On Recommending

Jonathan Furner
University of California, Los Angeles, 300 Young Drive North, Mailbox 951520, Los Angeles, CA 90036-6904
E-mail: jfurner@ucla.edu

The core of any document retrieval system is a mechanism that ranks the documents in a large collection in order of the likelihood with which they match the preferences of any person who interacts with the system. Given a broader interpretation of "recommending" than is commonly accepted, such a preference ordering may be viewed as a recommendation, made by the system to the information seeker, that is itself typically derived through synthesis of multiple preference orderings expressed as recommendations by indexers, information seekers, and document authors. The ERIn (Evaluation–Recommendation–Information) model, a decision-theoretic framework for understanding information-related activity, highlights the centrality of recommending in the document retrieval process, and may be used to clarify the respects in which indexing, rating, and citation may be considered analogous, as well as to make explicit the points at which content-based, collaboration-based, and context-based flavors of document retrieval systems vary.

Introduction and Overview

Since the early 1990s, the act of *recommending* (i.e., of making a recommendation) has become increasingly of interest to members of the information studies community, and especially to designers of document retrieval systems and digital libraries. In general, the core of any *recommender system* is a mechanism that carries out analysis of a large collection of real-life objects, so that the objects in the collection may be ranked in order of the likelihood with which they match the preferences of any person who interacts with the system. Typically, the highest ranked objects will be presented as recommendations or suggestions to the user. Such recommendations may be viewed as predictions of the user's preferences. Many recommender systems currently incorporate mechanisms for ranking documents on the basis of the extent to which they have received recommendations in the course of prior system usage,[1] and information seekers are thus given the opportunity to benefit from a form of *social feedback* or indirect *collaboration*.

Although the literature on specific implementations of recommendation-based retrieval mechanisms has proliferated, extended explorations of the *concept* of recommending are rare. In this article, a model of information-related activity is proposed in which the act of recommending is assigned a privileged position. It is argued that both indexing and citation can usefully be viewed as forms of recommendation, and that hybrid systems integrating content-based, collaboration-based, and context-based approaches to retrieval[2] may, consequently, be promoted as recommender systems *par excellence*.

In succeeding sections, then, a model of information-related activity that borrows from microeconomic theory is presented as that which underlies current conceptions of recommender systems; a clarification is provided of the dimensions on which such systems may be distinguished from their content-based counterparts; the analogy between recommendation, citation, and indexing is developed; and conclusions are drawn as to the implications of this exercise for design practice. First, however, a prefatory note on the terminology and literature of the field is required.

A Preliminary Note on Terminology

The application to systems of the term "recommenders" was introduced in an electronic mail message by Varian (1996), as a "better name than 'collaborative filtering'" for the "general category of interest" to attendees at a workshop held a few days earlier in Berkeley, CA, in March 1996 (see Table 1). The widespread usage of the term "recommender system" appears to date from the publication in 1997 of a special issue of *Communications of the ACM* (*CACM*),

Received October 10, 2001; Revised January 28, 2002; accepted January 28, 2002

[1] To take a well-known example, Amazon.com (see http://www.amazon.com/) implements a widely used system that recommends books and CDs to prospective purchasers in this manner.

[2] These approaches are defined in the unifying content–collaboration–context (CCC) model developed by Furner (in review).

TABLE 1. Workshops on recommender systems, 1996–2001.

Title (Organizer)	Workshop on Collaborative Filtering (University of California, Berkeley)
Location, date	Berkeley, CA, March 16, 1996
URL	http://www.sims.berkeley.edu/resources/collab/
Title (Organizer)	Workshop on Filtering and Collaborative Filtering (DELOS)
Location, date	Budapest, Hungary, November 10–12, 1997
URL	http://www.sztaki.hu/conferences/delosbudapest/schedule.jhtml
Title (Organizer)	Workshop on Recommender Systems (AAAI)
Location, date	Madison, WI, July 27, 1998
URL	http://www.aaai.org/Press/Reports/Workshops/ws-98-08.html
Title (Organizer)	Workshop: Interacting with Recommender Systems (ACM SIGCHI)
Location, date	Pittsburgh, PA, May 15–16, 1999
URL	http://www.darmstadt.gmd.de/rec99/
Title (Organizer)	Workshop on Machine Learning for Information Filtering (IJCAI)
Location, date	Stockholm, Sweden, August 1, 1999
URL	http://www-ai.cs.uni-dortmund.de/EVENTS/IJCAI99-MLIF/
Title (Organizer)	Workshop on Recommender Systems: Algorithms and Evaluation (ACM SIGIR)
Location, date	Berkeley, CA, August 19, 1999
URL	http://www.cs.umbc.edu/~ian/sigir99-rec/
Title (Organizer)	Workshop on Personalisation and Recommender Systems in Digital Libraries (DELOS/NSF)
Location, date	Dublin, Ireland, June 18–20, 2001
URL	http://www.compapp.dcu.ie/delos/programme.html
Title (Organizer)	Workshop on Recommender Systems (ACM SIGIR)
Location, date	New Orleans, LA, September 13, 2001
URL	http://www.cs.orst.edu/~herlock/rsw2001/

edited by Resnick and Varian (1997), with that topic as its theme. Prior to then, systems of the kind to which the term refers were often characterized as "collaborative filtering" systems, following the usage established by Goldberg, Nichols, Oki, and Terry (1992) in their report of Tapestry, a program that allowed users to filter incoming streams of electronic mail messages in novel ways.[3] Following the 1996 event in Berkeley, a succession of workshops on collaborative filtering or recommender systems have been held: in Budapest, Hungary (November 1997); Madison, WI (July 1998); Pittsburgh, PA (May 1999); Stockholm, Sweden (August 1999); Berkeley, CA (August 1999); Dublin, Ireland (June 2001); and New Orleans, LA (September 2001; see Table 1). Papers reporting on recommender-system research are also regularly contributed to the conferences of the Association for Computing Machinery's Special Interest Groups on computer–human interaction (SIGCHI),[4] information retrieval (SIGIR),[5] and electronic commerce (SIGecom),[6] inter alia. Readers seeking to understand current issues and trends in the field are referred to the latest proceedings in these series of workshops and conferences. Influential accounts of system design include those given by Hill, Stead, Rosenstein, and Furnas (1995), Shardanand and Maes (1995), and Konstan, Miller, Maltz, Herlocker, Gordon, and Riedl (1997); Pazzani (1999), Paepcke, Garcia-Molina, Rodriguez-Mula, and Chu (2000), and Schafer, Konstan and Riedl (2001) provide helpful overviews.

There are two separate senses in which we may classify systems as recommender systems. In one (less common) sense, we may ascribe the responsibility for recommending to the system itself; it is the function of the system to provide recommendations for its individual users as to the order in which the objects in a collection should be considered. In another (more common) sense, the recommending is done by the users of the system; it is the function of the system to synthesize multiple users' recommendations of objects in the form of a single ranking for the individual user.

This distinction may appear trivial, but it should be clear that the class of recommender systems is far more inclusive if we accept the first of these two readings. For in that case any retrieval system that implements a mechanism by which the objects in a collection are ranked in order of their predicted utility[7] for the searcher may be regarded as a recommender system; the set of objects retrieved comprise the recommendation—in other words, the suggestion or the prediction—made by the system. "Recommender system" may thus be viewed as a *synonym* for "retrieval system," and *document retrieval* (or document recommendation) *systems* defined, on the basis of the type of objects being retrieved or recommended, as systems that recommend documents or document-like objects to the searcher.

On the other hand, if we accept the second of the two readings, the class of recommender systems may be construed as a *subset* of the class of retrieval systems. Recommender systems are then those retrieval systems that effect retrieval specifically through analysis of the judgments of utility made by previous users, rather than or in addition to (for instance) direct analysis of the contents of items (as in *content-based* systems), or analysis of the relationships perceived to exist between items (as in *context-based* systems; Furner, in review).

[3] This earlier article was also a contribution to a themed issue of *CACM*, on that occasion covering the topic of information filtering (Loeb & Terry, 1992). The distinction between information filtering and information retrieval is considered by Belkin and Croft (1992) and revisited later in this article.

[4] See http://www.acm.org/sigchi/.

[5] See http://www.acm.org/sigir/.

[6] See http://www.acm.org/sigecom/.

[7] "Utility" is used here to denote the set of criteria on which the searcher bases her evaluation of objects when constructing a preference ordering of them. The significance of these concepts are discussed in more detail in a later section.

However, even if we do accept the second reading, we may still wish to argue that all retrieval systems are recommender systems (as in the first reading) by contending that the assignment of documents to terms (indexing) that goes on in content-based systems, and the linking of documents to one another (citing) that goes on in context-based systems, are both instances of recommending activity by which preference orderings are expressed in the form of judgments of utility, relevance, relatedness, or approval. Indeed, this is the argument that is made in this article.

Information-Related Activity: The ERIn Model

In this section, I present a descriptive model of certain kinds of information-related activity.[8] It is intended that the simplifications and emphases on which this model is based will be helpful ones to adopt in the framing of any subsequent analysis of the motivations for, influences on, and products of such activity. It is also intended that the particular analogies and distinctions that are suggested by this model will inspire designers of operational information systems to develop techniques and features that support information-related activity more successfully. In subsequent sections, some of these analogies and distinctions are described in detail; in particular, the act of recommending is identified as an operation that is fundamental to all information systems, and some implications of such a conceptualization for design practice are highlighted.

The base of the model is a distinction drawn between three types of information-related activity, as follows: [A] the act of determining the value, worth, or virtue of engaging in the examination of a given document; [B] the act of expressing an observable statement of the perceived worth of a given document; and [C] the act of informing oneself by examining the content of a given document.

We may call these categories of activity [A] *Evaluative*, [B] *Recommending*, and [C] *Informative* activity (ERIn), respectively. Adoption of this schema allows us conveniently to consider separately the tasks of [A] judging the "relevance" of documents, [B] "retrieving" them, and [C] "reading" them; and to highlight the analogies that may be drawn between, for instance, the actions of retrieving, recommending, citing, linking, and indexing as members of category [B].

Before examining the details of the model, it may be helpful to sketch, as an example, an outline of a familiar scenario to which the model may be applied. Let us imagine that the elements of this scenario are as follows:

(1) an *information seeker*—i.e., a person who desires to inform herself in some way (we might say that she desires to resolve some part of her "anomalous state of knowledge," or to "make sense" of some part of her world; we might also characterize this desire as an *information need*; Dervin & Nilan, 1986);
(2) a *collection of documents*, each of which may at any given time be judged by the information seeker to be related to her information need in some way and to some extent (again, we might say that the documents vary in the extent to which they are capable of satisfying that need, or in the degree to which—in the given context or situation—they are *relevant,* useful, valuable, or worthy of examination; Schamber, 1994); and
(3) an *information system* whose function is to enable the information seeker to locate and read those documents that she *would* judge to be most closely related to her need *if* (which is usually not the case) she were able and willing to examine all the available documents in the collection.

A familiar sequence of events that occurs when an information seeker interacts with an information system runs as follows: (1) the prediction, by the information seeker, that documents having a certain property or set of properties (such as their coverage of a certain topic or subject) would be relevant to her; (2) the specification, by the information seeker, of a query expressing her prediction; (3) the determination, by the system, of the degree to which every document in the collection has the property specified in the query; (4) the presentation, by the system, of a ranked list of pointers to (surrogates for, or representations of) the n documents it judges thus to be most "similar" to the query; (5) the revision, by the information seeker, of her initial prediction, in the light of the new evidence supplied in the form of the list of recommended documents; (6) the expression, by the information seeker, of a modified or expanded query, reformulated to reflect her revised prediction; repetition of steps (3), (4), and (5); (7) the examination, by the information seeker, of those documents that she predicts to be most relevant.

The mere specification of such a sequence of events is, of course, nothing new; rough approximations to the sequence specified here have been common starting points for many of the sophisticated models of retrieval developed in the literature of information retrieval (IR) since the 1950s (see, e.g., Maron & Kuhns, 1960; Robertson, 1977; Salton & McGill, 1983). The distinctive clarification made in the present article is that, in this sequence of events, the actions taken in steps (1), (3), and (5) may be conceived as *evaluative* acts of Type [A]; those in steps (2), (4), and (6) as *recommending* acts of Type [B]; and that in step (7) as an *informative* act of Type [C].

There are many other kinds of information-related act that may be identified as exemplifying one or other of the

[8] I use the inelegant compound "information-related" primarily to signal my concern with a range of activity that is wider than that typically covered by the term "information-*seeking*." Ideally, I would prefer to use "information" itself to refer to a process, just as "communication" is often so used; if it were conventional to think of information as the process of informing (i.e., of instigating a change in a cognitive state, either one's own or that of another), rather than as an object of transfer (e.g., a document), as the content of such an object (e.g., data), or as the meaning of such content ("knowledge"?), it might be easier to accept the phrases "information behavior" and "information activity," if only as tautologies.

three types [A], [B], and [C]; there are many other kinds of sequence of information-related events that may be recognized as cycles of acts of these types. One corollary of the argument developed in this article is that other sequences of information-related events—no less familiar than the one described above, but typically distinguished from it as instances not of information seeking but variously of recommending, citing, indexing, etc.—may be modeled in terms of the same basic categories of activity employed here.

One product of activity of any of these kinds (although of informative activity it happens to be only a secondary product) is assumed to be a *preference ordering* over a set of options (Cohen, Schapire, & Singer, 1999; Wong & Yao, 1990; Yao, 1995). The concept of preference is central to this model, as it is to models of consumer behavior in the theory of microeconomics (see, e.g., Hausman, 1992; Kreps, 1990). Equally fundamentally, activity of all three kinds is treated as activity in which *decisions* are made. Again, it is in economic theory—and in related subfields of economics, statistics, and philosophy, such as decision theory (Luce & Raiffa, 1957; Resnik, 1987), utility theory (Fishburn, 1970), game theory (Von Neumann & Morgenstern, 1944), and social choice theory (Arrow, 1951; Johnson, 1998)—that decision-making activity is typically characterized as activity in which individuals, with given beliefs, seek to satisfy whatever preferences they have for the consequences of their actions.

It might be considered that the ERIn model bears some affinity to *rational choice (RC) theory* (Becker, 1976),[9] because one of the important simplifying assumptions that are made is that people generally act "rationally," where rationality is equated with the instrumental intention to maximize the extent to which their personal preferences are satisfied.[10] Indeed, RC theory is commonly conceived as an approach to the explanation of social behavior that rests on this very assumption, already made explicit in microeconomics and the related fields mentioned above. Although RC theory employs the concepts of rationality and utility in much the same way as, for example, decision theory, it is distinct from the latter in that its (non-normative) aim is to explain and predict the behavior of groups of people; decision theory, in contrast, aims (normatively) to demonstrate how individuals ought to act if they were rational.

Another common thread in the following discussion is the general analogy that may be drawn between document retrieval systems and voting systems (Pennock, Horvitz, & Giles, 2000). Just as a voting system allows a community of voters to make a combined recommendation to policy makers as to the selection of policies or office holders (see, e.g., Dummett, 1984), it will be shown that a retrieval system correspondingly allows a community of judges of document relevance to pool their resources in providing guidance to information seekers making decisions to view or to read.

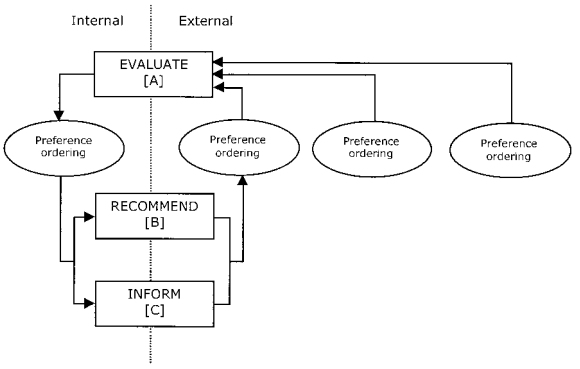

FIG. 1. An RC-theoretic model of information-related activity.

The relationships between the three kinds of information-related activity and their products are depicted in Figure 1; the rest of this section fleshes out the bare bones presented in that diagram.

The RC-Theoretic Foundations

In the rationalist models of decision theory, whose terminology is summarized in Table 2, individual people may be viewed as *agents* or actors (labeled [1] in the Table) in the essential sense that they voluntarily engage in interaction with the physical world and with other agents. Many of the *actions* [2] of humans may be characterized as goal-directed or instrumental, in that we may say that they are carried out in order to bring about some intended change in the prevailing state of affairs. The new state of affairs that is the intended *outcome* (consequence, result) [4] of such an action may be considered the goal of that action. On any given occasion, in any given situation, an agent may choose (i.e., make a selection) from a range of potential actions or courses of action. The members of this set of potential actions may thus be referred to as *options* or alternatives [3], and the decision that is made when resolving to act in a selected manner may be viewed as a choice among the available alternatives.

An agent comes to such a decision through consideration of a set of *beliefs* [5], primarily about the nature of the predicted outcome of each available option, but also about

[9] See Becker (1976) and Coleman (1990) for standard presentations of RCT; Hindess (1988) for a systematic criticism of the views it entails; and Elster (1986) for an overview of the philosophical problems it generates.

[10] "In rational behavior man contemplates the world of potentialities, evaluates them according to his value system, and chooses the best" (Boulding, 1956, p. 26). Formally, an agent is said to act rationally if and only if (a) her preferences are consistent (complete, transitive, and continuous), and (b) she prefers no option to the one she chooses.

TABLE 2. Decision-making activity.

Agent [1]	Decision maker; person; individual; actor
Action [2]	Decision; choice; selection
Option [3]	Alternative
Outcome [4]	State of affairs; consequence; result
Belief [5]	Perception of outcome–criterion–value
Criterion [6]	Attribute of agent–outcome pair; e.g., utility

the relatedness of particular options and particular outcomes, and the relative significance, priority, or relevance, of the different criteria on whose basis competing outcomes are evaluated. Beliefs of the first kind may be expressed as beliefs about the truth of propositions about outcomes; these propositions may in turn be expressed as object–attribute–value triples, where the outcomes are the objects, and the attribute-types may be viewed as the *criteria* [6] upon which decisions are made and actions are chosen. Each instance of a belief may thus be viewed as an agent–outcome–criterion–value quadruple.

For example: Suppose I (the agent [1]) am at home, it is now 8 a.m. on September 24, and I have to attend a meeting at UCLA at 8.30 a.m. There are a variety of routes (i.e., options [3]) I can take on my way to work, and I need to choose among them (this is the action [2]). The time that I arrive at work (i.e., the outcome [4]) will depend on the route that I choose. Let's say that if I take Sunset Boulevard, for instance, I can predict that I will arrive at the office at 8.45; if I take Beverly, on the other hand, I will arrive at 8.25. Clearly, I would prefer to arrive at 8.25 rather than at 8.45: in other words, my belief [5] is that, on the basis of a criterion [6] like usefulness (in allowing me to meet my goal of arriving in time for the meeting), the former outcome has a greater value for me than the latter. My belief may thus be expressed in the form of the quadruple "Jonathan—arrival at 8.25 on September 24—usefulness—1."

Stated in this way, the example is one of decision making under *certainty*, where we are certain that the choice of option x will have the outcome y. Choice theorists are also interested in decision-making under *risk*—where any given option x may have any of a range of outcomes of known probabilities of occurrence—and decision making under *uncertainty*—where the probabilities of occurrence of possible outcomes of option x are not known. To vary the circumstances of the example slightly, we might imagine—more realistically—that there are, in fact, several possible outcomes of traveling to work either via Sunset or via Beverly, whose occurrence depends on the level of traffic encountered. Indeed, in certain circumstances (if there happens to be construction work on Santa Monica, for instance), it might turn out to be faster to travel via Sunset. In this case of decision making under risk, the beliefs [5] that will need to be taken into account if we are to determine whether my choice is a rational one include not only my beliefs about the value to me of certain outcomes, but also about the probabilities with which those outcomes will occur. When we cannot be certain about what the outcomes of given options will be, the decision-making situation may be viewed as if it were a lottery, with outcomes as prizes. As we shall see, document evaluation may be viewed as an instance of this kind of decision making under uncertainty.

It is assumed that consideration of her beliefs will lead the agent to identify her preferred option—the one whose predicted outcome she believes to be "the best" (or in other words, the one whose predicted outcome she likes or desires the most). By extension, in any case in which there are two or more available options, the agent will have a preference ordering—a ranking of all options in order of their predicted outcome's degree of desirability, acceptability, utility, value, or worth for the agent, or the level of approval that the agent has for it. As a concept inherited from the 19th-century utilitarian philosophers, "utility" has often been used in the economic literature to refer to that generic criterion or property of any given agent–outcome pair on which the outcome's position in the agent's personal preference ordering is based.[11]

The agent's motivation to act is assumed to be a general desire to effect the desired change in the prevailing state of affairs, and in so doing to achieve her goal by satisfying her preferences—i.e., by ensuring that her outcome of choice is the one that actually occurs. The existence of factors outside the control of the individual agent, such as the actions of others, may mean that the agent finds it impossible to control events in precisely the manner that she desires, with consequences that are less acceptable. Nevertheless, it is the rationalist assumption of this model that individual agents invariably act to maximize the probability that their preferred outcomes prevail.

Preference orderings may be classified on at least three dimensions, as follows:

(1) *Level of measurement*. If an absolute value, acting as a measure of the degree to which an option is approved, is (or can be) assigned to each option, and a preference ordering is derived from inspection of those values, then we may speak of a *interval* ordering, and (for whatever it may be worth) reliably calculate absolute differences in degree of approval. If, on the other hand, our knowledge is simply of propositions of the form "A is preferred to B" (rather than, e.g., "A is x, B is $x - 1$"), we have a purely *ordinal* ranking.
(2) *Ordinal number*. If there are only two possible ranks in the ordering (e.g., preferred and nonpreferred, approved and nonapproved, top and bottom, 1 and 0), we may speak of a *binary* ordering. If, moreover, there is a requirement that only one option be assigned to the higher of the two ranks, we have a *special binary* ordering. If there are more than two possible ranks, we have a *non-binary* ordering.
(3) *Completeness*. If all options in a given set have been evaluated and ordered, we may speak of a *complete* ordering. If some of the options in the set remain outside the ordering, yet or never to be evaluated, we have an *incomplete* ordering.

[11] In modern economics, however, there is seldom any implication that utility is to be conceived as a property of chosen options; rather, it is seen simply as an index or measure of preference. "When economists say that individuals maximize utility, they are only saying that people do not rank any feasible option above the option they choose . . . An individual who is a utility maximizer just does what he or she most prefers. To say that individuals are utility maximizers says nothing about the nature of their preferences. *All it does is to connect preference and choice* . . . Rational individuals rank available alternatives and *choose* what they most *prefer*" (Hausman, 1992, p. 18; emphasis in original).

TABLE 3. Evaluative activity.

Agent [1]	Prospective recommender or information seeker
Action [2]	Evaluating; ranking; ordering; identifying preferences; weighing the evidence
Option [3]	Document being evaluated
Outcome [4]	State of knowledge after examining document
Belief [5]	Perception of option/outcome–criterion–value
Criterion [6]	Attribute of agent–option/outcome pairs; e.g., relevance, relatedness, quality–intrinsic or extrinsic
Evidence [7]	Source of justification for beliefs; e.g., others' recommendations–first-hand or testimonial
Method [8]	Procedure for synthesizing, combining evidence from multiple sources

In the following three subsections, we shall consider in turn how each of the three kinds of information-related activity in our model may be interpreted in the terms of the decision-theoretic framework outlined above.

Evaluative Activity

The first kind of information-related activity that we shall consider is (what we are calling) evaluative activity. Actions of both other kinds, recommending and informative, are dependent on preference orderings, which in turn, are based on certain beliefs held by the agent. Further, the agent may point to evidence of various kinds in justification of her holding those beliefs. *The action of weighing the available evidence in order to arrive at a preference ordering may be considered as evaluative activity* (see Table 3). Taking a gross input–output view, we may characterize the input to such activity as a set of beliefs about the options under consideration and (perhaps) about the preference orderings supplied by other agents. The output is a personal preference ordering, which in turn, directs the course of subsequent recommending or informative activity.

Formally, to engage in evaluative activity ([2] in Table 3) is for an individual agent[12] [1] to come to hold a set of beliefs [5]—beliefs about the likelihood with which each of the available options [3] is a member of a class specified by a particular agent–outcome[4]–criterion[6]–value quadruple—through consideration of the available evidence [7] that may involve a method [8] of combining evidence from multiple sources.

For example: suppose I (the agent [1]) have a paper to write. It is to be on the topic of the cultivation of cantaloupes, a subject of which I am wholly ignorant. I turn to Google for some inspiration. Google's database contains pointers to a very large number of documents (i.e., options [3]), and I need to choose among them (this is the action [2]). The extent to which I manage to resolve the anomaly in my knowledge of cantaloupes (the outcome [4]) will depend on the documents that I choose to retrieve. It might be considered that my most important belief [5] is that, on the basis of a criterion [6] like usefulness (in allowing me to meet my goal of writing the paper), a knowledge state in which my knowledge of cantaloupes is enhanced is of greater value than one in which my level of ignorance about musk melons remains stable. But my decision making in this situation is *uncertain* because, prior to either examining any of the documents in the database or using Google's recommendations to guide me, I have very little knowledge of the probability with which my choice of any given document will produce the outcome that I desire.

A more helpful interpretation of the beliefs on whose basis a preference ordering is constructed is that they are beliefs about the relationships that exist between the agent and each option (i.e., between the information seeker and each document). Each such relationship is *situational and contextual*, in that it is unique to the given agent–option pair; and it is *temporal and dynamic*, in that it is subject to change over time. In fact, we might choose to define such a relationship by a triple (or even a quadruple) that adds context (or both place and time) to the representation. Moreover, the agent's perceptions of, and beliefs about, the properties of each relationship are *personal and subjective*.

For instance, we may say that to rank [2] a set of documents [3] is for an information seeker, author, indexer, reviewer, or recommender/retrieval system [1] to determine the degree [5] to which each document is "relevant" [6], on the basis of the agent's knowledge [7] of the document and of any prior judgments of relevance. The action in every case consists in evaluating, given a criterion X such as relevance, the degree to which each of the available options is X—i.e., the degree to which each option is a member of the class whose label is specified by a particular agent–outcome–criterion–value quadruple.

The general criterion on which options are evaluated—"relevance"—may then be interpreted simply as the *degree of relatedness* of option to agent, or to a representative of the agent's current context that we may call a *query*. Some instances of an agent–option relationship are more "embedded"—we might say they are more rooted, more intrinsic, more persistent, or less transitory—than others, to the extent that the agent has no need to specify a temporary self-representation or query. There is likely to be, for example, a difference between the *degree of embeddedness* of the relationship between a particular candidate in an election and a voter, and that of the relationship between a particular document in a collection and an information seeker; both are subject to change over time, but the former is less likely to change with the rapidity and regularity that the latter will change in accordance both with changes in the day-to-day interests of the information seeker and with her awareness and usage of the given document. We may refer to criteria that are weakly embedded properties of query–document–context triples, as *extrinsic*—in contrast with those *intrinsic* criteria that are strongly embedded properties of agent–option pairs.

[12] Here "agent" is used to refer either to a person or to the retrieval mechanism of an information system. It is hoped that the ascription of "beliefs" to nonhumans will not seem unjustifiably anthropomorphic.

TABLE 4. Recommending activity.

Agent [1]	Recommender (consumer)		Voter (citizen)	Retrieval system
Action [2]	Rating		Voting	Retrieving
Option [3]	Good; product; service		Candidate	Retrievable document
Expression [4]	Rating		Vote	Retrieved document
Agent [1]	Indexer	Reviewer, editor	Citer (author)	Linker (author)
Action [2]	Indexing	(Peer) reviewing	Citing	Linking
Option [3]	Document	Submitted document	Citable document	Targetable document
Expression [4]	Index term	Accepted document	Citation	Link
Agent [1]	Information seeker	Information seeker	Relevance judge	
Action [2]	Searching	Giving feedback; rating	Making an expert judgment	
Option [3]	Document	Retrieved document	Document	
Expression [4]	Search term	Relevance judgment	Relevance judgment	

The evidence that the agent may point to in justification of her holding her beliefs may be said to vary in terms of the process by which such evidence is produced. Evidence of one kind is that produced *first hand* by direct observation, on the part of the agent, of the objects about which the beliefs are held. Evidence of a different kind, however, is produced by acceptance of the *testimony* of another agent, or of a synthesis of the testimony of multiple other agents.

In our previous examples, the agent has been assumed to be an information seeker. But the ERIn model is intended to cover the related activities not only of information seekers, but also of automated retrieval mechanisms, as well as of indexers, citers, and raters of documents. In a situation in which we may identify a retrieval mechanism as the agent, the testimony on which that mechanism's evaluation of documents is based may be the preference orderings of a single indexer or classifier, of a single citer or linker, or of a single rater (perhaps the current information seeker, providing "relevance feedback"); alternatively, the testimony may include preference orderings of multiple indexers, citers, or raters, or of any combination of those agents engaged in the ERIn process. In these latter cases, the problem for retrieval mechanisms becomes one of appropriately combining, often using sophisticated mathematical methods, multiple sources of evidence of document relevance that are supplied in the form of preference orderings established by the evaluative activity of indexers, citers, and raters.

Recommending Activity

When considering information-related activity in general, we may distinguish, on the basis both of motivation and of outcome, between *informative* activity—as that which is undertaken as an end in itself, directly to satisfy a preference ordering—and *recommending* activity—as that which is undertaken as a means to an end, *to record for subsequent consideration (by oneself or, more commonly, by others) an expression of a preference ordering*. Activity of both kinds results in the expression of a preference ordering: in the first case, however, this expression may be regarded as *implicit*, because the generation of an expression of the ordering is a goal that is secondary or incidental to the satisfaction of the ordering.

In Table 4, the properties of a range of recommending activities are summarized. To engage in recommending activity [2] is for an individual agent [1] to record an expression [4] of a preference ordering over the options [3] in a given set. For instance, to vote [2] is for an individual citizen (the voter) [1] to record an expression (the vote) [4] of a preference ordering over the candidates [3] on a slate. Correspondingly, to "search" [2] is for an information seeker [1] to record an expression (the query) [4] of a preference ordering over the documents [3] in a collection; the query acts as the information seeker's prediction of the properties that are held by documents that she would judge relevant if she had perfect knowledge of the collection.

To "retrieve" [2] is for a retrieval system [1] to present to its user a retrieval set of documents [4] expressive of a preference ordering over the documents [3] in a collection; to cite [2] is for an author (the citer) [1] to record an expression (the citation) [4] of a preference ordering over the documents [3] in a collection; to index [2] is for an indexer [1] to record an expression (in the form of the assignment of an index term) [4] of a preference ordering over the documents [3] in a collection; and so on. Of course, we do not need Table 4 to figure out that searchers giving feedback are essentially engaged in the same activity as the "experts" employed to make relevance judgments on the documents in collections used in experimental tests of retrieval systems. Equally uncontentious is the observation that citing and linking are analogous activities (Borgman & Furner, 2002). Commonalities that are recognized less often, however, are those that connect recommending, retrieving, and indexing. In fact, it may be argued that, in all of these cases, *any expression of a preference ordering may be viewed as a relevance judgment* (whether or not it is commonly referred to as such). The citation or link between a citing or source document and a cited or target document may be viewed as a judgment, made by the citer, of the relevance of the target to the source. The index term as-

signed to a document may be viewed as a judgment, made by the indexer, of the relevance of the index term to the document. The rating assigned to a document may be viewed as a judgment, made by the recommender, of the relevance of the document to the recommender herself.

In a similar sense, it may also be argued that, in all of these cases, *a judgment of relevance is a judgment of approval* (whether or not it is commonly recognized as such). All such judgments are personal decisions made by individuals based on subjective beliefs about (or interpretations or perceptions of) the degree to which each object under consideration is a member of that class of objects which, in some sense that only the judge herself may confirm, are related either directly to the judge or indirectly to a query (a representative of description of the judge's current context). The objective existence of relationships of this kind may, of course, be called into question by other judges; but such argument would be irrelevant because it is the existence of the subjective beliefs about the relationships that is of consequence. Even if we could prove the objective existence of relationships between documents and queries, it is not the case that such proof would logically determine our beliefs as to such existence; and relevance judgments (of all kinds—i.e., citations, assignments of index terms, document ratings) can thus be viewed as nothing more than statements of the form "Right now, I like *X*."

For instance, the decision to create a link between a target and a source document is based on the linker's belief about the degree to which the target document is a member of that class of documents that are "related," possibly in various ways, to the source document that describes the linker's current context. Similarly, the decision to assign a document to an index term is based on the indexer's belief about the degree to which the document is a member of that class of documents that are related to the index term. (The index term may in effect be viewed both as a label for that class of documents and as the indexer's query.) And the decision to assign a document to a positive rating is based on the recommender's belief about the degree to which the document is a member of that class of documents that are related to the positive rating—the class of documents, in other words, that are directly related to the recommender rather than to any intermediating query such as an index term or another document. The approval that is expressed in this last case is not context-free, just as no action is context-free; but we may assume that it is more deeply embedded than that expressed in any situation in which a query is used by the agent as a temporary representative.

Informative Activity

In one distinctively narrow sense, information-*seeking* behavior may be viewed as decision-making behavior of a particular kind, where the decisions that are being made are choices between documents (i.e., information-bearing objects). The primary drawback of pursuing this analogy is that the relationship between the information seeker and the context in which her desire to seek information arises is not explicitly modeled, and thus any understanding we might already have of this central part of the information-seeking process cannot be enhanced. Many would argue that the explanatory power of such a model is crucially hamstrung from the outset; others might explore ways in which the basic model can be extended to provide a more compelling account of the origins and motivational effects of information needs.

Putting such criticism to one side for a moment, we can sketch a RC-theoretic interpretation of information seeking as follows. The information seeker is an agent [1] engaged voluntarily in interaction with physical documents. Her actions [2] are directed towards the goal of effecting desirable change in her personal cognitive state of knowledge. These actions take the form of decisions to read certain documents that are available on given occasions; each such decision may be characterized as a choice from among a large set of options [3] (the documents in the collection); the information seeker perceives, with varying degrees of ease and at varying confidence levels (for this is a situation that the decision theorist would recognize as choice under uncertainty), the degree to which each available option is likely to contribute to bringing about a desirable outcome [4], and implicitly arrives at a preference ordering in which documents are ranked in accordance with their expected value.

An important consequence of this interpretation for designers of information retrieval systems—conceived here as mechanisms that provide automated support of various kinds for the information seeker—is that seekers will best be served by presentations of ranked sets of documents whose orders match those of the seekers' own personal preference orderings. As we saw earlier, the task for the system is thus to predict the particular ordering that the individual seeker *would* specify in the given context, *if* her knowledge of the content of the collection were complete.

Among our categories of information-related activity, information seeking is treated as the archetypal *informative* activity, because its essence is the act of informing oneself through the examination of document content. In a wider context, information seeking may be regarded as a *satisficing* activity, in the sense that the goal of the information seeker is to satisfy (rather than, for instance, to express) their preferences. Analogies between RC-theoretic interpretations of economic consumption, social choice, and information seeking as satisficing activities are summarized in Table 5.

Homo Informaticus

Roberts (1982) describes a model of "primitive information man," implicit in much research in information use studies, that bears certain similarities to *homo oeconomicus* in classical microeconomics. Each is an "improbably simple" abstraction that serves to highlight the fact that the "basic unit of study" in each discipline is the individual

TABLE 5. Satisficing activity.

		Economic consumption	Social choice	Information seeking
Agent [1]	Decision maker; individual; actor	Consumer	Policy maker (government)	Information seeker
Action [2]	Decision; choice; selection	Purchase	Election; legislation	Examination; reading
Option [3]	Alternative	Bundle of goods	Candidate (person; policy)	Document
Outcome [4]	State of affairs; consequence; result	State of economic welfare	State of social welfare	State of cognitive welfare (knowledge)

agent, whose behavior is assumed to be stable over time. Each operates in a narrow, artificial environment in which resources (informative or material) are scarce, and among which the agent is constrained to make choices, many of which involve interaction with other agents. Each has perfect knowledge of the environment; each is motivated to action only by considerations of a certain kind (information need or material need); each employs a simple, general criterion for assessment of performance (basically, "the securing of pleasure and the avoidance of pain," where pleasure is conceived as the satisfaction of desire); and each is assumed to proceed in a "rational" manner (by choosing whatever is most preferred). Precisely as such, "primitive information man" is "a construct of notable information unreality" (Roberts, 1982, p. 101).

A common criticism of RC theories of social behavior is that they fail to explain or to predict, and instead, allow simply for post hoc rationalization of what has already happened as the consequence of individuals' motivations and desires that are taken as given (Hindess, 1988). As Roberts says (1982, p. 98), "It is not unusual to discover data being explained by recourse to behavioural interpretations derived from assumptions held about such behaviour." RC theories typically tell us little about the content, origin, and cause of motivations and desires, and point us away from detailed qualitative studies, both of psychological and of sociological processes, that seek to understand those catalysts for information-related activity. If some sociologists were taken aback in the 1980s by the emergence of RC theory, its thoroughgoing methodological individualism, and its replacement of the Durkheimian model of *homo sociologicus* with that of *homo oeconomicus*, then any attempt to revive *homo informaticus* on the economic model will no doubt be met with equal suspicion.

Roberts points out that *homo oeconomicus* has long been regarded by economists as only a "very first approximation," and that more complex characterizations have emerged as economic explanations have become more sophisticated. "However, classical economic man retains his usefulness at the expository level, allowing of the treatment of the simple before the complex, and for application to certain types of economic analysis" (1982, p. 94). Notwithstanding Roberts's subsequent calls for the construction of more complex versions of *homo informaticus*, it might be considered that the RC-theoretic model of information behavior outlined above does retain "usefulness at the expository level." For the intention is not simply to spell out the qualities of the rationality that is assumed, from the outset, to drive the decision-making activity of information seekers; it is rather to draw, between certain elements of that activity, analogies, and distinctions that will ultimately be of service both in the construction of explanatory models of information need, and in the design of systems that support information seekers' attempts to satisfy needs of that kind. In this way, a descriptive, RC-based theory may be pressed into normative service in library and information science, not as a specification of how people should behave (rationally or otherwise; Ryan, 1991), but as a justification for certain system-design decisions.

I shall return to this theme in a later section. Before doing so, I wish to review the characteristics of document recommender systems that may be identified in the light of the ERIn model.

The Recommender/Retrieval System

As I asserted in the previous section, one result either of recommending or of informative activity is an expression of its underlying preference ordering, which in turn, may contribute to the body of evidence considered in a separate, later, evaluative event. The cyclical nature of this process was illustrated in Figure 1. On that basis, we may understand a document recommender/retrieval system as an agent engaged in evaluative [A] and recommending [B] activity with a view to supporting the evaluative [A] and informative [C] activity of its users. It is a system, operating on behalf of information seekers, that evaluates documents (whether through direct observation or through analysis of testimonials), arrives at preference orderings, and presents those orderings, so that searchers can come to their own relevance judgments and make their own reading decisions on the basis of the evidence supplied by the system.

In passing, it may also be noted that a recommender/retrieval system is analogous to a voting system in the sense that, in designing a system of either kind, specifications need to be made of the following components (inter alia):

(1) An eligibility criterion for agents (i.e., recommenders or voters)—On what basis is it decided which agents are eligible to recommend or vote?
(2) An eligibility criterion for options (i.e., documents or candidates)—On what basis is it decided which options are eligible to be recommended or voted for?
(3) The mechanism by which recommendations/votes are recorded—How do recommenders/voters physically record their recommendations/votes?

TABLE 6. Recommender systems vs. retrieval systems.

	Recommender system	Retrieval system
Collection	Dynamic	Static
Queries	Static	Dynamic
Feedback	Primary	Secondary
Mechanism	Unobtrusive	Obtrusive
Queries	Pre-defined	User-defined
Collaborativeness	High	Low
Personalization	Yes	No
Function	Purchase	Use

(4) The form of expression of preference orderings—Is this interval or ordinal?; binary, special binary, or nonbinary?; complete or incomplete?
(5) The method by which preference orderings are combined—Does it personalize or not?; and if it does, on which matrix manipulation procedure is it based?

Nevertheless, despite the similarity of the concerns of the recommender-system designer and those of retrieval-system designer, the systems are studied by two overlapping but distinct research communities. In practice, the archetypal recommender system differs from the archetypal retrieval system in a number of ways. These differences are summarized in Table 6, and described in more detail in the following subsections. Many of the differences are well understood by researchers: the distinctions between filtering and retrieval, and between implicit and explicit judgments, for instance, are treated often and authoritatively in the literature. The motivation for revisiting these distinctions in the present context is the perception of a need for a comprehensive summary, erected on the foundations provided by an appropriate model of information-related activity (viz., the ERIn model), so that a fully integrated view of recommending activity may be developed and applied to the design of effective information systems.

Filtering versus Retrieval

In the IR community, a distinction is commonly drawn between *retrieval* (a.k.a. "ad hoc" retrieval) and *filtering* (a.k.a. routing) on the basis of consideration of the degree of persistence through time of both (i) the collection of items being accessed and (ii) the queries used to express searchers' requirements (Foltz & Dumais, 1992; Malone, Grant, Turbak, Brobst, & Cohen, 1987).

A collection may be viewed as "static" if items are added to or deleted from it on a "batch" basis at discrete intervals. Given the persistent nature of a static collection, such as a database maintained by a global Web search service or by a library, it can generally be assumed that multiple queries will be addressed to that collection before items are added to or deleted from it; in such a context, where the individual queries addressed to the collection are likely to be different in content and to come from different searchers, those queries may be viewed as "dynamic." This is the familiar process of ad hoc retrieval.

On the other hand, a collection may be viewed as "dynamic" if items are added to or deleted from it in a continuous process, as in the case of a stream of e-mail messages or new journal articles. The queries that are addressed to such a dynamic collection are likely to be relatively "static" in the sense that a searcher may well wish for a single query, representing a semipermanent information need, to be successively compared with every new addition to the collection without requiring repeated intervention from the searcher. A query of this kind acts as a "filter" on incoming messages or on documents being added to the collections. The sense is that irrelevant documents, those not matching the query, are being filtered out (in contrast to a retrieval system that homes in on and picks out the relevant, matching documents). In a scenario where there are multiple users of the system, each characterized by a persistent "user profile" that acts as a personal filter, the sense is that different incoming documents will be "routed" to different people.

It should be clear that the static and dynamic states defined above occupy opposite ends of a continuous spectrum of possibilities, and that retrieval and filtering are thus "two sides of the same coin" (Belkin & Croft, 1992), the similarities between them being more important than the differences. Nevertheless, for many years, the two tasks were addressed by different research communities. Work on recommender systems originated in the domain of information filtering (IF), and, though informed by IR research, it has continued primarily to employ the concepts of the IF community.

Feedback

In the language of the IR system designer, a recommendation—an expression of the preference on the part of a given user on a given occasion for a given document over all others—is simply a relevance judgment—a record of the contextual decision made by the searcher as to the degree to which the given document is relevant. Furthermore, "relevance feedback" (or simply "feedback") is distinguished as any expression of a relevance judgment communicated *by the user to the system* (Spink & Losee, 1996). On the dimension of "extent to which the system's operations are based on the exploitation of feedback," we might imagine, at one pole, systems that make no use of feedback whatsoever, and at the other, systems that conversely make no use of sources of evidence of document relevance *other* than the feedback provided by users.

Many systems that may otherwise be characterized as "traditional," in that their primary source of evidence of document relevance is the similarity of *content* of documents and user-defined queries, also implement mechanisms that elicit and manipulate feedback from individual searchers. An example would be provided by any system that allows the searcher to identify, from among the docu-

ments in a set retrieved in response to an initial query, those that the searcher deems to be relevant, and to communicate such judgments to the system by tagging or marking the favored documents (perhaps by checking a box accompanying each record and pressing a "More like these . . ." button on completion). A number of algorithms have been suggested for the construction, automatically or semiautomatically, of a new or modified (reformulated or expanded) query based on analysis of the content of the documents marked as relevant (and sometimes also of those rejected by the searcher as non-relevant). One automatic (opaque) method of query expansion, for instance, might involve adding to the original query the top *n* most frequently-occurring (or most highly weighted) terms in the documents marked as relevant; a corresponding semiautomatic (transparent, interactive) method might present a list of such terms to the searcher, with a request that she select from the list those terms that she would like to add to the original query.

Given the proliferation of systems that incorporate at least some feedback-based technique, it seems more useful to distinguish between, on the one hand, systems whose exploitation of feedback is secondary to the main content-based process by which relevance is estimated, and on the other, systems in which feedback is the primary source of evidence of document relevance.

Obtrusiveness of Judgment Acquisition Mechanism

A user may communicate to the system a judgment on a document either *implicitly* or *explicitly*. An implicit judgment is communicated in the form of a request or instruction that the system manipulate the document in some way: for example, by transferring an electronic copy of the document to the user's machine for on-screen viewing, offline storage, or hard-copy printout; or by recording an order (which may or may not involve payment) for a printed copy of the document. An explicit judgment is communicated in the expectation that such action will be helpful in the system's subsequent efforts to predict users' personal preference orderings of the documents in the collection. The extent to which this expectation has an altruistic as well as a purely self-interested component will depend partly on the level of collaborativeness of the system (see below) and partly on the individual user's level of awareness that the effectiveness of a collaborative system relies on all users' active participation.

In the case of an implicit judgment, then, whatever the content of the request or instruction, it is likely that the action will be regarded by the user as an end in itself. In the case of an explicit judgment, in contrast, whatever the content of the message, the action is regarded as a means to an end. Moreover, in the implicit case, any treatment by the system of the user's action as evidence of relevance is secondary to the main task, which is to respond to the instruction. We might say that such treatment is a side effect, coincidental, and carried out *en passant*; we might also say that the acquisition or collection of relevance judgments in such a manner is *unobtrusive* (just as the collection of relevance judgments through inspection of the interdocument citations made by document authors is also often said to be unobtrusive). In the explicit case, the function of the judgments is a dedicated one; they are supplied intentionally, explicitly, and specifically for the purpose of helping the system build up a richer body of evidence of relevance, and thus to improve its future approximations to users' preference orderings. The mechanism used to acquire judgments is *obtrusive* in the sense that it intrudes into the user's primary line of activity, the ultimate goal of which is to satisfy an information need, not to rate documents.

We can thus distinguish between systems that require (or, as some would say, allow) users to supply relevance judgments explicitly, and those that are able to collect judgments by drawing inferences from observation of actions of other kinds, prompted by decisions to view, download, print, order, buy, etc. Many systems implement both methods; the archetypal retrieval system can be contrasted with the archetypal recommender system on the basis of its reliance solely on judgments of the explicit kind (if, of course, it exploits feedback at all).

User-Defined versus Predefined Queries

A query may be understood as a set of attribute–value pairs identified by an agent for use as self-representation, and communicated as such to the system. We may distinguish immediately between (a) a representation that has existed for some period of time prior to its selection as a query by the agent, and (b) one that is formulated by the agent herself, at the time of its selection. We may call queries of the first kind predefined (or perhaps "queries-by-example"), and queries of the second kind user-defined. It should be noted that all queries, of either kind, are user-*selected*; the focus on predefinition and user-definition partly captures a related distinction, often made in the literature of human–computer interaction (HCI), between selection-by-recognition and selection-by-recall.

Examples of predefined queries include the following:

(1) *Existing document representations*. For instance, a system might allow the searcher to select an existing document representation as a query by providing a "More like this . . ." or "Find related records" button next to every record listed in a retrieval set; on activation of the button, the system would compare every document in the collection with the selected query document, and produce a ranking of documents in order of document–document coindexing, coapproval, coupledness, cocitedness, or distance, depending on the kind of attribute–value pairs that are used to represent the documents. An example of a different kind is provided by the system that allows the searcher to take advantage of a so-called "dynamic linking" facility by direct selection of terms within currently viewed documents; many systems offer

similar opportunities for direct selection, as query terms, of the labels of classes within a thesaurus or classification scheme.

(2) *Existing agent representations*. For instance, rather than allowing the searcher to make a selection from a range of existing agent representations (which might be considered to raise concerns over agents' personal privacy), a system might make the reasonable assumption that the most useful such representation is that of the searcher herself, and proceed simply by providing, either on a continuous basis or on explicit request, a ranking of documents based on an analysis of agent–agent consistency. Each agent representation will consist of a set of relevance judgments, either inferred from prior usage of the system to view, download, print, order, or purchase documents, or supplied explicitly by the searcher on previous occasions.

Correspondingly, examples of user-defined queries include the following:

(1) *Sets of query terms*. All database searchers will be familiar with the process of formulating queries made up of sets of terms describing the subject matter, or perhaps the form or structure, of documents predicted to be relevant. Such a set of terms may conveniently be treated as if it were a new document, and thus comparable directly with the documents already existing in the collection.

(2) *Sets of relevance judgments*. In any system that implements a feedback mechanism, the searcher may create a query consisting of a set of relevance judgments about, or ratings of, the documents whose surrogates or contents she views. Such a query may then be used either (a) to build a composite document representation consisting of the most highly weighted terms in the most highly rated documents, which may then be compared with the content-based representations of other documents in the collection; or (b) in comparison with the sets of relevance judgments provided by other users of the system, so that documents that come highly rated by those with similar tastes to the present user, but that have not yet been rated by the present user, are retrieved.

Incidental to this discussion, it is interesting to note a distinction of a different but related kind that is often made between "browsing" (or sometimes "navigating") and "searching" (or sometimes "querying"). Typically, the drawing of this distinction results from a placing of particular emphasis on the difference that is perceived to exist between (a) the act of retrieving documents by activating the links, citations, or pointers that have been inserted between related documents, and (b) the act of retrieving documents by specifying a set of keywords against which other sets of keywords (i.e., documents) may be compared. On this reading, browsing is characterized as a query-less process, of retrieval of a succession of single documents, that requires little special knowledge of system or vocabulary on the part of the searcher—to be contrasted with query-based, keyword searching that requires the searcher to pay attention to the selection of terms (values) and possibly of attributes (fields) also.

Although it may sometimes be convenient to make this particular distinction, it seems to conflate several more-fundamental points of differentiation. Here I would argue that searching is the broadest category, that all searching necessarily involves a query of some kind, and that searching, querying, and navigation are most usefully regarded as near-synonyms. As described above, queries may be classified as predefined or user-defined; and the act of following a hypertext link from one (source) document to another (target) document may be considered as a search in which the query (either the source document as a whole, or the link anchor by itself) is predefined, and which results in a weak, binary ranking of the collection in which the retrieved document (the target of the link) is preferred to all others.

One connotation of "browsing" in its ordinary usage is that the browser is exploiting the proximity of objects in space. The principle accepted by the browser is that, if one object is next, close or near to another in spatial distance, then the two objects are more likely to be related in content, form, structure, or whatever feature is being used as a criterion for judgments of relevance. For the browser, proximity or distance is the crucial source of evidence of document relatedness. But in the theory of document network analysis, notions of "space" and "distance" have no meaning other than those that can be applied to certain structural features of graphs; "proximity" describes just one of the ways in which two objects can be interpreted as being similar or related to another, and there seems no a priori reason why "browsing" should not be applied, for instance, to the act of following links between documents related not by proximity but by cocitation or coapproval.

At this point, we might care to reintroduce a distinction between the searching of document (object) space, and the searching of term (class) space. The term "browsing" may then be reserved to refer to the latter, to describe the particular process of retrieving documents through navigation of the links between the labels for classes of documents that comprise a hierarchical thesaurus or classification scheme, regarded as an acyclic graph.

Extent of Collaborativeness

The ugly word "collaborativeness" is used here to preserve the distinction between (a) the act or concept of collaborating (for which we reserve "collaboration"), and (b) the capability of any system to offer an opportunity for its users to collaborate or otherwise to take advantage of collaboration.

The extent of collaborativeness of most recommender systems is high. The most common means by which users of recommender systems may take advantage of the collaboration of others is by taking note of the recommendations made by the system through analysis of the ratings or relevance judgments supplied by other users. Herein lies the

essential difference between the archetypal recommender system and the archetypal retrieval system: stripped of the facility to synthesize multiple users' ratings, the recommender system becomes no more than a retrieval system, albeit one that may implement a sophisticated feedback mechanism. Resnick and Varian (1997) clarify the relationship between the broader term "recommender system" and the narrower term "collaborative filtering system" by pointing out that some systems make recommendations neither on the basis of collaboration, nor by a process of filtering; if we take their assertion at face value, then it would seem that Resnick and Varian see "recommender system" as a synonym for "retrieval system" in the broadest sense (see the section on terminology, above).

Extent of Personalization

At least as revealing a comment as Resnick and Varian's may be made about personalization, to the effect that not all recommender systems make recommendations that are personalized to the current user. For instance: a system that synthesizes multiple users' ratings of documents simply by summing the ratings received by each document and ranking documents in order of rating-sums (normalized or otherwise) is essentially engaged in presenting, in the form of a "hit parade" or "bestseller" list, the results of a popularity contest. The same results would be presented in the same order to any user requesting the same service at the same time, and such a service could not be said to be "personalized" in any sense. Content-based retrieval systems are similarly lacking in a facility for personalization, in the sense that different users selecting the same queries at the same time would ordinarily expect to receive the same responses. In contrast, any system that effectively takes into regular account the particular characteristics, needs, and history of the individual user—as recommender systems commonly do—is highly capable of providing results that, in principle, are highly personalized.

Function

An important historical feature of the work of the recommender-system community is that it has been concerned with the retrieval of objects *of certain kinds*, and the supply of those objects *in certain ways*. The documents that are recommended to system users have been movies, music albums, and books; in cases where the documents are textual, the content has been popular rather than scholarly; where the documents are short, the type has been the email message rather than the journal article. Documents have often been made available for purchase and supplied for payment, rather than being available for loan or supplied for free. Nevertheless, the success of systems like those employed by Amazon.com and CDNOW[13] would not seem to be based on any feature that is necessarily limited to documents (or document-supply mechanisms) of the particular types they offer, but rather on the willingness of users to "make an investment" in their usage of the system by expending time and effort in providing initial ratings. The present time would appear to be ripe for the exploitation of recommender-system technology by on-line library catalogs and other bibliographic retrieval systems that provide access to networks of academic books,[14] scholarly papers, and scientific reports;[15] it should be borne in mind, however, that the performance of such systems is likely to be highly dependent on the level of incentive that is provided to suppliers of ratings.

Other Points of Variance

Other respects, commonly reported in the literature, in which recommender/retrieval systems may vary include the following:

(1) *Scale of ratings*. The ratings supplied by users on the system's invitation may be binary (approval/nonapproval), or on a multipoint Likert scale (e.g., "Rate this item from 1 to 5, where 5 indicates your highest level of approval").
(2) *Persistence of ratings*. The history of an individual's usage of the system, and the ratings that she has supplied, may be stored for analysis only during the current session; alternatively, they may be retained in a database of previous ratings for any length of time beyond the current session.
(3) *Type of correlation used in prediction*. A collaboration-based system may produce personalized rankings of two kinds (Furner, in review). One such ranking is possible once the searcher has selected a query document, which the system can represent as a vector of the approval ratings of that document supplied previously by different judges. In the simplest form of ranking, this query vector may then be compared with every other document vector in the judge–document approval matrix, by which a score indicating degree of similarity is indicated for each query–document pair. A list of the documents in the collection may then be produced in which documents are ranked in order of their computed similarity scores. We might say that top-ranked documents in this list are related to the query document in the sense that "people who approved the query document also approved" the documents with the highest

[13] See http://www.cdnow.com/.

[14] It should be noted that Amazon.com's recommender-based facilities have always supported access to this kind of material.

[15] NEC's experimental "scientific literature digital library," ResearchIndex (formerly CiteSeer; accessible on the web at http://researchindex.org/; see Lawrence, Giles, & Bollacker, 1999), incorporates several recommender-based features in its provision of access to scholarly papers that are available in PDF or Postscript form on the Web. ResearchIndex collects from its users both implicit and explicit ratings of documents viewed; searchers may peruse rankings both of highly popular (i.e., "most accessed") documents, and (with the facility labeled "Users who viewed this document also viewed . . .") of highly coapproved documents.

scores.

An alternative ranking of documents may be produced by a two-step process, in which the searcher is treated as the initial query, and represented by the judgments that they have previously made about documents in the collection. The result of the first step of this process is the computation of a row of the judge–judge consistency matrix, and thus a ranking of judges in order of their similarity to the "query" judge. We might say that top-ranked judges in this list are related to the query judge in the sense that "documents approved by the query judge are also approved by" the judges with the highest scores. The product of the second step is a list of documents ranked in order of the degree to which they are approved by the judges found (in the first step) to be most similar to the query judge. We might say that top-ranked documents in this list are related to the ones already approved by the query judge, or that "people like the query judge also approved" the documents with the highest scores.

Summary

In summary, then, a system identified as a recommender system is typically one that is:

(1) a filtering system, in which the collection is dynamic and queries are static;
(2) feedback-based, relying on its users to provide information about their preferences among items already known to them;
(3) unobtrusive, i.e., able to infer users' preferences from their other activities without requesting them to supply explicit relevance judgments;
(4) "query-free," if only in the sense that users are able to choose from a selection of predefined queries, rather than having to create new ones;
(5) collaborative or social, in that the system provides recommendations on the basis of multiple users' feedback;
(6) personalized, in that recommendations are intended to match the preferences of the individual user;
(7) and e-commercial, in that the items recommended are products or services offered by a business, for purchase (rather than by a service, for use).

However, recommender systems vary in the degree to which they meet each of these "requirements." For example:

(1) the collection need not be particularly dynamic, and the queries may be;
(2) the recommendations made by the system need not be based on prior usage;
(3) the system may require effort on the part of its users that they may consider intrusive on their primary goal;
(4) the system may allow users to create their own queries;
(5) in making recommendations to an individual, the system may use only feedback supplied by that individual;
(6) the system may provide only "average" ratings (e.g., lists of best sellers) that are not personalized; and
(7) the items that it recommends need not be for sale.

At the other end of the spectrum, then, a traditional IR system is a recommender system that typically:

(1) carries out ad hoc retrieval,
(2) with minimal consideration of user feedback,
(3) while requiring much user effort
(4) not least in the formulation of queries,
(5) and taking no account of collaboration
(6) in the provision of the same results to different users
(7) free-of-charge.

My suggestion (not a novel one; see, e.g., Belkin, 2000) is that the historical distinction between retrieval and recommender systems is not as useful as one that we may draw between content-based and collaboration-based mechanisms, both of which we may view as subclasses of a general category of systems that make recommendations of documents to users. In a single system, of course, we may well find content-based and collaboration-based elements. In the next, concluding section, I indicate how a fully integrated view of recommending activity may lead us to an expanded vision of such hybrid systems.

Citation, Indexing, Rating: An Integrated View

We have seen, from Table 4, that the acts of citation,[16] indexing, and rating (inter alia) may be viewed as species of the same genus—recommending. The following commentary is deliberately repetitive to emphasize the commonalities among these activities.

The actions that the author of a document takes when (a) identifying and (b) specifying target documents to cite from the one that she is writing may be considered as [A] evaluative and [B] recommending, respectively. In the former case, she is determining, on each occasion that she feels the need or desire to cite a target document, a binary preference ordering over the collection of citable documents in which at least one document (the one she decides to cite) is preferred to most others of which she is aware. Moreover, the source document's bibliography (i.e., its full list of citations) may be considered in aggregate as a record of the evaluative decisions, made over a period of time by the

[16] The related terms "link" and "linking" are commonly associated, in the scholarly, professional, and popular literatures alike, with hypertext in general, and the World Wide Web in particular. Historically, bibliometricians interested in the explicit expressions of connections between documents have preferred to talk about "citations" and "citing." These latter terms, and the area of study known as "citation analysis," have come to be used in a somewhat narrower sense, in which (a) the documents being linked are typically scholarly papers published in academic journals; (b) the principal mode of distribution of the documents is in hard-copy format; (c) the links take the form of bibliographic references collected in lists at the end of citing documents; and (d) the links are identified by the authors of the citing documents. It is only relatively recently, with the emergence of Web-accessible databases of full-text scholarly papers whose citations are instantly navigable by readers, that researchers have begun in earnest to apply the well-established methods of citation analysis to the electronic environment (Borgman & Furner, 2002).

author, as to the degree of relatedness between the source document and each target document. Given the context represented by the source document (which includes the goal that the author had in writing it, its subject matter, and its manner of creation), the documents that the author chose as targets were preferred by her to the ones that she did not choose. In this sense, her particular choice of target documents is taken as *evidence* of her preferences; her preferences are assumed to be *revealed* by her choice.

Similarly, the actions that an indexer takes when (a) identifying and (b) specifying target documents to which to assign a given index term may be considered as [A] evaluative and [B] recommending, respectively. In the former case, she is determining, on each occasion that she is required to assign a term, a binary preference ordering over the collection of indexable documents in which at least one document (the one to which she decides to assign the term) is preferred to most others of which she is aware. Moreover, the index term's list of postings (i.e., the full list of documents to which it is assigned) may be considered in aggregate as a record of the evaluative decisions, made over a period of time by the indexer, as to the degree of relatedness between the index term and each target document. Given the context represented by the term (which includes its meaning, and its paradigmatic and syntagmatic relations with other terms in the indexing language), the documents that the author chose as targets were preferred by her to the ones that she did not choose. Again, her preferences are assumed to be revealed by her choice.

Finally, the actions that an information seeker takes when (a) identifying and (b) specifying target documents that meet with her approval may also be considered as [A] evaluative and [B] recommending, respectively. In the former case, she is determining, on each occasion that she needs or desires to rate or judge the relevance of a document, a binary preference ordering over the collection of retrievable documents in which at least one document (the one she decides to rate most highly) is preferred to most others of which she is aware. Moreover, the set of judgements made or ratings assigned by the information seeker may be considered in aggregate as a record of the evaluative decisions, made over a period of time by the searcher, as to the degree of relatedness between her and each judged document. In such cases, the context is supplied by the current information need, expressed or otherwise, of the searcher. Once again, her preferences are assumed to be revealed, either by her explicit statement of those preferences or, implicitly, by her decisions to purchase or otherwise manipulate certain documents.

Implications for Design Practice

Although "traditional" IR systems (such as those that provide access to on-line library catalogs, bibliographic databases of scholarly papers, and collections of Web pages) may be viewed as special-purpose recommender systems in the sense outlined earlier, the designers of such systems have been relatively slow to implement the techniques typically associated with the recommender systems used in e-commerce. Specifically, very few systems exist that offer the user the opportunity to take advantage of these techniques in combination with those offered by both content-based *and* context-based systems.

In the preceding discussion, we have developed a distinction between content-based (retrieval) and collaboration-based (recommender) systems; and several proposals have been made of hybrid systems that combine important features of both (Balabanovic & Shoham, 1997; Basu, Hirsh, Cohen, & Nevill-Manning, 2001; Pazzani, 1999). But we may, in fact, distinguish between three basic types of recommender system, each of which uses a different matrix as base data for further manipulation:

(1) *content*-based systems, whose base data is the content of a document–term content matrix (Fairthorne, 1956; Salton & McGill, 1983);
(2) *collaboration*-based systems, based on a judge–document approval matrix (Resnick, Iacovou, Bergstrom, & Riedl, 1994; Soboroff & Nicholas, 2000); and
(3) *context*-based systems (also known as link-analytic or hypertext IR systems), based on a document–document adjacency matrix (Croft & Thompson, 1987; Kleinberg 1999).

For convenience, we may refer to the unifying graph-theoretic model that encompasses systems of each of these three types as the *CCC (content—collaboration—context) model* (Furner, in review). In all of these systems, the base data is manipulated for the same purpose (that is, to produce document rankings), and in the same ways (that is, using the same few, simple matrix operations). Additionally, the effective operation of systems of each kind relies on their being supplied with recommendations—in the form of index terms, approval ratings, and citations—that are generated prior to the information seeker's current interaction with the system. As we saw above, such recommendations may be viewed as different sources of evidence of the preferences of indexers, raters, and citers, and thus of document relevance.

Of course, there would be little practical point either in asserting the equivalence of, or even in drawing an analogy between, certain of the elements that comprise the ERIn and CCC models if by doing so we were unable to contribute to the normative project of deriving principles for the design of systems that help to improve people's access to information. My suggestion, then, is that it would appear to make sense to examine the following hypothesis: that information seekers may benefit from being given the opportunity to exploit multiple sources of evidence of document relevance—i.e., that we may, in certain circumstances, improve the effectiveness of information retrieval by bringing multiple retrieval techniques to bear on single instances of information need.

In the Department of Information Studies at the University of California, Los Angeles, we are currently working

towards the development of a prototype, "proof-of-concept" system that will allow us to test this hypothesis. The prototype is called UCXtra, the idea being that, as the searcher, "you see extra" documents—documents that you might not get to see if you had less control over the retrieval process than this system allows. The principle guiding our design is that information seekers may additionally benefit from being allowed to maintain control over the ways and combinations in which multiple sources of evidence of document relevance are exploited in any given context (Furner, in review).

In summary, then, I began with the observation that the core of any document retrieval system is a mechanism that ranks the documents in a large collection in order of the likelihood with which they match the preferences of any person who interacts with the system. I have argued that, given a broader interpretation of "recommending" than is commonly accepted, such a preference ordering may be viewed as a recommendation, made by the system to the information seeker, that is itself typically derived through synthesis of multiple preference orderings expressed as recommendations by indexers, information seekers, and document authors. I have introduced the ERIn (Evaluation–Recommendation–Information) model, a decision-theoretic framework for understanding information-related activity, that highlights the centrality of recommending in the document retrieval process; and I have demonstrated how the ERIn model may be used to clarify the respects in which indexing, rating, and citation may be considered analogous, as well as to make explicit the point at which content-based, collaboration-based, and context-based (CCC) flavors of document retrieval systems vary. Evaluation of the UCXtra system, whose design is based on principles derived from the ERIn and CCC frameworks, will be reported in subsequent articles.

Acknowledgments

The author would like to thank Marcia Bates, Christine Borgman, Natalie Cole, and the anonymous referees for their helpful comments on an earlier draft of this article, and Shiferaw Assefa, Karen Chiang, and Andra Darlington for bibliographic assistance. The work described in this article is funded in part by a grant from the Academic Senate's Council on Research, University of Califonia, Los Angeles.

References

Arrow, K.J. (1951). Social choice and individual values. New York: John Wiley.

Balabanovic, M., & Shoham, Y. (1997, March). Fab: Content-based, collaborative recommendation. Communications of the ACM, 40(3), 66–72.

Basu, C., Hirsh, H., Cohen, W.W., & Nevill-Manning, C. (2001). Technical paper recommendation: A study in combining multiple information sources. Journal of Artificial Intelligence Research [On-line serial], 14, 231–252. Available: http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume14/basu01a.pdf [September 24, 2001].

Becker, G.C. (1976). The economic approach to human behavior. Chicago: Chicago University Press.

Belkin, N.J. (2000, August). Helping people find what they don't know. Communications of the ACM, 43(8), 58–61.

Belkin, N.J., & Croft, W.B. (1992, December). Information filtering and information retrieval: Two sides of the same coin? Communications of the ACM, 35(12), 29–38.

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), Annual review of information science and technology (vol. 36; pp. 3–72). Medford, NJ: Information Today.

Boulding, K.E. (1956). The image. Ann Arbor, MI: University of Michigan Press.

Cohen, W.W., Schapire, R.E., & Singer, Y. (1999). Learning to order things. Journal of Artificial Intelligence Research, 10, 243–270.

Coleman, J. (1990). Foundations of social theory. Cambridge, MA: Cambridge University Press.

Croft, W.B., & Thompson, R.H. (1987). I^3R: A new approach to the design of document retrieval systems. Journal of the American Society for Information Science, 38, 389–404.

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M.E. Williams (Ed.), Annual review of information science and technology (vol. 21, pp. 3–33). White Plains, NY: Knowledge Industry Publications.

Dummett, M. (1984). Voting procedures. Oxford, UK: Clarendon Press.

Elster, J. (1986). Introduction. In J. Elster (Ed.), Rational choice (pp. 1–33). New York: New York University Press.

Fairthorne, R.A. (1956). The patterns of retrieval. American Documentation, 7, 65–70.

Fishburn, P.C. (1970). Utility theory for decision making. New York: John Wiley.

Foltz, P.W., & Dumais, S.T. (1992, December). Personalized information delivery: An analysis of information filtering methods. Communications of the ACM, 35(12), 51–60.

Furner, J. (in review). Exploiting multiple sources of evidence of document relatedness in hybrid search engines: A unifying model and design proposal. Journal of the American Society for Information Science and Technology.

Goldberg, D., Nichols, D., Oki, B.M., & Terry, D. (1992, December). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61–70.

Hausman, D.M. (1992). The inexact and separate science of economics. Cambridge, UK: Cambridge University Press.

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. In I.R. Katz, R. Mack, L. Marks, M.B. Rosson, & J. Nielsen (Eds.), CHI "'95: Proceedings of the ACM conference on human factors in computing systems, Denver, CO, May 7–11, 1995 (pp. 194–201). New York: ACM Press.

Hindess, B. (1988). Choice, rationality, and social theory. London, UK: Unwin Hyman.

Johnson, P.E. (1998). Social choice: Theory and research. Thousand Oaks, CA: Sage.

Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46, 604–632.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., & Riedl, J. (1997, March). GroupLens: Applying collaborative filtering to Usenet news. Communications of the ACM, 40(3), 77–87.

Kreps, D.M. (1990). A course in microeconomic theory. Hemel Hempstead, UK: Harvester Wheatsheaf.

Lawrence, S., Giles, C.L., & Bollacker, K. (1999, June). Digital libraries and autonomous citation indexing. IEEE Computer, 32(6), 67–71.

Loeb, S., & Terry, D. (Eds.). (1992, December). Information filtering [Special issue]. Communications of the ACM, 35(12).

Luce, R.D., & Raiffa, H. (1957). Games and decisions. New York: John Wiley.

Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A., & Cohen, M.D. (1987). Intelligent information-sharing systems. Communications of the ACM, 30, 390–402.

Maron, M.E., & Kuhns, J.L. (1960). On relevance, probabilistic indexing and information retrieval. Journal of the Association for Computing Machinery, 7, 216–244.

Paepcke, A., Garcia-Molina, H., Rodriguez-Mula, G., & Cho, J. (2000). Beyond document similarity: Understanding value-based search and browsing technologies. SIGMOD Record, 29, 80–92.

Pazzani, M. (1999). A framework for collaborative, content-based, and demographic filtering. Artificial Intelligence Review, 13, 393–408.

Pennock, D.M., Horvitz, E., & Giles, C.L. (2000). Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In AAAI 2000: Proceedings of the 17th National Conference on Artificial Intelligence, Austin, TX, July 30–August 3, 2000 (pp. 729–734). Menlo Park, CA: AAAI Press.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In CSCW'94: Proceedings of the conference on computer supported cooperative work, Chapel Hill, NC, October 22–26, 1994 (pp. 175–186). New York: ACM Press.

Resnick, P., & Varian, H.R. (Eds.). (1997, March). Recommender systems [Special issue]. Communications of the ACM, 40(3).

Resnik, M.D. (1987). Choices: An introduction to decision theory. Minneapolis, MN: University of Minnesota Press.

Roberts, N. (1982). A search for information man. Social Science Information Studies, 2, 93–104.

Robertson, S.E. (1977). The probability ranking principle in IR. Journal of Documentation, 33, 294–304.

Ryan, A. (1991, October 10). When it's rational to be irrational. New York Review of Books, 38(16), 19–22.

Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.

Schafer, J.B., Konstan, J., & Riedl, J. (2001). E-commerce recommendation applications. Data Mining and Knowledge Discovery, 5, 115–153.

Schamber, L. (1994). Relevance and information behavior. In M.E. Williams (Ed.), Annual review of information science and technology (vol. 29, pp. 3–48). Medford, NJ: Learned Information.

Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating word of mouth. In I.R. Katz, R. Mack, L. Marks, M.B. Rosson, & J. Nielsen (Eds.), CHI '95: Proceedings of the ACM conference on human factors in computing systems, Denver, CO, May 7–11, 1995 (pp. 210–217). New York: ACM Press.

Soboroff, I., & Nicholas, C. (2000). Collaborative filtering and the generalized vector space model. In N.J. Belkin, P. Ingwersen, & M.-K. Leong (Eds.), SIGIR 2000: Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval, Athens, Greece, July 24–28, 2000 (pp. 351–353). New York: ACM Press.

Spink, A., & Losee, R.M. (1996). Feedback in information retrieval. In M.E. Williams (Ed.), Annual review of information science and technology (vol. 31, pp. 33–78). Medford, NJ: Information Today.

Varian, H. (1996, March 18, 1996). "Collaborative filtering" name game: How about "recommenders"? [Electronic mailing list]. Available: http://www.sims.berkeley.edu/resources/mailing-lists/collab/0001.html [August 20, 2001].

Von Neumann, J., & Morgenstern, O. (1944). Theory of games and economic behavior. Princeton, NJ: Princeton University Press.

Wong, S.K.M., & Yao, Y.Y. (1990). Query formulation in linear retrieval models. Journal of the American Society for Information Science, 41, 334–341.

Yao, Y.Y. (1995). Measuring retrieval effectiveness based on user preference of documents. Journal of the American Society for Information Science, 46, 133–145.