, 1994)                                  | E    |
|      | (Janes, 1991a)                       | E    |      | (Janes, 1994)                                   | E    |
|      | (Janes, 1991b)                       | E    |      | (Ottaviani, 1994)                               | CT   |
|      | (Schamber, 1991a)                    | E    |      | (Park, 1994)                                    | C    |
|      | (Schamber, 1991b)                    | E    |      | (Schamber, 1994)                                | S    |
|      | (Su, 1991)                           | E    |      | (Sebastiani, 1994)                              | T    |
| 1992 | (Bruza & van der Weide, 1992)        | T    |      | (Smithson, 1994)                                | E    |
|      | (Burgin, 1992)                       | E    |      | (Soergel, 1994)                                 | C    |
|      | (Froehlich & Eisenberg, 1992)        | CS   |      | (Su, 1994)                                      | E    |
|      | (Harter, 1992)                       | C    |      | (Sutton, 1994)                                  | C    |
|      | (Janes & McKinney, 1992)             | E    |      | (Wang, 1994)                                    | E    |
|      | (Lalmas & van Rijsbergen, 1992)      | T    | 1995 | (Brajnik, Mizzaro, & Tasso, 1995)               | E    |
|      | (Nie, 1992)                          | T    |      | (Crestani & van Rijsbergen, 1995a)              | T    |
|      | (Park, 1992)                         | E    |      | (Crestani & van Rijsbergen, 1995b)              | T    |
|      | (Su, 1992)                           | E    |      | (Nie, Brisebois, & Lepage, 1995)                | T    |
| 1993 | (Barry, 1993)                        | E    | 1996 | (Brajnik, Mizzaro, & Tasso, 1996)               | E    |
|      | (Bruza, 1993)                        | T    |      | (Ellis, 1996)                                   | C    |
|      | (Cool, Belkin, & Kantor, 1993)       | E    |      | (Harter, 1996)                                  | CS   |
|      | (Janes, 1993)                        | E    |      | (Lalmas, 1996)                                  | T    |
|      | (Lalmas & van Rijsbergen, 1993)      | T    |      | (Lalmas & van Rijsbergen, 1996)                 | T    |

time and not relevant later, or vice versa. The dependencies among documents are particularly studied: The first seen documents can affect the relevance of the next ones.

### Expression

Many kinds of human judgments are intrinsically inconsistent, and this is true also for relevance judgment.

Hence, the problem is relevance judgment *expression*: Which is the best way for the judges to express in a consistent manner their relevance judgment? Many alternatives have been proposed and used: The standard dichotomous (yes/no) relevance judgments; the *category rating scales*, in which the relevance judgment is expressed using a value taken from a finite scale containing typically 3–11 elements; and *magnitude estimation*, in

which every positive rational number can be used. For the magnitude estimation, an important parameter is how the judgment is physically expressed; common methods are: Numeric estimation (higher numbers indicate higher relevance), line-length (the longer the line drawn by the judge, the higher the relevance of a document), and force hand grip (the higher the strength measured by a dynamometer, the higher the relevance of a document).

### *Subjectiveness*

Relevance is subjective: Different judges may express different relevance judgments. Thus, it is important to understand if, when, and how much: (i) The relevance judgments expressed by different judges (or groups of judges) are consistent, and (ii) user's relevance judgments agree with non-user's judgments (judgments expressed by a person different from the user).

A further subsection ("The end of the period") terminates the sections. Obviously, a paper treating more than one aspect appears more than once, in different sections. Equally obviously, it is impossible to describe here all the aspects of each work: Only a brief synthesis is given.

### **Before 1958**

The history of relevance might start centuries ago with the first libraries: The first library users are already concerned with the problem of finding *relevant* information. This period is featured by implicitness: The notion of relevance lies behind a lot of studies, but it never comes to the surface, and almost nobody speaks explicitly of this subject. The main events are:

- In the 17th Century, with the publication of the first scientific journals, the communication mechanism that modern science still adopts nowadays arises, and the notion becomes more central, though never mentioned.
- In our century, a lot of studies, by Lotka (1926), Bradford (1934), Zipf (1949), Urquhart (1959), Price (1965) on what will be called, years later, *bibliometrics* (Pritchard, 1969) are seen by Saracevic (1975) as the first formal basis of relevance.
- In the 1930s and 1940s, according to Saracevic (1975), S. C. Bradford is the first one to talk about articles *relevant to a subject*.
- In the 1950s, the IR pioneers Mooers (1950), Perry (1951), Taube (1955), and Gull (1956) build the first IRSs, and note that not all the items retrieved are relevant.

It is clear that the notion of relevance is "somewhere out there," behind scientific literature search, bibliometric studies, IRSs, and so on. But it is not explicitly recognized; it is hidden, implicit. This period ends in 1958, with the International Conference for Scientific Information (ICSI) in which the concept is explicitly recognized.

### **1959–1976**

Vickery's presentations at the 1958 ICSI debate (Vickery, 1959a, 1959b) are widely recognized as a landmark in relevance history, and give rise to a consistent amount of studies in the period "1959–1976." This period is well documented in some surveys by Saracevic (1970a, 1970c, 1975, 1976) and also in Schamber et al. (1990). Hence, I do not go into the details of the work of this period.

As said above, in order to improve the comprehensibility, this section is divided into subsections, each one presenting a particular aspect of the research about relevance. Each subsection is closed by a brief summary of the corresponding line of research.

### *Foundations*

The papers exploring the foundations of relevance:

**Maron and Kuhns** (1960) call for the adoption of probability in the definition of relevance, and claim that relevance is not a yes/no decision.

**Doyle** (1963) states that relevance is too elusive for being a reliable criterion for IRSs evaluation.

**Hillman** (1964) starts from the definitions of "concept," "concept formation," and "conceptual relatedness" for defining relevance. An experiment shows that the formal definition of a concept cannot be exploited on the basis of human similarity-judgments of documents, so an alternative approach is sketched.

**Rees** (1966) notes that the definition of relevance should rely on concepts as the information conveyed by a document, the "previous knowledge" of the user, and the "usefulness" of the information to the user.

**Goffman and Newill** (1967) (Goffman, 1970) compare the spreading of ideas with the spreading of disease, and treat relevance as a measure of the "effectiveness of the contact." They mathematically prove that relevance is an equivalence relation (because more than one answer to a request is possible), and that the database is partitioned in equivalence classes.

**Wilson** (1968) notes that a topical document may not be judged interesting by the user if, for example, he already knows the document, or its contents.

**Saracevic** (1970a, 1970b, 1975, 1976) synthesizes some statistical distributions studied in bibliometrics and presenting the common feature that in a set, a small subset of elements appear more often, while the largest part of the elements appear only seldom. This is true when substituting "elements appear" with: "Documents are retrieved," "words appear in a document," "authors and bibliographic citations appear in the literature," and so on. Saracevic suggests that relevance is the concept underlying such phenomena.

**Cooper** (1971) defines relevance on the basis of notions borrowed from mathematical logic, namely entailment and minimality. First of all, Cooper defines

that a sentence  $s$  is relevant to another sentence  $r$  (or to its logical negation  $\neg r$ ) if  $s$  belongs to a minimal set of premises  $M$  entailing  $r$ . In symbols: *Relevant* ( $s, r$ ) iff  $\exists M(s \in M \wedge M \models r \wedge M - s \not\models r)$ . Then, a document  $D$  is seen as a set of sentences  $D = \{s_1, s_2, \dots, s_n\}$ , and its relevance to a request  $r$  is defined as: *Relevant*( $D, r$ ) iff  $\exists i$  (*relevant*( $s_i, r$ )).

**Wilson** (1973) tries to improve Cooper's definition, and uses the term *situational relevance*. He introduces the "situation," the "stock of information," and the goals of the user, and claims that probability and inductive logic, in addition to the deductive one used by Cooper, have to be used in defining relevance.

**Kochen** (1974) defines a mathematical function that assigns a utility value to each document, given a user and a request. He also notes the limitations of his definition that does not take into account the "situation," that may affect the preferences of the user.

The definitional issues are afforded through various mathematical instruments, and the basis for the future work are established: Maron and Kuhns (1960) for probabilistic retrieval; Cooper (1971) and Wilson (1973) for the use of mathematical logic; and Rees (1966) and Wilson (1968) for the importance of the user's stock of knowledge for the relevance of a document.

### *Kinds*

The papers about the existence of various kinds of relevance:

**Vickery** (1959a, 1959b) presents at the ICSI debate a distinction between "relevance to a subject" (the relevance of a document to a query for what concerns the topical component) and "user relevance" (that refers to what the user needs).

**Bar-Hillel** (1960) questions topicality (intended as the relevance with respect to the topic component), maintaining that the distance between documents (or topics) cannot be measured.

**Maron and Kuhns** (1960) note that the relevance of a document to a request is different from the relevance of a document to an information need, although the two are supposedly related (and such hypothesis is experimentally verified).

**Fairthorne** (1963) maintains that relevance has to be measured only on the basis of the words in the document and in the request (the relevance of a document to a request). If the individuality of the user is taken into account, then any text is relevant to any request from some point of view.

**Taube** (1965) criticizes the notion of relevance adopted in the Cranfield Studies, the relevance of a surrogate to a request.

**Goffman and Newill** (1966) distinguish between "relevance" (intended as the relevance of a document

to a request) and "pertinence" (the relevance of a document to an information need).

**Rees and Saracevic** (1966) remark that the relevance to a request is different from the relevance to an information need.

**Rees and Schultz** (Rees, 1966; Rees & Schulz, 1967) distinguish between "relevance" (the relevance of information to the information need) and "usefulness," that comprises individual characteristics of the judge.

**O'Connor** (1968), discussing the expression "satisfying a requester's information need," implicitly speaks of three kinds of relevance, namely the relevance of (1) a surrogate to the query, (2) information to the information need, and (3) information to the problem.

**Paisley** (1968) distinguishes "perceived relevance" and "perceived utility" which includes things like how easy it is to obtain and read the document.

**Foskett** (1970, 1972) distinguishes between relevance to a request, that he calls "relevance," and relevance to an information need, that he calls "pertinence." The former is seen as a "public," "social" notion, that has to be established by a general consensus in the field, the latter as a "private" notion, depending solely on the user and his information need.

**Cooper** (1973a, 1973b) distinguishes between his "logical relevance," or topicality (relevance for what concerns the topical component), and "utility" (relevance for all three components). He argues that it is the second one that must be used in evaluating an IRS.

**Wilson** (1973) explicitly distinguishes the relevance of information to an information need (his "situational relevance") and the relevance of information to a problem.

**Kemp** (1974) continues Foskett's work, remarking that relevance is objective, while pertinence is not.

The existence of many kinds of relevance is early recognized, though often in a short-sighted way if contrasted with the scenario presented in the section on the framework. Many authors simply note it (Vickery, 1959a, 1959b; Maron & Kuhns, 1960; Goffman & Newill, 1966; Rees, 1966; Rees & Saracevic, 1966; Rees & Schulz, 1967; O'Connor, 1968; Paisley, 1968; Foskett, 1970, 1972; Wilson, 1973); others maintain the inadequacy of some kind of relevance (Bar-Hillel, 1960; Taube, 1965); and others claim that one kind of relevance is better than another (Fairthorne, 1963; Cooper, 1973a, 1973b; Kemp, 1974).

### *Surrogates*

The works aimed at understanding how various forms of surrogate affect relevance judgment:

**Rath, Resnick, and Savage** (1961) (Resnick, 1961; Resnick & Savage, 1964) explore, by an experimental study, the differences among the relevance judg-

ments expressed on the basis of various kinds of surrogates (title, abstract, keywords, bibliographic citation, extract, bibliographic citation plus abstract, and bibliographic citation plus keywords) and the whole document. They find no significant differences.

**Hagerty** (1967) compares the relevance judgments expressed on the basis of: (1) A document, (2) a surrogate made up by the title, and (3) various surrogates made up by abstracts of different lengths (30, 60, and 300 words). She finds that the quality of the surrogate (defined in the previous section) increases with its length.

**Katter** (1967) compares the relevance judgments on the basis of a surrogate made up by keywords with the relevance judgments on the basis of the whole document.

**Weis and Katter** (1967) compare the relevances of various kinds of surrogates (abstract, keywords, extract, title) and find that abstracts and extracts have a higher quality than titles.

**Saracevic** (1969) compares the relevance judgments on the basis of two kinds of surrogate (title and abstract) and of the whole document, finding significant differences.

**Belzer** (1973), on the basis of Shannon and Weaver's *information theory*, shows how the *entropy* of various types of surrogates (abstracts, first paragraphs, and last paragraphs) can be used as a prediction of the quality of a surrogate.

**Thompson** (1973) studies how the presence or absence of the abstract affects the correspondence of a quick preliminary relevance judgment with the final relevance judgment, and the time needed to express it. He finds no difference.

Some of these studies (for instance, Hagerty, 1967; Weis & Katter, 1967) suggest that the quality of a surrogate increases with its length, while others (for instance, Rath et al., 1961; Resnick, 1961; Resnick & Savage, 1964; Thompson, 1973) maintain that there is no significant difference. Anyway, all of them agree that increasing the length of the surrogate does not make its quality worse.

### Criteria

Only a few studies analyze the beyond-topical criteria adopted by users in judging the relevance:

**Rees and Saracevic** (1963) hypothesize on the variables and conditions under which the relevance judgment would achieve a high degree of agreement.

**Cuadra and Katter** (1967a, 1967b, 1967c) find 38 variables (for instance style, specificity, and level of difficulty of documents) that affect the relevance judgment.

**Rees and Schultz** (1967) note that relevance judgments are inconsistent and affected by about 40 variables.

They also find that as more information is given to the judge, relevance judgments become more stringent.

**Cooper** (1971, 1973a, 1973b) suggests that "utility" depends on many non-topical factors, among which are: Accuracy, credibility, source of publication, recency, authorship, and so on.

The studies by Cuadra and Katter and by Rees and Schultz show how relevance judgment depends on many beyond-topical variables: (1) The kind of document representation, (2) the way the request is expressed, (3) features of the judge like his knowledge of the subject, (4) the mode for expressing the judgment, and (5) the situation/context in which the judgment is expressed. But this line of research will have a huge expansion in the next period, as we will see below.

### Dynamics

The works that study the dynamic nature of relevance judgment:

**Goffman** (1964) proves, using the mathematical theory of *measures*, that relevance is not a relation only between the request and each single document: For relevance being a measure, the relations among documents must be taken into account.

**Rees and Saracevic** (1966) claim that the relevance judgment (for a single user) depends on time.

**Kochen** (1974) notes that the presentation order of documents can affect the preferences of the user.

The studies regarding the dynamic nature of relevance are very few in this period. It is anyway noted that a relevance judgment may depend on time and on the presentation order of the documents.

### Expression

The studies that explore the issue of relevance judgment expression:

**Cuadra and Katter** (1967a, 1967b) show that human relevance judgments are affected by a number of surrounding conditions, thus questioning the reliability of human relevance judgment. The authors also find that the judges, when using category rating scales, prefer to have a high number of categories among which to choose.

**Rees and Schultz** (1967) study the effect of different scaling techniques on the reliability of judgments.

**Weis and Katter** (1967) use a nine-points category rating scale for measuring the correspondence of relevance judgments expressed on the basis of different document representations.

**Katter** (1968) compares rating methods with ranking methods, and category-assignment methods with magnitude-estimation methods, finding no reliable method.

These studies do not establish "the most" reliable

method for expressing relevance judgment. They are any-way important, as they stand as the basis of the studies of the next period, and reveal some phenomena like: Different kinds of methods may produce different judgments; when using category rating scales, the judge prefers a high number of categories and uses mainly the end points of the scales; and so on.

### *Subjectiveness*

The studies that analyze the variation of relevance judgments among different judges:

**Barhydt** (1964, 1967) introduces the following measures for the similarity between users' and non-users' relevance judgments: *Sensitivity* (among the documents judged relevant by the user, the percentage judged relevant also by the non-user) and *specificity* (among the documents judged non-relevant by the user, the percentage judged nonrelevant also by the non-user). *Effectiveness* is the synthesis of these two measures into a single one:

$$\text{effectiveness} = \text{sensitivity} + \text{specificity} - 1.$$

Moreover, the author compares the (dichotomous) relevance judgments by subject experts and IRS experts, finding a 0.35 average effectiveness.

**Hoffman** (1965) studies the consistency of relevance judgments among different groups of judges and among judges of the same group.

**Rees and Saracevic** (1966) note that relevance judgment is subjective and not inherent to a document, and conclude that the relevance of a document for a user can be judged only by himself.

**Rees and Schultz** (Rees, 1966; Rees & Schulz, 1967) find about 40 variables affecting relevance judgment, among which are "features of the judge" and "quantity of information available" (more scientifically oriented judges and more information cause lower relevance ratings).

**Cuadra and Katter** (1967a, 1967b, 1967c) study 38 variables affecting relevance judgment, grouped into five classes comprising judge, judgment situation, and mode of expression of judgment.

**O'Connor** (1967) studies the effects of unclear requests on the relevance judgment: He suggests that if the request is unclear, then different judges will interpret it differently, and hence the agreement among them will be low. Different type of unclear requests are studied, and some suggestions for formulating clear requests are given.

**Goffman and Newill** (1967) (Goffman, 1970) mathematically prove, comparing the spreading of ideas with the spreading of disease, that relevance depends on what the judge already knows.

**Lesk and Salton** (1968) find a 30% agreement between

users' and non-users' relevance judgment. They define a *strong hypothesis* (differences in relevance judgments cannot affect the assessment of retrieval performance) and a *weak hypothesis* (differences in relevance judgments cannot affect the comparison of performances of different retrieval methods). Both the hypotheses are supported by experimental data. Moreover, the authors note that relevance judgments are more stringent as the subject knowledge of the judges increases.

**Gifford and Baumanis** (1969) show that the agreement of relevance judgments can be explained on the basis of co-occurrence of terms in the abstracts.

**O'Connor** (1969) suggests, on the basis of experimental evidence, that a discussion among judges changes relevance judgments and can resolve disagreements.

The problem of subjectiveness is noted, but no solution is proposed in this period. Anyway, some useful concepts are early established: Effectiveness, sensitivity, and specificity (Barhydt, 1964, 1967); consistency between and within groups of judges (Hoffman, 1965); weak and strong hypothesis (Lesk & Salton, 1968); and discussion among judges for resolving disagreements (O'Connor, 1969) (but see Gull, 1956; Harter, 1996).

### *The End of the Period*

This period is closed, in the middle of the 1970s, by some surveys by Saracevic. They summarize the work done and stand as a basis for the research of the following years:

**Saracevic** (1970a, 1970c, 1975, 1976) reviews the papers on relevance published in the "1959–1976" period, and proposes a framework for classifying the various notions of relevance proposed until then.

The studies by Cuadra and Katter (1967a, 1967b, 1967c) and by Rees and Schultz (1967) are surely the most important of this period. They appear in more than one of the above subsections, and (together Saracevic's surveys) will be the most cited in the papers of the next period.

### **1977–Present**

The last period of relevance history begins in 1977 (just after the above described surveys by Saracevic) and continues until today. This section describes this period and, as said, is divided into the same subsections as the previous one.

#### *Foundations*

A lot of papers continue to discuss the foundational issues:

**Maron** (1977) discusses *aboutness*, a central concept of indexing. He uses the expression "subjective about"

and maintains that aboutness is very complex, not understood, subjective, and not measurable.

- Robertson** (1977) synthesizes some work on probabilistic interpretations of relevance.
- Cooper and Maron** (1978) note, from the standpoint of utility-theoretic indexing theory, that relevance is a matter of degrees, not a dichotomous decision.
- Tessier** (1981) proposes a “summary measure of relevance” that evaluates the whole process of information seeking as an average of satisfaction scores.
- Ellis** (1984) questions the use of relevance as a criterion for assessing IRS performance. He speaks of a “paradox of relevance,” that can be summarized by: The more one uses the “real” relevance, the less one can measure it.
- van Rijsbergen** (1986a, 1986b, 1989) again brings to attention the use of mathematical logic for modeling relevance (and IR in general), after more than 10 years (since Cooper, 1971; Wilson, 1973): If the document and the query are represented by the logical formulae  $d$  and  $q$ , respectively, then the document is relevant to the query if the logical implication  $d \rightarrow q$  is true.
- Nie** (1988, 1989, 1992) relies on *modal logic* and Kripke’s *possible world semantics* (Chellas, 1980) for modeling the relevance of a document (represented by a possible world) to a query (represented by a formula) using the accessibility relationship among possible worlds. This approach allows one to model thesaural information and query expansion.
- Janes** (1989) suggests that the relevance judgment process can be seen, borrowing concepts from *search theory*, as a *detection* process: The more time and information a searcher has, the more likely he is to make a correct decision (judgment).
- Schamber, Eisenberg, and Nilan** (1990) join the user-oriented (as opposed to the system-oriented) view of relevance. They maintain that relevance is a multi-dimensional, cognitive, and dynamic concept, and feel that it is both systematic and measurable.
- Bruza and van der Weide** (1991, 1992) (Bruza, 1993) propose a logical approach in which the derivability relation is weakened. Two types of derivations are defined: “Strict” (represented by  $\vdash$ ) and “plausible” (represented by  $\vdash\sim$ ); each of them has its own axioms, and a document  $d$  is relevant to a query  $q$  if at least  $d \vdash\sim q$  holds.
- Gordon and Lenk** (1991) challenge from a mathematical standpoint, using (i) signal detection plus decision theory, and (ii) utility theory, the *probability ranking principle* in IR. Usually, a (probabilistic) IRS assigns, on the basis of a user’s query, a “retrieval status value” (a prediction of how much the document will be relevant for the user) to each document in a collection. Then, the  $n$  documents with the highest values are presented to the user. This policy is questioned, and is advisable (and optimal) only if

three conditions are met: (1) The retrieval status values are indeed the *probabilities* that the documents are relevant; (2) such values are reported without uncertainty (they are numbers, not intervals); and (3) the user’s judgments of relevance are mutually independent. If these conditions do not hold, then an alternative strategy might reduce the “risk” of presenting to the user non-relevant documents.

- Su** (1991, 1992, 1994) considers 20 measures (divided into 4 groups: Relevance, efficiency, utility, user satisfaction) for the evaluation of an IRS, and creates a fifth measure (success, representing the overall success of the search as judged by the user). She finds in an operational environment that seven (of 20) measures are correlated with success, while precision (a relevance-based measure) is not.
- Harter** (1992) applies the theory of *psychological relevance*, proposed by Sperber and Wilson, to the concept of relevance in information science. He obtains an elegant framework and draws some very interesting conclusions for IR and bibliometrics.
- Lalmas and van Rijsbergen** (1992, 1993, 1996) (Lalmas, 1996) use *situation theory* (Devlin, 1991) for modeling relevance in a similar way to Nie (1988, 1989, 1992): A document is a situation  $s$ , a query is a type  $\varphi$ , and the document is relevant to the query if there exists a flow of information arising from the situation  $s$  and leading to a situation  $s'$  such that  $s' \models \varphi$ .
- Meghini, Sebastiani, Straccia, and Thanos** (1993) (Sebastiani, 1994) use terminological logic: A document is an individual, a concept is a class of documents, a query is a concept, relations among concepts and documents are modeled by axioms, and the relevance is modeled by the “instance assertion” operator (a document is relevant to a concept if the document is an instance of the concept). Probability is added (Sebastiani, 1994) in this model in order to take into account the probability of relevance.
- Wilson** (1993) discusses the issue of efficiency in scientific communication. He starts from the assumption that for having efficiency it is necessary that *relevant* information, and not merely information, is communicated. The question of whether scientific communication is efficient or not is deemed a fundamental and unanswered one.
- Park** (1994) calls for the adoption of a *naturalistic paradigm of inquiry* (as opposed to the *rationalistic* one) in studying relevance. She claims that the focus must be on users, not on systems, and that, in order to make it possible to understand user’s information behavior, a qualitative approach seems unavoidable.
- Crestani and van Rijsbergen** (1995a, 1995b) use *logical imaging* (Harper, Stalnaker, & Pearce, 1981) (in which the evaluation of a conditional takes place using the “nearest” possible worlds). Possible worlds model terms, formulae model documents and

queries, and degrees of relevance are calculated on the basis of semantic relations among terms.

**Nie, Brisebois, and Lepage** (1995) use *logical imaging* (Harper et al., 1981) to model the relevance of a document to a query. This approach takes into account a model of the user (knowledge, background, and so on): Each possible world models a possible knowledge state of the user, a document is a formula, and it is relevant if it is compatible with the state of knowledge associated to the world.

Observing this line of research, it is easy to note the growing presence (especially near the end of the “1977–present” period) from the one side, of user-oriented, cognitive approaches (Schamber et al., 1990; Harter, 1992; Park, 1994; Su, 1991, 1992, 1994), and, from the other side, of attempts to define a logic for IR: The seminal work by van Rijsbergen (1986a, 1986b, 1989) gives rise to a considerable amount of studies (Nie, 1988, 1989, 1992; Bruza & van der Weide, 1991, 1992; Bruza, 1993; Lalmas & van Rijsbergen, 1992, 1993, 1996; Lalmas, 1996; Meghini et al., 1993; Sebastiani, 1994; Crestani & van Rijsbergen, 1995a, 1995b; Nie et al., 1995) that in the following years continue to propose more refined and complex modelizations. Moreover, while in the “1959–1976” period there seem to be two extreme positions, synthesized by the “paradox of relevance” by Ellis (1984), in the last period the cognitive approaches try to tackle the “subjective, not measurable” relevance in a more optimistic way (Schamber et al., 1990). Finally, the first studies that consider the relevance of a set of documents instead of a single document appear (Gordon & Lenk, 1991).

### *Kinds*

The analysis of various kinds of relevance continues:

**Bookstein** (1977, 1979) distinguishes between the “relevance” of a document, assigned by the user, and the “prediction” of the relevance of a document, assigned by the IRS and called “Retrieval Status Value” (RSV). He notes that RSV and relevance might not coincide, and so the documents with the highest RSVs might not be the most relevant ones.

**Swanson** (1977, 1986) defines two “frames of reference” for relevance. Frame of reference one sees relevance as a relation between “the item retrieved” and the user’s *need*; frame of reference two is based on the user’s *query*. In frame of reference two, relevance is identified with topicality (the relevance of a document to a query for what concerns the component topic), and retained more objective, observable, and measurable. In frame of reference one, topicality is not enough for assuring relevance, a more subjective and elusive notion (the relevance of a document to an information need for what concerns all three components).

**Lancaster** (1979) defines “pertinence” as the relation

between a document and a request as judged by the user, and “relevance” as the same relation, but judged by an external judge.

**Boyce** (1982) presents an analysis of relevance in which it is divided into “topicality” and “informativeness.” On the basis of such analysis, he proposes that the retrieval process should take place in two stages, a first one devoted to retrieve all the topical documents, and a second one for individuating the informative documents (i.e., the documents that give information to the user) among the topical ones.

**Regazzi** (1988) defines for his experimental study “relevance,” intended as topicality, or the relevance of a surrogate to a request for what concerns the topical component, and “utility,” intended as the relevance of a surrogate to a request for what concerns all three components. He finds (in an artificial setting) no significative difference between judgment of “relevance” and “utility.”

**Saracevic, Kantor, Chamis, and Trivison** (1988) (Saracevic & Kantor, 1988a, 1988b) present a comprehensive study of the information seeking behavior, analyzing many factors (grouped in 14 categories) affecting the evaluation of a search. Among the various measures used in the experiment, the authors define “relevance,” intended as the relevance of a surrogate to a request for what concerns the topical component, and “utility,” intended as the global usefulness for the user of the results of the search, and they find a good correlation between these two measures.

**O’Brien** (1990) discusses the importance of relevance in the evaluation of Online Public Access Catalogues (OPACs). She points out that relevance is a central feature, but the major uncertainty and dynamicity of the OPACs’ world (no intermediary, heterogeneous users, no time restrictions, relevance as a part of a complex information seeking situation, and so on) with respect to IRSs’ world, makes “subjective relevance” (the relevance of information to information need for all three components) harder to obtain and measure, and “objective relevance” (relevance of a document to a query on the topic component) less important.

**Sandore** (1990) suggests that the relevance to analyze is the relevance of a document judged by the user once his problem is solved. She collects about 200 relevance judgments expressed by the users approximately 2 weeks after the search took place. In this way, the users are allowed to review the search results, and their judgments should be more reliable. Sandore finds a moderate correlation between user satisfaction and precision of the search.

**Froehlich** (1991) claims that relevance is a “natural category,” acquired through experience, and not through definition. He also speaks of the “polarity” (“duality” might be another term) of relevance: On the

one side, there are topicality and “its universe” (a more “social” view) and on the other side there is the user’s context (a more “individual” view).

**Harter** (1992) derives, from the application of “psychological relevance” to IR, that “being on the topic” is neither a necessary nor a sufficient condition for “being relevant.”

**Hersh** (1994) explores the limitation of topicality (relevance of a document to a query on the topic) and situational relevance (relevance of information to an information need on all three components) in the medical domain, finding that none of them taken alone is adequate for evaluating an IRS. He then suggests a framework for the evaluation of IRSs that uses both kinds of relevance: Topicality for assessing different approaches to indexing and retrieval, and situational relevance to measure the impact of the IRS on the users. Moreover, a third aspect “that leads beyond the notions of relevance entirely” must be explored, in order to measure the outcome of the whole system-user interaction.

**Soergel** (1994) summarizes, in the introduction of his paper about indexing, some previously proposed definitions of topical relevance, pertinence, and utility. An entity is “topically relevant” if it can, in principle, help to answer user’s question. An entity is “pertinent” if topically relevant and “appropriate” for the user (i.e., the user can understand it and use the information obtained). An entity has “utility” if pertinent and if it gives to the user “new” (not already known) information.

**Brajnik, Mizzaro, and Tasso** (1995, 1996) present the evaluation of an intelligent user interface to an IRS, in which: (i) Not only “relevance” (the relevance of a document to a request on the topic), but also “utility” (relevance of a document to a request on topic and task) is measured; and (ii) the quality of the system-user interaction is explored.

The above presented framework shows how all the distinctions proposed by these authors are short-sighted: Many studies (of both “1959–1976” and “1977–present” periods) mistake system-relevance for topic-relevance, do not consider all the existing kinds of relevance, and so on. Anyway, in this period some studies (Regazzi, 1988; Saracevic et al., 1988; Saracevic & Kantor, 1988a, 1988b; Sandore, 1990; Brajnik et al., 1995, 1996) try to measure the until then retained unmeasurable relevances, sometimes comparing two different types of relevance; in this way, they get closer to the “top” relevance of Figure 1.

### Surrogates

Some studies continue the work of the 1960s aimed at understanding how relevance judgments are affected by different types of surrogate:

**Marcus, Kugel, and Benenfeld** (1978) compare the rele-

vance judgment expressed on the basis of various kinds of surrogates (title, abstract, keywords, keywords that appear in the query) and on the basis of the whole document. The findings suggest a *length hypothesis*: The quality of a surrogate is directly proportional to its length.

**Wilson** (1978) notes that the relevance to a surrogate cannot be the same as the relevance to a document, because *any* feature of a document may make it relevant to the user, and a surrogate is necessarily shorter than a document. He also calls for beyond-topical IRSs.

**Kazhdan** (1979) compares the relevance judgments obtained by seven types of surrogates (title, abstract, title plus extract, and so on), finding some “non-negligible” difference but noting that the relative ranking of the quality of the seven different surrogates does not change.

**Rorvig** (1987) finds that the relevance judgment expressed on the basis of the document and of the surrogate (abstract) are similar if the document is an image and the surrogate a textual description.

**Janes** (1991b) studies how the relevance judgment changes as more information becomes available to the user. Users see in sequence three surrogates of the same document, chosen from title, abstract, keyword, and bibliographic citation surrogates, in some order. The changing of judgment as the new abstract is presented is measured by a *motion index*. Abstracts are found to be by far the most important surrogate kind, followed by titles and by bibliographic information and keywords. The length hypothesis (Marcus et al., 1978) is negated by these findings.

The distinction “1959–1976” versus “1977–present” does not seem to affect the studies on the surrogates: The history regarding this issue apparently flows continuously from 1959 until present (though there are no studies on this topic in the last 5 years). In my opinion, two are the milestones of this line of research: The “length hypothesis” (Marcus et al., 1978) and the study by Janes (1991b), probably one of the most complete and systematic.

In the major part of these studies (in both periods), surrogate-based relevance judgments tend to become similar to full-document judgments as the surrogate is enriched: The quality of title surrogates is the lowest, followed by keywords, extract, and abstract surrogates. Nevertheless, some authors find different results, for instance Janes (1991b). This suggests that the “length hypothesis” (Marcus et al., 1978) seems too superficial: Also the quality of words, besides their quantity, should be taken into account.

### Criteria

During the “1959–1976” period, the beyond-topical criteria were individuated by experts. This line of research

is continued, especially at Syracuse University, with a new approach: The criteria are elicited directly from the users. The expressions *user defined criteria* and *document characteristics* (on which the criteria are based) are often used:

**Taylor** (1986) proposes the following six criteria adopted by users in choosing information (that is, expressing dichotomous relevance judgments): (1) Ease of use, (2) noise reduction, (3) quality, (4) adaptability (ability to respond to user's problem), (5) time saving, and (6) cost saving.

**Halpern, Nilan, Peek, and Snyder** (Halpern & Nilan, 1988; Nilan, Peek, & Snyder, 1988) sketch a methodology for (i) eliciting the criteria adopted by users in the evaluation of the *source* of the information, and (ii) individuating the time points at which the criteria are applied. In a preliminary study, they find about 40 criteria, among which are: Credentials, reputation, trust, expertise, love, financial considerations, time considerations, and so on.

**Regazzi** (1988) asks 32 judges to rate the importance of five document attributes (author, title, abstract, source of publication, and date of publication) and of six information attributes (accuracy, completeness, subject, suggestiveness, timeliness, and treatment), finding very different preferences among judges.

**Schamber** (1991a, 1991b) finds 10 criteria grouped in three categories mentioned by users of weather information evaluating the (multimedial) information received: (1) Information (accuracy, currency, specificity, geographic proximity), (2) source (reliability, accessibility, verifiability through other sources), and (3) presentation (dynamism, presentation quality, clarity).

**Park** (1992, 1993) elicits from academic users some criteria affecting relevance judgment, grouped into three categories: (1) "Internal context," containing criteria pertaining to the user's prior experience (for instance, expertise in subject literature, educational background); (2) "external context," factors concerning the search that is taking place (for instance, purpose of the search, stage of research); and (3) "problem (content) context," representing the motivations and the intended use of the information (for instance, obtaining definitions of something, or frameworks).

**Barry** (1993, 1994) identifies 23 "criterion categories," classified in the following seven "criterion category groups": (1) Information content of the document, (2) user's background/experience, (3) user's beliefs and preferences, (4) other information and sources within the environment, (5) sources of the documents, (6) document as a physical entity, and (7) user's situation. Three criteria are original: Effectiveness of a technique presented within the document, consensus within the field, and user's relationship with the author.

**Cool, Belkin, and Kantor** (1993) elicit from about 300 subjects 60 criteria underlying documents usefulness judgment, grouped in six categories: (1) Topic, (2) content/information (concepts and facts contained in the document), (3) format, (4) presentation, (5) values, and (6) oneself (personal need or use of the document).

**Thomas** (1993) identifies 18 factors, grouped in four categories, affecting Ph.D. students relevance judgment in an unfamiliar environment: (1) Information and knowledge sources, (2) feelings of uncertainty (affective, pragmatic and cultural, and academic), (3) endurance and coordination, and (4) establishing professional relationships.

**Bruce** (1994) individuates some "document characteristics" (author, title, keywords, source of publication, and date of publication) and "information attributes" (accuracy, completeness, content, suggestiveness, timeliness, treatment), that the judges might use when expressing their relevance judgments. He suggests that the importance ascribed to each of these parameters by the judge changes during the seeking of information.

**Howard** (1994) uses the psychological *personal construct theory* developed by Kelly (1955) to elicit the "personal constructs" (i.e., criteria) used in relevance judgment. She elicits from five judges the individual justifications for their relevant/nonrelevant judgments of 14 documents, and then asks three other subjects to group such criteria in two ways: By similarity and by "foci" ("topicality" and "informativeness"). Three (one for each subject) sets of groups of similar personal constructs are found, with cardinality seven, 13, and 12, respectively. Two of the three subjects agree that 39 constructs are labeled as topical ones, and 24 as "informative" (i.e., pertaining to the user and to the use of information) ones; only five constructs are not labeled. The results suggest to Howard that topicality and informativeness appear in the mental model of relevance, and that topicality seems more important than informativeness.

**Schamber** (1994) remarks that the percentage of "new" criteria discovered in recent studies is very low, and thus: (i) There seem to be a finite set of such criteria for users in all types of information problem situations, and (ii) probably almost all criteria have already been identified.

**Wang** (1994) performs a think aloud experiment with 25 real users in order to individuate and rank user criteria based on "document information elements" and on user personal knowledge. The criteria, ranked in decreasing importance order, and the corresponding "document information elements" are: Topicality (title, abstract, geographic location), orientation/level (title, abstract, author, journal), quality (author, journal, document type), subject area (author's

subject area, journal), novelty (title, author), recency (publication date), authority (author), and relation/origin (author).

This line of research has been latent for many years, until the publication of the seminal papers (Halpern & Nilan, 1988; Nilan et al., 1988). From 1988 to 1994, there has been a consistent increase of papers, due to the empirical studies (often doctoral dissertations) by Schamber (1991a, 1991b, 1994), Park (1992, 1993), Thomas (1993), Cool et al. (1993), Barry (1993, 1994), Bruce (1994), Howard (1994), and Wang (1994). The importance of these studies should be manifest: The existence of factors beyond topicality affecting user's relevance judgment is confirmed; the criteria directly elicited from users agree with the ones proposed by, or elicited from, experts in the studies of the "1959–1976" period; it seems that the users can individuate and discuss the beyond-topical criteria; such criteria seem to have been identified, and they can (and ought to) be taken into account in the construction of new-generation, beyond-topical IRSs. Anyway, most of the studies are by the authors' own admission, exploratory or preliminary ones. This, together with the recency of these studies, calls for caution and further work in this direction.

### *Dynamics*

Many studies of the "1977–present" period analyze how relevance judgments are time dependent, especially because of the previously seen documents:

**Brookes** (1980) notes that the documents retrieved by an IRS are similar, hence the relevance judgment of a document is unavoidably affected by the previously seen documents. He assumes that the utility of the following documents can only be diminished (not augmented) by the previous ones.

**Bookstein** (1983) defines a mathematical model, based on the statistical decision theory, that takes into account interdocument interaction, instead of considering each document in isolation from the others.

**Meadow** (1985, 1986) emphasizes the fact that the request changes while the user interacts with the intermediary and the IRS. In his opinion, this prevents the measurement of relevance.

**Eisenberg and Barry** (Eisenberg, 1986; Eisenberg & Barry, 1986; Eisenberg, 1988; Eisenberg & Barry, 1988) present experimental evidence of the *presentation order effect*, i.e., that the order of document presentation affects relevance and relevance judgments. This effect is more evident using a category rating scale score and less evident using a magnitude estimation score. The studies are anyway not definitive, and further research is invoked.

**Regazzi** (1988) ascribes to learning effects some of the dynamic nature of relevance found in his experiment.

**Swanson** (1988) in his third "postulate of impotence,"

claims that the relevance of a document depends on the other documents seen by the user.

**Tiamiyu and Ajiferuke** (1988) propose a mathematical model for taking into account interdocument dependencies. They define a "total relevance function" that assigns the relevance to a set of documents not merely summing up the relevances of each single document of the set, but considering "substitutability" (i.e., the relevance of a document is diminished by another document) and "complementarity" (i.e., the relevance of a document is increased by another document) relationships among documents.

**Katzer and Snyder** (1990) criticize the assumption that the information need of the user does not change during the interaction with the IRS. On this basis, they sketch a methodology for the evaluation of an IRS in which the user is asked to write three versions of his information need while interacting with the IRS. Such three versions are compared for finding differences, and the last version is the one used for the relevance judgment.

**Purgailis, Parker, and Johnson** (1990) experimentally find that if the user of an IRS is presented less than 15 documents, the presentation order effect discovered by Eisenberg and Barry, does not appear (with a three-point scale relevance judgment). It seems to appear when the users have to examine more than 15 documents.

**Harter** (1992) derives, from the application of "psychological relevance" to IR, that the user's mental state, and hence relevance, changes while reading citations, and that the real relevance judgment can take place only after the reading of the entire document.

**Bruce** (1994) describes a framework to observe the temporal evolution of the importance attributed by the user to some parameters affecting relevance judgment. He proposes three key points: The time in which the user has a problem, the time interval from the first request to the last query, and the time in which the user has solved his problem.

**Ottaviani** (1994) proposes a mathematical model of relevance judgments, based on *fractal theory*: The information received by a user interacting with an IRS forces him to change his question; then new information is received, the question is changed again, and so on, in a fractal-like manner.

**Smithson** (1994) uses as subjects 21 students preparing a dissertation, and measures three kinds of relevance: At the end of the online search, at the end of the research project, and by examining the citations in the final dissertations of the subjects. The relevance judgments for each subject change radically.

**Sutton** (1994) studies the information seeking behavior of *attorneys* when interacting with a full-text legal IRS. He describes the *mental model* of the law that an attorney builds and maintains, and he shows how this model is modified by the interaction between

the attorney on the one side, and (legal) information and IRS on the other.

**Wang** (1994) experimentally finds that the criteria used by a user for assessing the relevance of a document, and the “document information elements” on which such criteria are based, vary from document to document.

Also for these studies (like for those regarding the criteria) there is a consistent increase in number in the last years (since 1986). The main research topics in this last period are: (i) The existence of a presentation order effect (Brookes, 1980; Eisenberg, 1986; Eisenberg & Barry, 1986; Eisenberg, 1988; Eisenberg & Barry, 1988; Swanson, 1988; Purgailis, Parker, & Johnson, 1990); (ii) the dynamic nature of query, request, information need, and problem justifies at least in part the dynamic nature of relevance (Meadow, 1985, 1986; Katzer & Snyder, 1990; Ottaviani, 1994); (iii) cognitive considerations based on learning (Regazzi, 1988), mental models (Harter, 1992; Sutton, 1994), and criteria (Wang, 1994) can explain the variations in relevance judgments; (iv) the time point at which relevance is measured (Bruce, 1994; Smithson, 1994) is a key factor; (v) some mathematical models are proposed (Bookstein, 1983; Tamiyu & Ajiferuke, 1988).

These studies have important consequences for the construction of IRSs. In fact, they support the position, maintained by many researchers (Bates, 1989, 1990; Belkin, Oddy, & Brooks, 1982a, 1982b; Ingwersen, 1992), that we need *iterative* and *interactive* IRSs, i.e., systems that, stimulating a continued, iterative bidirectional communication of information, achieve an effective interaction with the user.

### Expression

The studies of the “1959–1976” period found no satisfactory answer to the problem of relevance judgment expression. This issue is again explored in the following works:

**Koll** (1979, 1981) again brings to attention the issue of relevance judgment expression, after about 10 years in which it has been neglected. He shows that intervally scaled relevance judgments may be used to compare hypotheses on alternative systems.

**Rorvig** (1985) demonstrates that it is possible to obtain transitive, interval measures of human judgments for documents whenever desirable.

**Eisenberg** (1986, 1988) finds that magnitude estimation is appropriate for the measurement of relevance and that it seems to be more robust, with respect to context variations, than category rating scales. Eisenberg finds a *context effect*: The relevance judgment for a particular document seems affected by the other documents being judged.

**Eisenberg and Hu** (Eisenberg, 1986; Eisenberg & Hu, 1987) examine dichotomous relevance judgments,

and find that when using category rating scales, the break between relevance and nonrelevance expressed by judges is below the midscale value.

**Foster** (1986) reviews Rorvig’s work (1985), and criticizes some of its conclusions.

**Halpern, Nilan, Peek, and Snyder** (Halpern & Nilan, 1988; Nilan et al., 1988) propose a methodology, derived mainly from Dervin’s *Sense-Making* (1983), for eliciting the criteria that users adopt when evaluating the information source. Such a methodology is claimed to be effective.

**Rorvig** (1988) surveys the development of psychometrics and the applications of psychometric measurement techniques in IR. He emphasizes previous unproductive work and some mistakes, both caused by an incomplete knowledge of psychometrics, and underlines the importance of some neglected work at System Development Corporation (Weis & Katter, 1967).

**Rorvig** (1990) proposes to substitute the usual relevance judgments with “preference” judgments, i.e., judgments of preference of one document over another one. He shows some experimental results that seem to confirm the reliability of this approach.

**Janes** (1991a) confirms the findings in (Eisenberg, 1986; Eisenberg & Hu, 1987): When judges collapse their scaled judgments into dichotomous judgments, the break between relevant and not relevant is below the middle of the scale.

**Janes and McKinney** (Janes, 1991b; Janes & McKinney, 1992; Janes, 1994) use line-length magnitude estimation for exploring the consistency of relevance judgment, and their work confirms the reliability of this method.

**Janes** (1993) analyzes the relevance judgments expressed by means of category rating scales in his own studies and in the 1960s studies. He notes that the judges use mainly the end points of the scales, and concludes that relevance seems to be mainly dichotomous.

**Bruce** (1994) empirically finds that magnitude estimation (numeric estimation and hand grip) is appropriate to let the judge express the importance ascribed to various characteristics of documents and information, and to measure how such importance is time dependent.

Many of these studies approach the issue of the relevance judgment expression through the application of psychometric and psychologic instruments (Rorvig, 1988), obtaining more encouraging results than the studies of the previous period. As a matter of fact, many studies of the “1977–present” period (Eisenberg, 1986; Eisenberg, 1988; Janes, 1991b; Janes & McKinney, 1992; Janes, 1994; Bruce, 1994) seem to demonstrate that magnitude estimation (numeric estimation, line length, and force hand grip) is an effective and reliable method for expressing relevance judgments, and that it is preferable

to both category rating scales and dichotomous judgments.

### Subjectiveness

The subjectiveness of relevance judgments is analyzed by the following researchers:

**Davidson** (1977) experimentally finds that almost all the subjectiveness in relevance judgment is systematic and depends on two variables: (1) Judge's expertise and interest in the area of the search and (2) judge's openness to information, i.e., judge's aptitude for perceiving messages as informative ones.

**Tessier, Crouch, and Atherton** (1977) note that many other features besides relevance affect user satisfaction (for instance, kind of interaction with the intermediary, library location, and so on).

**Figueiredo** (1978) finds a 57.2% agreement between librarians' and users' relevance judgments on a three-point category rating scale.

**Kazhdan** (1979) finds experimental evidence to support the weak hypothesis of Lesk and Salton, but not to support their strong hypotheses.

**Regazzi** (1988) maintains that the characteristics of judges explain much of the difference of relevance judgments. He compares eight different groups of four judges, groups obtained combining three parameters: "Type" (either researcher or student), "level" (either senior or junior), and "specialty" (either biomedicine or social science). The relevance judgments are affected by the group to which the judge belongs: The most important of the three parameters is specialty, followed by level and type.

**Swanson** (1988) claims that topical relevance judgments can be inconsistent, especially when expressed by non-users, and that the results of the evaluation of IRSs depend more on the circumstances of the retrieval judgment than on the system itself.

**Burgin** (1992) finds good agreements (from 40 to 55%) among judges of four different groups (users, online searching experts, and two kinds of subject experts, "more" and "less" expert) judging full-text documents.

**Janes and McKinney** (1992) compare users' relevance judgments with non-users' (information/library studies students and psychology students), finding a 0.62 specificity and a 0.68 sensitivity.

**Janes** (1994) compares users' relevance judgment with non-users' judgment of: Relevance (not defined), topicality (similarity to the topic), and utility (usefulness to the user). The judges belong to three different groups: Incoming students to a school of information/library science, experienced students in that school, and academic librarians. The study is an exploratory one (no claim of statistical reliability, and lack of definition of relevance, utility and topicality),

and the results are synthesized in the following table, adapted from Janes (1994):

|                      | Sensitivity | Specificity |
|----------------------|-------------|-------------|
| Incoming students    | 0.861       | 0.557       |
| Experienced students | 0.778       | 0.844       |
| Library staff        | 0.694       | 0.773       |

**Wang** (1994) experimentally finds that the criteria used by the users for assessing the relevance of a document, and the "document information elements" on which such criteria are based, vary from user to user.

**Ellis** (1996) studies the problem of measurement of the performance of IRSs, and maintains that using relevance judgments for measuring retrieval effectiveness is different from using an instrument (e.g., a thermometer) for measuring a physical quantity (temperature): In the first case, psychological methods are more suited.

**Harter** (1996) analyzes the literature concerning the factors affecting the relevance judgments and the experimental evaluation of IR systems. He derives that the assumption that the variations in relevance judgments do not significantly affect the measurement of IR systems performance (on which the Cranfield-like experiments are based) is not supported. He suggests a new approach to evaluation experiments, in which different "problem types" (different types of searches, request, and relevant documents) are evaluated separately.

The subjectiveness of relevance judgments seems less worrying in the "1977-present" period than in the "1959-1976" one, also in virtue of the studies that help to understand why and when this phenomenon manifests itself, i.e., which are the conditions (features of the judges, but also criteria and dynamics, see previous subsections) that lead to inconsistency (Davidson, 1977; Regazzi, 1988; Burgin, 1992; Janes, 1994). This line of research has obviously important consequences for the evaluation of IRSs: They are analyzed in Harter (1996).

### The End of the Period

The "1977-present" period of relevance history is closed by some surveys:

**Schamber, Eisenberg, and Nilan** (1990) survey the work on relevance and classify the various approaches under the labels "multidimensional," "cognitive," and "dynamic." Then the authors individuate the assumptions underlying the analyzed works and propose an alternative perspective based on different assumptions.

**Froehlich and Eisenberg** (1992) are the moderators of a forum about relevance. The works presented there are later published in JASIS (1994).

TABLE 2. Number of studies for each category for each year.\*

|       | Fo | Ki | Sr | Cr | Dy | Ex | Sj | Total |
|-------|----|----|----|----|----|----|----|-------|
| 59    |    | 2  |    |    |    |    |    | 2     |
| 60    | 1  | 2  |    |    |    |    |    | 3     |
| 61    |    |    | 2  |    |    |    |    | 2     |
| 62    |    |    |    |    |    |    |    | 0     |
| 63    | 1  | 1  |    | 1  |    |    |    | 3     |
| 64    | 1  |    | 1  |    | 1  |    | 1  | 4     |
| 65    |    | 1  |    |    |    |    | 1  | 2     |
| 66    | 1  | 3  |    |    | 1  |    | 2  | 7     |
| 67    | 1  | 1  | 3  | 4  |    | 4  | 7  | 20    |
| 68    | 1  | 2  |    |    |    | 1  | 1  | 5     |
| 69    |    |    | 1  |    |    |    | 2  | 3     |
| 70    | 3  | 1  |    |    |    |    | 1  | 5     |
| 71    | 1  |    |    | 1  |    |    |    | 2     |
| 72    |    | 1  |    |    |    |    |    | 1     |
| 73    | 1  | 3  | 2  | 2  |    |    |    | 8     |
| 74    | 1  | 1  |    |    | 1  |    |    | 3     |
| 75    | 1  |    |    |    |    |    |    | 1     |
| 76    | 1  |    |    |    |    |    |    | 1     |
| Total | 14 | 18 | 9  | 8  | 3  | 5  | 15 | 72    |
| 77    | 2  | 2  |    |    |    |    | 2  | 6     |
| 78    | 1  |    | 2  |    |    |    | 1  | 4     |
| 79    |    | 2  | 1  |    |    | 1  | 1  | 5     |
| 80    |    |    |    |    | 1  |    |    | 1     |
| 81    | 1  |    |    |    |    | 1  |    | 2     |
| 82    |    | 1  |    |    |    |    |    | 1     |
| 83    |    |    |    |    | 1  |    |    | 1     |
| 84    | 1  |    |    |    |    |    |    | 1     |
| 85    |    |    |    |    | 1  | 1  |    | 2     |
| 86    | 2  | 1  |    | 1  | 3  | 3  |    | 10    |
| 87    |    |    | 1  |    |    | 1  |    | 2     |
| 88    | 1  | 4  |    | 3  | 5  | 4  | 2  | 19    |
| 89    | 3  |    |    |    |    |    |    | 3     |
| 90    | 1  | 2  |    |    | 2  | 1  |    | 6     |
| 91    | 3  | 1  | 1  | 2  |    | 2  |    | 9     |
| 92    | 5  | 1  |    | 1  | 1  | 1  | 2  | 11    |
| 93    | 4  |    |    | 4  |    | 1  |    | 9     |
| 94    | 3  | 2  |    | 5  | 5  | 2  | 2  | 19    |
| 95    | 3  | 1  |    |    |    |    |    | 4     |
| 96    | 2  | 1  |    |    |    |    | 2  | 5     |
| Total | 32 | 18 | 5  | 16 | 19 | 18 | 12 | 120   |
| Total | 46 | 36 | 14 | 24 | 22 | 23 | 27 | 192   |

\* Fo, foundations; Ki, kinds; Sr, surrogates; Cr, criteria, Dy, dynamics; Ex, expression; Sj, subjectiveness.

**Froehlich** (1994) introduces the special topic issue of the *Journal of the American Society for Information Science* on the topic of relevance (JASIS, 1994), listing six common themes of the articles in that issue: (1) Inability to define relevance; (2) inadequacy of topicality; (3) variety of user criteria affecting relevance judgment; (4) the dynamic nature of information seeking behavior; (5) the need for appropriate methodologies for studying the information seeking behavior; and (6) the need for more complete cognitive models for IRS design and evaluation. Then, he proposes a synthesis of the articles: (1) User-relevance cannot be defined in a precise, “Cartesian” sense; (2) relevance is a natural cate-

gory, derived from experience; (3) the distinction between relevance and pertinence is not the same in users and librarians; (4) topicality is at the core of relevance; (5) relevance judgment is based on a finite set of criteria; (6) hermeneutics can give a framework for modeling user criteria and IRSs.

**Schamber** (1994) proposes, in the first ARIST (Annual Review of Information Science and Technology) chapter devoted entirely to relevance, three fundamental themes and related questions: (1) Behavior (What factors contribute to relevance judgments? What processes does relevant assessment entail?); (2) Measurement (What is the role of relevance in IR system evaluation? How should relevance judgment be measured?); and (3) Terminology (What should relevance, or various kinds of relevance, be called?). She does not answer these questions, but reviews the literature on these issues (concentrating on the period 1983–1994). The review is divided into five sections: (1) Background, in which she proposes three different views of relevance (system, information, and situation views) on the basis of a classical IR interaction model; (2) Evaluation and measurement, in which she discusses recall, precision, utility, and satisfaction; (3) Factor and effects, in which she describes the factors affecting relevance judgments; (4) User criteria, in which she reports recent results of the criteria identified by the users; and (5) Models and contexts, in which she sketches the interdisciplinary models and the theoretical approaches to relevance, and discusses some methodological problems.

The main feature of this period is a shift from system-oriented studies to studies, often based on the works by Belkin, Oddy, and Brooks (1982a, 1982b), Dervin (1983), and MacMullin and Taylor (MacMullin & Taylor, 1984; Taylor, 1986) (see also Dervin & Nilan, 1986) that take a more user-oriented, cognitive perspective.

## Discussion

I have already sketched an analysis of the various papers at the end of the subsections for each category, and of the whole periods at the end of the sections for each period. Here I continue such analysis from a more general point of view.

The total number of papers discussed in this work is 157 (see Table 1 at the beginning of the article). The mean number of studies for each year is higher in the “1977–present” period: For the “1959–1976” it is  $54/18 = 3$ ; for the “1977–present” it is  $103/20 = 5.15$ . Relevance is still an interesting topic of research.

A more refined analysis can be made on the basis of the seven categories of papers. Table 2 reports the number of studies for each year and for each category, with the totals for each period; the labels on the columns have the following meaning: “Fo” stands for foundations, “Ki”

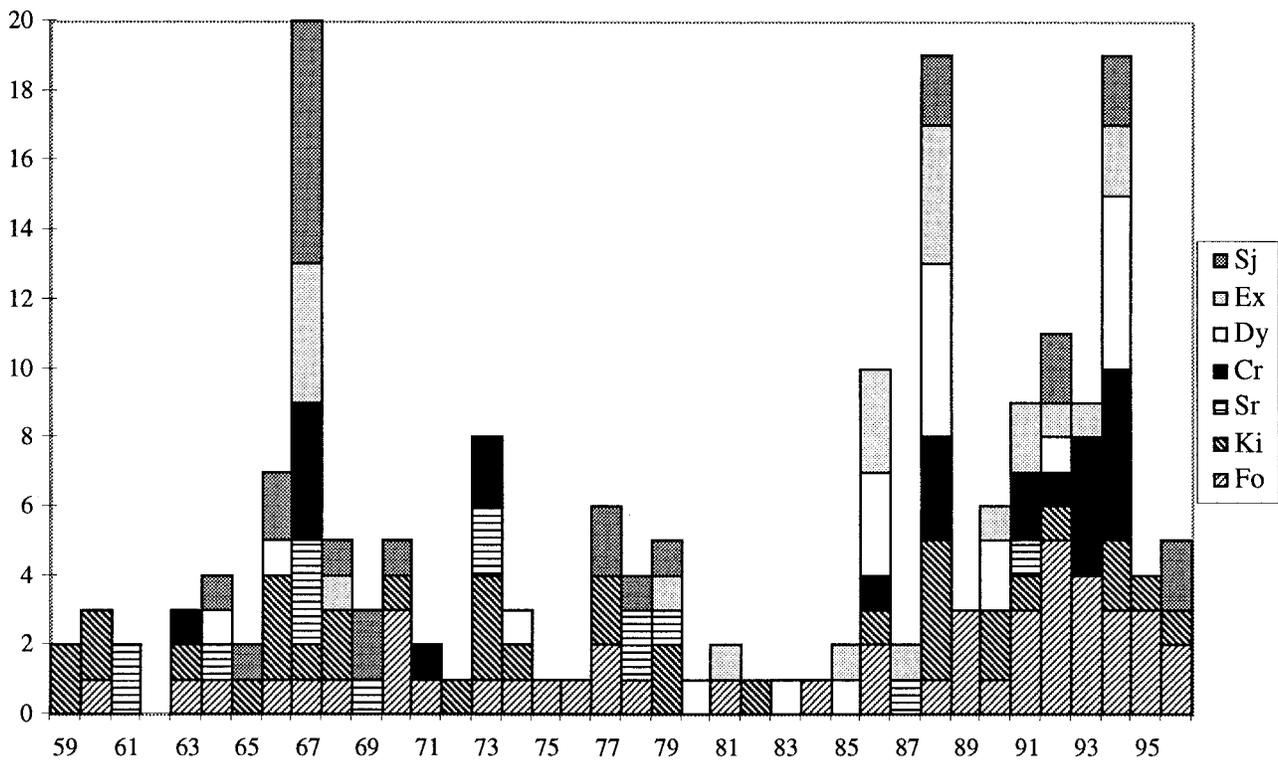


FIG. 2. Number of studies for each category for each year.

for kinds, “Sr” for surrogates, “Cr” for criteria, “Dy” for dynamics, “Ex” for expression, and “Sj” for subjectiveness. The total is 192, greater than 157, because some papers fall into more than one category. The data in the Table 2 are represented in graphical form in Figures 2 and 3. It is easy to note that:

- There is an increase in the number of studies in the middle of the 1960s and in the last 10 years or so;
- The studies on foundations (Fo) and kinds (Ki) are the most numerous; the studies on surrogates (Sr) are the least numerous; and the other categories are similar;
- The studies concerning foundations (Fo), criteria (Cr), dynamics (Dy), and expression (Ex) show a consistent increasing in number in the last period, especially in the last 10 years or so, while the papers in the other categories have a more uniform development.

It is difficult to understand and correctly interpret the history of “something” while such “something” is still going on. That is the difference between an historian and a reporter. Notwithstanding that, I try to interpret the above quantitative data to obtain some qualitative conclusion. In writing this article, I have implicitly assumed that we are at the end of a period. This feeling is confirmed, since a lot of studies have been published recently and some surveys have appeared: Schamber et al. (1990), Froehlich (1994), Schamber (1994), and this article. To-

day’s situation is very similar to the one in middle 1970s, that lead to Saracevic’s surveys.

Moreover, it seems clear that the “1959–1976” period is more oriented towards a relevance inherent in document and query: Some problems are noted, but operationally supposedly negligible. In the “1977–present” period, these problems are tackled, and the researchers try to understand, formalize, and measure a more subjective, dynamic, and multidimensional relevance: The relevance research is climbing the lattice of Figure 1.

## Conclusions

After the definition of a framework, the history of relevance has been presented, through a hopefully complete survey of the literature. The history has been divided into three conventional periods: “Before 1958,” “1959–1976,” and “1977–present.” Only a brief sketch of the first period has been presented, while the papers published during the “1959–1976” and “1977–present” periods have been analyzed and classified under seven different aspects (foundations, kinds, surrogates, criteria, dynamics, expression, and subjectiveness). A further section summarizes and analyzes the research on relevance from a general point of view.

In addition to presenting the history of relevance, this work should have shed some light on relevance itself. In order to emphasize how fundamental and not yet understood relevance is, I would like to conclude this article

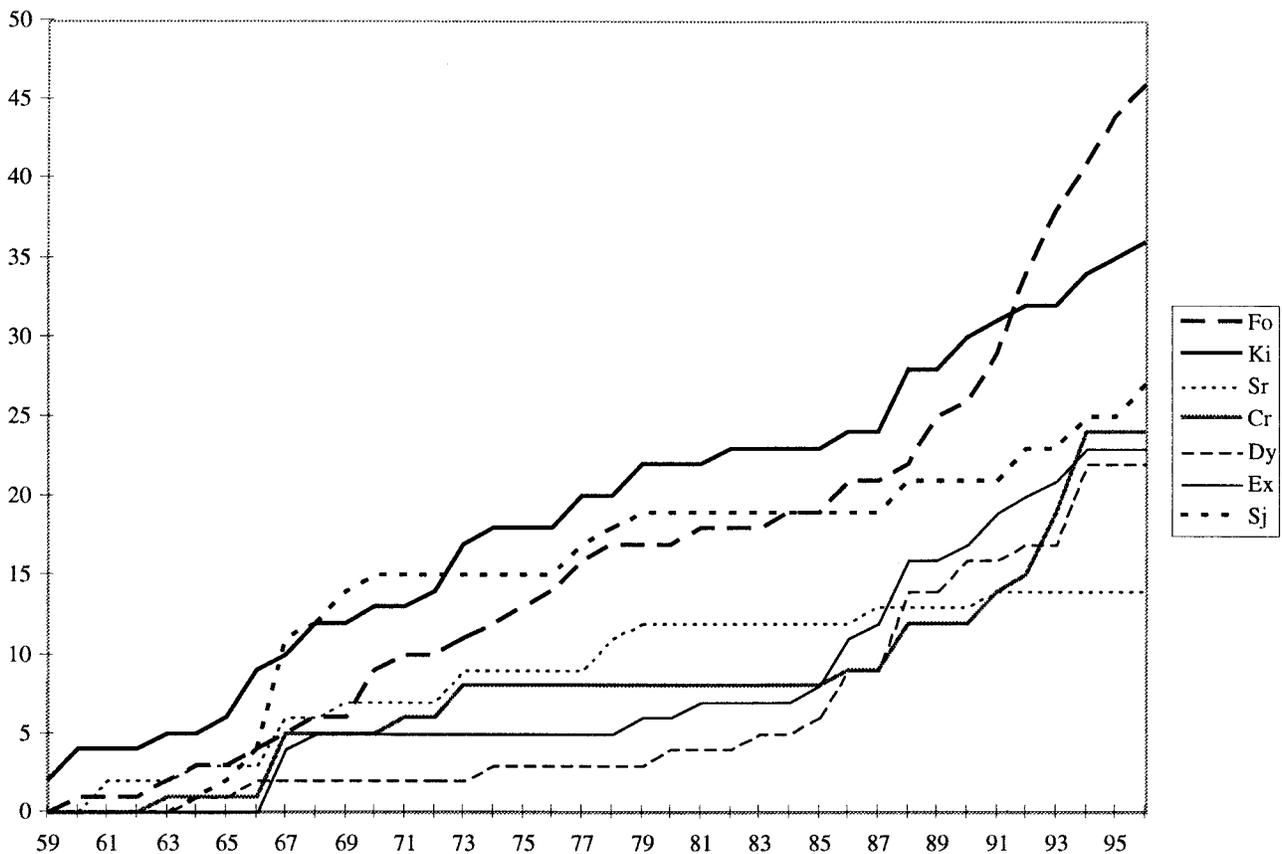


FIG. 3. Cumulative number of studies for each category.

with the closing sentences of the three major surveys on relevance that have appeared in the literature: “Our understanding of relevance in communication is so much better, clearer, deeper, broader than it was when information science started after the Second World War. But there is still a long, long way to go” (Saracevic, 1975, p. 339). “We consider the pursuit of a definition of relevance to be among the most exciting and central challenges of information science, one whose solution will carry us into the 21st century” (Schamber et al., 1990, p. 774). “Relevance is a necessary part of understanding human information behavior. The field should be encouraged by commonalities across perspectives, not discouraged by disagreements. Relevance presents a frustrating, provocative, rich, and—undeniably—relevant area of inquiry” (Schamber, 1994, p. 36).

### Acknowledgments

I am indebted to Mark Magennis for long and useful discussions, to Carlo Tasso and Giorgio Brajnik for many wise suggestions, and to Trudi Bellardo Hahn, Michael Buckland, two anonymous referees, and, again, Mark Magennis, for many useful comments on a draft of this paper.

### References

- Bar-Hillel, Y. (1960). *Some theoretical aspects of the mechanization of literature searching* (NSF Rep. AD-236 772 and PB-161 547). Washington, DC: Office of Technical Services.
- Barhydt, G. C. (1964). A comparison of relevance assessments by three types of evaluator. *Proceedings of the American Documentation Institute* (pp. 383–385). Washington, DC: American Documentation Institute.
- Barhydt, G. C. (1967). The effectiveness of non-user relevance assessments. *Journal of Documentation*, 23(2,3), 146–149, 251.
- Barry, C. L. (1993). *The identification of user relevance criteria and document characteristics: Beyond the topical approach to information retrieval*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, 45, 149–159.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13, 407–424.
- Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5), 575–591.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982a). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61–71.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982b). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation*, 38(3), 145–164.
- Belzer, J. (1973). Information theory as a measure of information con-

- tent. *Journal of the American Society for Information Science*, 24, 300–304.
- Bookstein, A. (1977). When the most “pertinent” document should not be retrieved—An analysis of the Swets model. *Information Processing & Management*, 13(6), 377–383.
- Bookstein, A. (1979). Relevance. *Journal of the American Society for Information Science* 30, 269–273.
- Bookstein, A. (1983). Information retrieval: A sequential learning process. *Journal of the American Society for Information Science*, 34, 331–342.
- Boyce, B. (1982). Beyond topicality: A two stage view of relevance and the retrieval process. *Information Processing & Management*, 18, 105–109.
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85–86.
- Brajnik, G., Mizzaro, S., & Tasso, C. (1995). Interface intelligenti a banche di dati bibliografici [Intelligent interfaces for bibliographic data bases]. In D. Saccà (Ed.), *Sistemi evoluti per basi di dati* (pp. 95–128). (In Italian) Milan, Italy: Franco Angeli.
- Brajnik, G., Mizzaro, S., & Tasso, C. (1996). Evaluating user interfaces to information retrieval systems: A case study on user support. *ACM SIGIR '96, 19th International Conference on Research and Development in Information Retrieval*, August 1996, Zurich, Switzerland. (pp. 128–136). New York: ACM Press.
- Brookes, B. C. (1980). Measurement in information science: Objective and subjective metrical space. *Journal of the American Society for Information Science*, 31, 248–255.
- Bruce, H. W. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science*, 45, 142–148.
- Bruza, P. D. (1993). *Stratified information disclosure: A synthesis between hypermedia and information retrieval*. Unpublished doctoral dissertation, University of Nijmegen, Nijmegen, The Netherlands.
- Bruza, P. D., & van der Weide, T. P. (1991). The modelling and retrieval of documents using index expression. *SIGIR Forum*, 25(2), 91–103.
- Bruza, P. D., & van der Weide, T. P. (1992). Stratified hypermedia structures for information disclosure. *The Computer Journal*, 35(3), 208–220.
- Burgin, R. (1992). Variations in relevance judgments and the evaluation of retrieval performance. *Information Processing & Management*, 28(5), 619–627.
- Chellas, B. F. (1980). *Modal logic*. Cambridge, MA: Cambridge University Press.
- Cool, C., Belkin, N. J., & Kantor, P. B. (1993). Characteristics of texts affecting relevance judgments. In M. E. Williams (Ed.), *Proceedings of the 14th National Online Meeting* (pp. 77–84). Medford, NJ: Learned Information, Inc.
- Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7(1), 19–37.
- Cooper, W. S. (1973a). On selecting a measure of retrieval effectiveness, part 1: The “subjective” philosophy of evaluation. *Journal of the American Society for Information Science*, 24, 87–100.
- Cooper, W. S. (1973b). On selecting a measure of retrieval effectiveness, part 2: Implementation of the philosophy. *Journal of the American Society for Information Science*, 24, 413–424.
- Cooper, W. S., & Maron, M. E. (1978). Foundation of probabilistic and utility-theoretic indexing. *Journal of the Association for Computing Machinery*, 25(1), 67–80.
- Crestani, F., & van Rijsbergen, C. J. (1995a). Information retrieval by logical imaging. *Journal of Documentation*, 51(1), 3–17.
- Crestani, F., & van Rijsbergen, C. J. (1995b). Probability kinematics in information retrieval. *ACM SIGIR, 18th International Conference on Research and Development in Information Retrieval*, Seattle, WA (pp. 291–299).
- Cuadra, C. A., & Katter, R. V. (1967a). *Experimental studies of relevance judgements* (NSF Rep. TM-3520/001, 002, 003, 3 vols.). Santa Monica, CA: Systems Development Corporation.
- Cuadra, C. A., & Katter, R. V. (1967b). Opening the black box of “relevance.” *Journal of Documentation*, 23(4), 291–303.
- Cuadra, C. A., & Katter, R. V. (1967c). The relevance of relevance assessment. *Proceedings of the American Documentation Institute* (Vol. 4, pp. 95–99). Washington, DC: American Documentation Institute.
- Davidson, D. (1977). The effect of individual differences of cognitive styles on judgments of document relevance. *Journal of the American Society for Information Science*, 28, 273–284.
- Dervin, B. (1983). *An overview of sense-making research: Concepts, methods and results to date*. Paper presented at the International Communication Association Annual Meeting, May 1983, Dallas, TX.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual review of information science and technology* (Vol. 21, pp. 3–33). White Plains, NY: Knowledge Industry.
- Devlin, K. (1991). *Logic and information*. Cambridge, UK: Cambridge University Press.
- Doyle, L. B. (1963). Is relevance an adequate criterion for retrieval system evaluation? *Proceedings of the American Documentation Institute* (pp. 199–200). Washington, DC: American Documentation Institute.
- Dym, E. D. (1967). Relevance predictability: I. Investigation background and procedures. In A. Kent et al. (Eds.), *Electronic handling of information: testing and evaluation* (pp. 175–185). Washington, DC: Thompson Book Co.
- Eisenberg, M. (1986). *Magnitude estimation and the measurement of relevance*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Eisenberg, M. B. (1988). Measuring relevance judgments. *Information Processing & Management*, 24(4), 373–389.
- Eisenberg, M., & Barry, C. (1986). Order effects: A preliminary study of the possible influence of presentation order on user judgments of document relevance. *Proceedings of the American Society for Information Science* (pp. 80–86). Medford, NJ: Learned Information.
- Eisenberg, M., & Barry, C. (1988). Order effects: A study of the possible influences of presentation order on user judgment of document relevance. *Journal of the American Society for Information Science*, 39, 293–300.
- Eisenberg, M., & Hu, X. (1987). Dichotomous relevance judgments and the evaluation of information systems. *Proceedings of the American Society for Information Science* (pp. 66–69). Medford, NJ: Learned Information.
- Ellis, D. (1984). Theory and explanation in information retrieval research. *Journal of Information Science*, 8, 25–38.
- Ellis, D. (1996). The dilemma of measurement in information retrieval research. *Journal of the American Society for Information Science*, 47, 23–36.
- Fairthorne, R. A. (1963). Implications of test procedures. In A. Kent (Ed.), *Information retrieval in action* (pp. 109–113). Cleveland: Case Western Reserve University Press.
- Figueiredo, R. C. (1978). Estudo comparativo de julgamentos de relevância do usuário e não-usuário de serviços de D.S.I. *Ciência da Informação—Rio de Janeiro*, 7, 69–78.
- Foskett, D. J. (1970). Classification and indexing in the social sciences. In *ASLIB Proceedings* (Vol. 22, pp. 90–100).
- Foskett, D. J. (1972). A note on the concept of “relevance.” *Information Storage and Retrieval*, 8(2), 77–78.
- Foster, J. (1986). Book review: An experiment in human preferences for documents in a simulated information system. *Canadian Journal of Information Science*, pp. 66–68.
- Froehlich, T. J. (1991). Towards a better conceptual framework for understanding relevance for information science research. *Proceedings of the American Society for Information Science*, Washington, DC (pp. 118–125). Medford, NJ: Learned Information.
- 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, 124–133.

- Froehlich, T. J., & Eisenberg, M. (1992). SIG/FIS—Relevance: A dialogue about fundamentals. *Proceedings of the American Society for Information Science*, Pittsburgh, PA. (pp. 330–331).
- Gifford, C., & Baumanis, G. J. (1969). On understanding user choices: Textual correlates of relevance judgments. *American Documentation*, 20(1), 21–26.
- Goffman, W. (1964). On relevance as a measure. *Information Storage and Retrieval* 2(3), 201–203.
- Goffman, W. (1970). A general theory of communication. In T. Saracevic (Ed.), *Introduction to information science* (chap. 13, pp. 726–747). New York: Bowker.
- Goffman, W., & Newill, V. A. (1966). Methodology for test and evaluation of information retrieval systems. *Information Storage and Retrieval*, 3(1), 19–25.
- Goffman, W., & Newill, V. A. (1967). Communication and epidemic processes. *Proceedings of the Royal Society*, 298, 316–334.
- Gordon, M. D., & Lenk, P. (1991). A utility theoretic examination of the probability ranking principle in information retrieval. *Journal of the American Society for Information Science*, 42, 703–714.
- Gull, C. D. (1956). Seven years of work on the organization of materials in special library. *American Documentation*, 7, 320–329.
- Hagerty, K. (1967). *Abstracts as a basis for relevance judgment*. Unpublished master's thesis, Graduate Library School, University of Chicago, Chicago, IL.
- Halpern, D., & Nilan, M. S. (1988). A step toward shifting the research emphasis in information science from the system to the user: An empirical investigation of source-evaluation behaviour in information seeking and use. *Proceedings of the American Society for Information Science* (pp. 169–176). Medford, NJ: Learned Information.
- Harper, W. L., Stalnak, R., & Pearce, G. (Eds.). (1981). *Ijs*. Dordrecht, The Netherlands: D. Reidel Publishing Company.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43, 602–615.
- Harter, S. P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science*, 47, 37–49.
- Hersh, W. (1994). Relevance and retrieval evaluation: Perspectives from medicine. *Journal of the American Society for Information Science*, 45, 201–206.
- Hillman, D. J. (1964). The notion of relevance (I). *American Documentation*, 15(1), 26–34.
- Hoffman, J. M. (1965). *Experimental design for measuring the intra- and inter-group consistence of human judgment for relevance*. Unpublished master's thesis, Georgia Institute of Technology, Atlanta, GA.
- Howard, D. L. (1994). Pertinence as reflected in personal constructs. *Journal of the American Society for Information Science*, 45, 172–185.
- Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.
- Janes, J. W. (1989). *Towards a search theory of information*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Janes, J. W. (1991a). The binary nature of continuous relevance judgments: A case study of users' perceptions. *Journal of the American Society for Information Science*, 42(10), 754–756.
- Janes, J. W. (1991b). Relevance judgments and the incremental presentation of document representations. *Information Processing & Management*, 27(6), 629–646.
- Janes, J. W. (1993). On the distribution of relevance judgments. *Proceedings of the American Society for Information Science* (pp. 104–114). Medford, NJ: Learned Information.
- Janes, J. W. (1994). Other people's judgments: A comparison of user's and other's judgments of document relevance, topicality, and utility. *Journal of the American Society for Information Science*, 45(3), 160–171.
- Janes, J. W., & McKinney, R. (1992). Relevance judgments of actual users and secondary judges. *Library Quarterly*, 62, 150–168.
- JASIS (1994). Special topic issue: Relevance research. *Journal of the American Society for Information Science*, 45(3).
- Katter, R. V. (1967). Study of document representations: Multidimensional scaling of index terms (Tech. Rep., 100 p.). Santa Monica, CA: Systems Development Corporation.
- Katter, R. V. (1968). The influence of scale form on relevance judgment. *Information Storage and Retrieval*, 4(1), 1–11.
- Katzer, J., & Snyder, H. (1990). Toward a more realistic assessment of information retrieval performance. *Proceedings of the American Society for Information Science* (pp. 80–85). Medford, NJ: Learned Information.
- Kazhdan, T. V. (1979). Effects of subjective expert evaluation of relevance on the performance parameters of a document-based information-retrieval system. *Nauchno-Tekhnicheskaya Informatsiya, Seriya 2*, 13, 21–24.
- Kelly, G. A. (1955). *The psychology of personal constructs: Volume 1: A theory of personality*. New York: W. W. Norton.
- Kemp, D. A. (1974). Relevance, pertinence and information system development. *Information Storage and Retrieval*, 10(2), 37–47.
- Kochen, M. (1974). *Principles of information retrieval*. Los Angeles, CA: Melville.
- Koll, M. B. (1979). *The concept space in information retrieval systems as a model of human concept relations*. Unpublished doctoral dissertation, Syracuse University, School of Information Studies, Syracuse, NY.
- Koll, M. B. (1981). Information retrieval theory and design based on a model of the user's concept relations. In R. N. Oddy, S. E. Robertson, C. J. van Rijsbergen, & P. W. Williams (Eds.), *Information retrieval research* (pp. 77–93). London: Butterworths.
- Lalmas, M. (1996). *Theories of information and uncertainty for the modelling of information retrieval: An application to situation theory and Dempster-Shafer's theory of evidence*. Unpublished doctoral dissertation, University of Glasgow, Glasgow, UK.
- Lalmas, M., & van Rijsbergen, C. J. (1992). A logical model of information retrieval based on situation theory. *BCS 14th Information Retrieval Colloquium*, Lancaster (pp. 1–13).
- Lalmas, M., & van Rijsbergen, C. J. (1993). Situation theory and Dempster-Shafer's theory of evidence for information retrieval. *Proceedings of Workshop on Incompleteness and Uncertainty in Information Systems*, Concordia University, Montreal, (pp. 62–67).
- Lalmas, M., & van Rijsbergen, C. J. (1996). Information calculus for information retrieval. *Journal of the American Society for Information Science*, 47, 385–398.
- Lancaster, F. W. (1979). *Information retrieval systems: Characteristics, testing and evaluation* (2nd ed.). New York: John Wiley & Sons.
- Lesk, M. E., & Salton, G. (1968). Relevance assessments and retrieval system evaluation. *Information Storage and Retrieval*, 4(3), 343–359.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, 16(12), 317–323.
- MacMullin, S. E., & Taylor, R. S. (1984). Problem dimensions and information traits. *The Information Society*, 3(1), 91–111.
- Marcus, R. S., Kugel, P., & Benenfeld, A. R. (1978). Catalog information and text as indicators of relevance. *Journal of the American Society for Information Science*, 29, 15–30.
- Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28, 38–43.
- Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing, and information retrieval. *Journal of the Association for Computing Machinery*, 7(3), 216–244.
- Meadow, C. T. (1985). Relevance? [Letter to the editor]. *Journal of the American Society for Information Science*, 36, 354–355.
- Meadow, C. T. (1986). Problems of information science research. *Canadian Journal of Information Science*, 11, 18–23.
- Meghini, C., Sebastiani, F., Straccia, U., & Thanos, C. (1993). A model of information retrieval based on a terminological logic. *ACM SIGIR, 16th International Conference on Research and Development in Infor-*

- ation Retrieval, Pittsburgh, PA (pp. 298–307). New York: ACM Press.
- Mizzaro, S. (1995). Le differenti relevance in information retrieval: Una classificazione. *Atti del Congresso Annuale AICA '95* (Vol. I, pp. 361–368) [“The various relevances in information retrieval: A classification. In *Proceedings of the Annual Conference AICA '95*]. (In Italian)
- Mizzaro, S. (1996a). A cognitive analysis of information retrieval. In P. Ingwersen and N. O. Pors (Eds.), *Information Science: Integration in Perspective—Proceedings of COLISZ* (pp. 233–250). Copenhagen, Denmark: The Royal School of Librarianship.
- Mizzaro, S. (1996b). *Relevance: The whole (hi)story* (Research Rep. UDMI/12/96/RR). Department of Mathematics and Computer Science, University of Udine, Udine, Italy.
- Mooers, C. S. (1950). Coding, information retrieval, and the rapid selector. *American Documentation*, 1(4), 225–229.
- Nie, J. Y. (1988). An outline of a general model for information retrieval systems. *ACM SIGIR, 11th International Conference on Research and Development in Information Retrieval*, Grenoble, France (pp. 495–506). Presses Universitaires de Grenoble.
- Nie, J. Y. (1989). An information retrieval model based on modal logic. *Information Processing & Management*, 25(5), 477–491.
- Nie, J. Y. (1992). Towards a probabilistic modal logic for semantic-based information retrieval. *ACM SIGIR, 15th International Conference on Research and Development in Information Retrieval*, Copenhagen, Denmark (pp. 140–151). New York: ACM Press.
- Nie, J. Y., Brisebois, M., & Lepage, F. (1995). Information retrieval as counterfactual. *The Computer Journal*, 38(8), 643–657.
- Nilan, M. S., Peek, R. P., & Snyder, H. W. (1988). A methodology for tapping user evaluation behaviors: An exploration of users’ strategy, source and information evaluating. *Proceedings of the American Society for Information Science* (pp. 152–159). Medford, NJ: Learned Information.
- O’Brien, A. (1990). Relevance as an aid to evaluation in OPACs. *Journal of Information Science*, 16, 265–271.
- O’Connor, J. (1967). Relevance disagreements and unclear request forms. *American Documentation*, 18(3), 165–177.
- O’Connor, J. (1968). Some questions concerning “information need.” *American Documentation*, 19(2), 1 200–203.
- O’Connor, J. (1969). *Some independent agreements and resolved disagreements about answer providing documents. American Documentation*, 20(4), 311–319.
- Ottaviani, J. S. (1994). The fractal nature of relevance: A hypothesis. *Journal of the American Society for Information Science*, 45, 263–272.
- Paisley, W. J. (1968). Information needs and uses. In *Annual Review of Information Science and Technology*, 3, 1–30.
- Park, T. K. (1992). *The nature of relevance in information retrieval: An empirical study*. Unpublished doctoral dissertation, School of Library and Information Science, Indiana University, Bloomington, IN.
- Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *Library Quarterly*, 63, 318–351.
- 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.
- Perry, J. W. (1951). Superimposed punching of numerical codes on handsorted, punch cards. *American Documentation*, 2(4), 205–212.
- Price, D. J. (1965). Networks of scientific papers. *Science*, 149(3683), 510–515.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of the American Society for Information Science*, 27, 292–306.
- Purgailis, Parker, L. M. P., & Johnson, R. E. (1990). Does order of presentation affect users’ judgement of documents? *Journal of the American Society for Information Science*, 41, 493–494.
- Rath, G. J., Resnick, A., & Savage, T. R. (1961). Comparisons of four types of lexical indicators of content. *American Documentation*, 12(2), 126–130.
- Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval systems. *ASLIB Proceedings*, 18, 316–324.
- Rees, A. M., & Saracevic, T. (1963). Conceptual analysis of questions in information retrieval systems. *Proceedings of the American Documentation Institute* (pp. 175–177). Washington, DC: American Documentation Institute.
- Rees, A. M., & Saracevic, T. (1966). The measurability of relevance. *Proceedings of the American Documentation Institute* (pp. 225–234). Washington, DC: American Documentation Institute.
- Rees, A. M., & Schulz, D. G. (1967). *A field experimental approach to the study of relevance assessments in relation to document searching* (2 vols., NSF Contract No. C-423). Center for Documentation and Communication Research, School of Library Science, Case Western Reserve University, Cleveland, OH.
- Regazzi, J. J. (1988). Performance measures for information retrieval systems—An experimental approach. *Journal of the American Society for Information Science*, 39, 235–251.
- Resnick, A. (1961). Relative effectiveness of document titles and abstracts for determining relevance of documents. *Science*, 134(3483), 1004–1006.
- Resnick, A., & Savage, T. R. (1964). The consistence of human judgments of relevance. *American Documentation*, 15(2), 93–95.
- Robertson, S. E. (1977). The probabilistic character of relevance. *Information Processing & Management*, 13, 247–251.
- Rorvig, M. E. (1985). *An experiment in human preferences for information in a simulated information system*. Unpublished doctoral dissertation, The University of California, School of Library and Information Studies, Berkeley, CA.
- Rorvig, M. E. (1987). The substitutibility of images for textual descriptions of archival materials. In K. D. Lehmann & H. Strohl-Goebel (Eds.), *The applications of micro-computers in information, documentation and libraries* (pp. 407–415). Amsterdam, The Netherlands: North-Holland.
- Rorvig, M. E. (1988). Psychometric measurement and information retrieval. *Annual Review of Information Science and Technology*, 23, 157–189.
- Rorvig, M. E. (1990). The simple scalability of documents. *Journal of the American Society for Information Science*, 41, 590–598.
- Sandore, B. (1990). Online searching: What measures satisfaction? *Library and Information Science Research*, 12, 33–54.
- Saracevic, T. (1969). Comparative effects of titles, abstracts and full texts on relevance judgements. *Proceedings of the American Society for Information Science*, Washington, DC (pp. 293–299).
- Saracevic, T. (1970a). The concept of “relevance” in information science: A historical review. In T. Saracevic (Ed.), *Introduction to information science*, (pp. 111–151). New York: R. R. Bowker.
- Saracevic, T. (1970b). On the concept of relevance in information science. Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, Ohio.
- Saracevic, T. (1970c). Ten years of relevance experimentation: A summary and synthesis of conclusions. *Proceedings of the American Society for Information Science* (pp. 33–36).
- Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, 321–343.
- 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.
- Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39, 177–196.
- Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39, 197–216.
- Saracevic, T., Kantor, P., Chamis, A., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39, 161–176.

- Schamber, L. (1991a). Users' criteria for evaluation in a multimedia environment. *Proceedings of the American Society for Information Science*, Washington, DC (pp. 126–133). Medford, NJ: Learned Information.
- Schamber, L. (1991b). *Users' criteria for evaluation in a multimedia information seeking and use situation*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*, 29, 3–48.
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management*, 26(6), 755–776.
- Sebastiani, F. (1994). A probabilistic terminological logic for information retrieval. *ACM SIGIR, 17th International Conference on Research and Development in Information Retrieval*, Dublin, Ireland (pp. 122–130). London: Springer Verlag.
- Shirey, D. L., & Kurfeerst, H. (1967). Relevance predictability. II Data reduction. In A. Kent et al., (Eds.), *Electronic handling of information: Testing and evaluation* (pp. 187–198). Washington, DC: Thompson Book Co.
- Smithson, S. (1994). Information retrieval evaluation in practice: A case study approach. *Information Processing & Management*, 30, 205–221.
- Soergel, D. (1994). Indexing and retrieval performance: The logical evidence. *Journal of the American Society for Information Science*, 45, 589–599.
- Su, L. T. (1991). *An investigation to find appropriate measures for evaluating interactive information retrieval*, Unpublished doctoral dissertation, Rutgers, the State University of New Jersey, New Brunswick, NJ.
- Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing & Management*, 28(4), 503–516.
- Su, L. T. (1994). The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3), 207–217.
- Sutton, S. A. (1994). The role of attorney mental models of law in case relevance determinations: An exploratory analysis. *Journal of the American Society for Information Science*, 45, 186–200.
- Swanson, D. R. (1977). Information retrieval as a trial-and-error process. *Library Quarterly*, 47(2), 128–148.
- Swanson, D. R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *Library Quarterly*, 56, 389–398.
- Swanson, D. R. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science*, 39(2), 92–98.
- Taube, M. (1955). Storage and retrieval of information by means of the association of ideas. *American Documentation*, 6(1), 1–17.
- Taube, M. (1965). A note on the pseudo-mathematics of relevance. *American Documentation*, 16(2), 69–72.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29, 178–194.
- Taylor, R. S. (1986). *Value-added processes in information systems*. Norwood, NJ: Ablex Publishing.
- Tessier, J. A. (1981). *Toward the understanding of user satisfaction: A multivariate study of user evaluations of computer-based literature searches in medical libraries*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Tessier, J. A., Crouch, W. W., & Atherton, P. (1977). New measures of user satisfaction with computer-based literature searches. *Special Libraries*, 68(11), 383–389.
- Thomas, N. P. (1993). Information seeking and the nature of relevance: Ph.D. student orientation as an exercise in information retrieval. *Proceedings of the American Society for Information Science*, Columbus, OH, pp. 126–130. Medford, NJ: Learned Information.
- Thompson, C. W. N. (1973). The functions of abstracts in the initial screening of technical documents by users. *Journal of the American Society for Information Science*, 24, 270–276.
- Tiamiyu, A. M., & Ajiferuke, I. Y. (1988). A total relevance and document interaction effects model for the evaluation of information retrieval processes. *Information Processing & Management*, 24(4), 391–404.
- Urquhart, D. J. (1959). *Use of scientific periodicals*. *Proceedings of the International Conference on Scientific Information*, (Vol. 1, pp. 277–290). Washington, DC: National Academy of Sciences.
- van Rijsbergen, C. J. (1986a). A new theoretical framework for information retrieval. *ACM SIGIR, International Conference on Research and Development in Information Retrieval*, Pisa, Italy (p. 200).
- van Rijsbergen, C. J. (1986b). A non-classical logic for information retrieval. *The Computer Journal*, 29, 481–485.
- van Rijsbergen, C. J. (1989). Towards an information logic. *ACM SIGIR, 12th International Conference on Research and Development in Information Retrieval*, Cambridge, MA, (pp. 77–86).
- Vickery, B. C. (1959a). The structure of information retrieval systems. *Proceedings of the International Conference on Scientific Information* (Vol. 2, pp. 1275–1290) Washington, DC: National Academy of Sciences.
- Vickery, B. C. (1959b). Subject analysis for information retrieval. *Proceedings of the International Conference on Scientific Information* (Vol. 2, pp. 855–865). Washington, DC: National Academy of Sciences.
- Wang, P. (1994). *A cognitive model of document selection of real users of information retrieval systems* (Doctoral dissertation), University of Maryland, College of Library and Information Science, College Park, MD. (University Microfilms No. AAI9514595).
- Weis, R. L., & Katter, R. V. (1967). *Multidimensional scaling of documents and surrogates* (Tech. Memorandum SP-2713, 29 p.). Santa Monica, CA: Systems Development Corporation.
- Wilson, P. (1968). Relevance. In *Two kinds of power* (pp. 45–54). Berkeley: University of California Press.
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9(8), 457–471.
- Wilson, P. (1978). Some fundamental concepts of information retrieval. *Drexel Library Quarterly*, 14, 10–24.
- Wilson, P. (1993). Communication efficiency in research and development. *Journal of the American Society for Information Science*, 44, 376–382.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.