ELSEVIER

Information Processing and Management 40 (2004) 1–19

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

Relevance as resonance: a new theoretical perspective and a practical utilization in information filtering

Christophe Brouard [a,*], Jian-Yun Nie [b]

[a] *Equipe MRIM, Laboratoire CLIPS-IMAG, B.P. 53, Grenoble cedex 9 38041, France*
[b] *Dept. d' Informatique et Recherche Opérationnelle (DIRO), Université de Montréal, C.P. 6128, Succursale CENTRE-VILLE, Montreal, Qué., Canada H3C 3J7*

Received 30 April 2002; accepted 8 August 2003

Abstract

This paper presents a new adaptive filtering system called RELIEFS. This system is based on neural mechanisms underlying an information selection process. It is inspired from the cognitive model adaptive resonance theory [Biol. Cybernet. 23 (1976) 121] that proposes a neural explanation of how our brain selects information from its environment. In our approach, resonance, the key idea of this model is used to model the notion of relevance in information retrieval and information filtering (IF). The comparison of resonance with the previous models of relevance shows that resonance captures the very core of most existing models. Moreover, the notion of resonance provides a new angle to look at relevance and opens new theoretical perspectives. The proposed mechanism based on resonance has been directly implemented and tested on the TREC-9 and TREC-11 IF data. The experimental results show that this approach can result in a high effectiveness in practice.

Keywords: Information retrieval; Relevance; Resonance; Cognitive science; Information filtering

1. Introduction

The goal of an information retrieval (IR) system is to find *relevant* information to a user's information need. In order to implement an IR system, several fundamental questions are raised: How can we judge if a document is relevant or not? What are the characteristics of a relevant document? What is the relation between a user's information need and a document relevant to this

[*] Corresponding author. Tel.: +334-7663-5855; fax: +334-7644-6675.
E-mail address: christophe.brouard@imag.fr (C. Brouard).

0306-4573/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2003.08.004

need? Immediately, "relevance" appears as the central notion of IR and the need for its formalization arises. [1]

A large number of studies have been devoted to this notion, from the early stage of IR (Saracevic, 1975) till now (JASIS, 1994; Mizzaro, 1997, 1998; Saracevic, 1996; Schamber, 1994). Still we are far from reaching a formal and clear definition of relevance (Froehlich, 1994). The difficulty comes in part from the situational and the dynamic aspects of relevance (Schamber, Eisenberg, & Nilan, 1990). That is, relevance depends on the situation in which it is judged and it changes in time.

It comes also from the existence of different points of view on this notion. Indeed, a recurrent debate consists in opposing the system and the user points of view. When a mathematical formalization of relevance is proposed without integrating explicitly the user's point of view (e.g., criteria beyond-topical aspects, psychological mechanisms of relevance estimation and so on), a legitimate question concerns whether this formalization is coherent with the user's point of view. We should not ignore the fact that this information is selected for the user, and the way it is selected by the system should be consistent with the way the user would select it. Of course, an external evaluation of the formalization can be used to compare the system responses to the user relevance judgments. However, this external evaluation is done for a particular corpus in a particular situation. It does not provide a direct examination of formalization itself. It is then difficult to claim for general theoretical quality of the proposed formalization from such an evaluation.

For these reasons, we believe that it could be beneficial to adopt a cognitive approach and to consider "natural" mechanisms, that is to say, psychological or neural mechanisms responsible for information selection. Thus, in this paper, we take into account cognitive studies which deal with semantic memory (knowledge organization et knowledge access) or with attention (selection of information from our environment). We consider the models developed in these studies as sources of inspiration for the IR systems. Especially, we propose a formulation of relevance by adapting the notion of resonance developed in adaptive resonance theory (ART) (Grossberg, 1976). This notion is simple and intuitive. As we will show, it captures the essence of most previous formalisms for relevance.

ART provides a formal theory to model cognitive activities in human. This theory has been used to model different cognitive processes, from perception, attention, to categorization (Grossberg, 1999a). The resonance mechanism is able to simulate the way that certain information from our environment is selected whereas others are not. For example, the ART provides an interesting explanation on why it is possible to follow a conversation in a noisy environment—some sound signals relevant to the conversation can be selected while the background noise is not (Grossberg, 1999b).

The notion of resonance may be applied to IR as a possible formulation of relevance. This paper will offer a comparison of the two notions. It turns out that not only there is a strong relationship between relevance and resonance, but also several formulations of relevance in IR can be reformulated in terms of resonance. This suggests that resonance may be a natural for-

[1] The situation in information filtering is similar. The user wants the system to keep only the relevant documents and reject the irrelevant ones among a flow of incoming documents. The key problem is also that of determining the relevance of a document.

mulation of relevance in IR. Moreover, this new consideration of relevance sheds new light on the meaning of relevance and opens new theoretical perspective.

In addition of a theoretical comparison, this formulation is further tested in practice in an information filtering (IF) task. Our system resulted in a quite good performance. This further shows that the idea of resonance is usable in practice.

This paper is organized as follows. In the next section, we will first describe the ART and the underlying concept of resonance. In the third section, it will be compared with several formulations of relevance in IR. Then, in the following section, an adaptation of resonance will be proposed to IR and IF. We will describe the experimental results obtained in TREC-9 and TREC-11 tasks. Finally, a general discussion carries on the theoretical perspectives of this work.

2. Resonance—a new formulation of relevance

This section deals with the formulation of relevance. We propose to bridge a gap between relevance and resonance. First, we give a brief description of the ART in which the notion of resonance in a network has been originally proposed. Next, we describe different formulations of relevance in IR literature, and we emphasize the two implications between document and query. Finally, we describe more precisely the analogy between resonance and the two implications.

2.1. Principle of adaptive resonance theory

This theory was initiated to cope with the interaction of a system with its environment. It uses a situated approach (Clancey, 1997) by which one tries to understand the cognitive system within the world in which it evolves. The key idea in this approach is the following coupling: A system S takes into account the information from its environment E while applying its own knowledge for the selection of this information (Fig. 1). In Fig. 1, the interaction between S and E is represented as two arrows. The arrow from E to S (E → S) represents the impact of outside information on the cognitive system (it is called a bottom-up process). The second arrow (E → S) represents the application of the cognitive system expectations to the environment (called a top-down process).

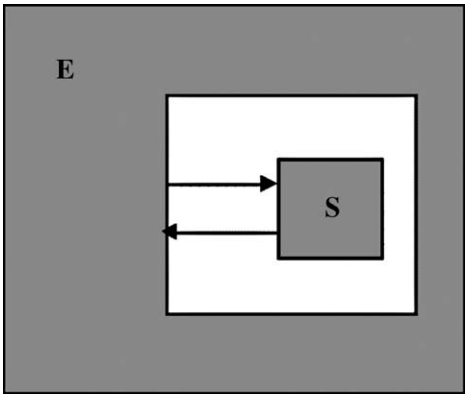

Fig. 1. Illustration of the coupling of a system (S) with its environment (E) (from Clancey, 1997, p. 283).

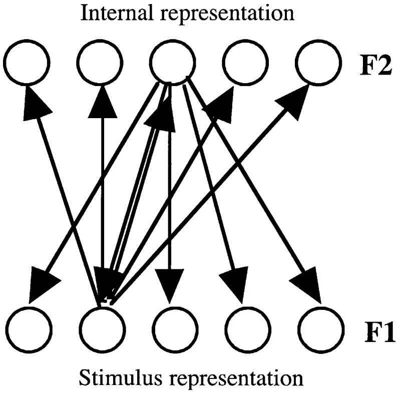

Fig. 2. ART neural network architecture. Layer F1 is used to represent stimuli. Layer F2 is used to represent internal representations. The connection between nodes of F1 and F2 are oriented.

ART is implemented as a two-layer neural network (Fig. 2). The entry layer F1 allows us to represent the stimuli from the environment. The other layer F2 corresponds to internal representations of the system. Connections between F1 and F2 are oriented. So, between a node in F1 and a node in F2, one may have a connection leading from the former to the latter and another one in reverse direction. As in all connectionist approaches, learning in ART consists of updating weights of the connections. In ART, this update is based on an association principle: if an F1 node and an F2 node are activated simultaneously, the connection between them is reinforced.

Without going into details, the computation of the reinforcement results in a measure of the strength of the following assertion "if the node A is activated then the node B is also activated". So, globally, the connection weights can be considered as an evaluation of the entailment strength of information B (represented by the incoming node of the connection) by information A (represented by the outcoming node of the connection). [2]

Considering this architecture, when an external stimulus arrives, it activates some of the nodes in the layer F1. In other words, the stimulus is represented by an activation pattern in F1. This activation is then spread to F2 according to the weights of the top-down connections, and then after internal competition at F2, [3] back-propagated to F1 according to the weights of the bottom-up connections. If the back-propagation matches well the original activation of F1, (i.e., a strong back-propagation to a node in F1 that is originally strongly activated), then the activation will spread again on the same nodes in F2 and so on. The network locks into a resonant state. In summary, resonance is a state in which two nodes strongly activate each other. The notion of resonance as it appears in the neural network, is not disconnected to the general meaning. It always refers to the back-propagation of a signal to its source involving the amplification or the prolongation of the source signal. The echo of a sound is a concrete illustration of this phenomenon.

[2] See Smolensky (1987) for a discussion on the relation between computed connection weights in neural network and logical implication.

[3] The F2 layer is a competitive layer. Every node in this layer has inhibiting connections to the other nodes. As a result, only the node with the largest input has an output.

This resonant state can or cannot occur depending on the incoming stimulus. In ART, the resonance mechanism is presented as a mechanism of information selection. The stimulus for which emerges a resonant state is taken into account for learning (connection weights are updated). In the contrary case, the system does not pay attention to it or process it in a different way. A vigilance parameter is used to determine how much mismatch will be tolerated (mismatch means that the activation received by F2 is not totally back-propagated to the nodes in F1 originally activated).

The resonance mechanism has been used to explain different phenomena. For example, Grossberg (1999b) uses this resonant processing to explain how we can keep track of our conversation with a friend in a crowded noisy room even though the sound emitted by the voice of the friend may be overlapped by the sounds emitted by other speakers (this is called the cocktail party problem). In the model of Grossberg (called ARTSTREAM), each time a sound arises, its different frequency components are redundantly represented by multiple neurons (in F1). Each neuron is associated with a particular voice and the problem is to amplify the activation of the neurons corresponding to the "good" voice and inhibit the other ones. Without going into details, this problem is resolved in ART by using an abstract representation (neurons in an F2 layer) of the friend's voice. The representation of this voice is more activated than the other ones if the frequencies of the emitted sound correspond well to the friend's voice. It will win the competition with the abstract representation of the other voices (in F2) and will act (back-propagation) so as to amplify the activation of the neurons associated with the friend's voice (in F1).

More globally, Grossberg gives an evolutionist interpretation of this mechanism. He argues that resonance allows humans to preserve the knowledge previously acquired, and simultaneously continue to learn (i.e., modifying some previous knowledge). This is what Grossberg calls the stability/plasticity dilemma (Grossberg, 1999a). The connections from F1 to F2 (the bottom-up pathways) allow the system to take the environment into account. If they are removed, then the system will be unable to react to the outside world. On the other hand, the connections from F2 to F1 (the top-down pathways) allow for the application of the knowledge of the system to the case under consideration. Without them, the system will not be able to use the knowledge it accumulated previously. The selection by resonance takes into account both aspects simultaneously. A new stimulus will have an impact on the learning process only if it matches well an internal representation. This rule makes the system relatively stable (it depends on how the mismatch is tolerated) while being able to learn. Thus, resonance appears as an information selection mechanism and it is justified from a more global perspective of the interaction between the system and its environment.

2.2. Formulating relevance in IR

The study of relevance refers to the identification of the relation associating an entity considered to be relevant (e.g., a document *D*) to a requirement entity (e.g., an information need specified as a query *Q*). Prior to any study about this relation, one has to question about the features of the entities that have to be taken into account in this relation.

According to the way relevance is studied, a distinction can be made between two groups. The first group tries to define directly the relevance relation between the two entities. For example, Green and Bean (Green, 1995; Green & Bean, 1995) analyze the relation between the topics of a thematic guide and the referenced passages in that guide, and try to define the relevance relation from it. This group of studies usually is restricted to the topics of the entities, i.e., the aspects beyond-topical aspects are not taken into account. The second group tries to define the evaluation criteria instead, that is, the entity's features that have to be taken into account. Numerous studies have been carried out on evaluation criteria. The results of these studies often overlap (Barry & Schamber, 1998). For example, Barry (1994) interviewed a group of users about their judgments of relevance, and analyzed their answers. The study extracted seven different categories of criteria involved in user's relevance judgments: the contents of the document, the user's experience (including his knowledge), the user's belief and preference, the information coming from other sources (e.g., consensus and external verification), the source of the document (the quality of the source), the document as a physical object (the cost and the ease to obtain the document), and the context in which the user judges relevance (e.g., the time constraint).

Our study is related to the identification of the relation and can be consequently classified in the first group. Among the studies belonging to the first group, beside the kind of study mentioned above (Green, 1995; Green & Bean, 1995), there are also a number of attempts trying to create formal formulations of relevance. Logic and probability are the two general frameworks in which proposals have been made.

2.2.1. The logical approach

The logical framework has a long tradition in IR. One of the most used models—Boolean model, is in fact based on Boolean logic. In this framework, relevance is modeled by the logical implication "$D \rightarrow Q$" (see (Lalmas, 1997), for an overview). In his paper, Cooper (1971) already stressed the close link between relevance and logical consequence:

"One might, on first reflection, suppose that logical consequence could be distantly related to relevance, but it will be argued shortly that when problem is posed in terms of declarative sentences, logical consequence and relevance are very intimately connected" ((Cooper, 1971), p. 22).

The logical formulation is further developed with the formulation of van Rijsbergen (1986). In his formalization, relevance relation is modeled as an uncertain logical implication between document contents D and the information need (the query) Q, expressed as "$D \rightarrow Q$". If this logical implication is valid, then we say that the document is relevant to the query. This formulation includes the classical cases of the Boolean model. For example, if a document talking about A, B and C is relevant to a query looking for A and B, i.e., the logical implication A&B&C → A&B is valid. It corresponds to the fact that the document satisfies the requirement. As pointed out in Nie (1989), the degree of satisfaction of the requirement by the document contents is only one of the two important aspects of relevance. Another important aspect is whether the document contents are exclusively related to the requirement. If the requirement is satisfied by two documents, we would certainly prefer the one whose contents are exclusively related to the requirement, rather than the one of which only part of the contents is related. This second aspect can be formulated by the reverse implication "$Q \rightarrow D$", i.e., the degree to which all the contents of the document are within the scope of the query. It is then suggested that the degree of relevance should be expressed as a function F of both aspects: $F(P(D \rightarrow Q), P(Q \rightarrow D))$ where P is a certain evaluation function of the uncertainty of implications. The criterion "$D \rightarrow Q$" determines the exhaustivity of the document to the requirement of the query, i.e., whether the entire requirement is satisfied. The second criterion "$Q \rightarrow D$" is related to the specificity of the document to the requirement of the query. While the logical aspects are still formulated as "$D \rightarrow Q$" and "$Q \rightarrow D$", Dempster–

Shafer theory is used in Lalmas and Ruthven (1998) to cope with uncertainty, and logical imaging is used in Crestani and van Rijsbergen (1995). Nie and Brisebois (1996) use a fuzzy modal logic in which the calculation of the uncertainty is carried out in a fuzzy logic framework.

2.2.2. The probabilistic approach

A large number of methods have been proposed for the calculation of the uncertainty of relevance. As a particular application of the uncertainty principle van Rijsbergen (1986) suggests to evaluate $P(D \rightarrow Q)$ as the conditional probability $P(Q|D)$ instead of considering "$D \rightarrow Q$" as material implication. [4] The conditional probabilities are also commonly used in the probabilistic models (see (Crestani, Lalmas, Rijsbergen, & Campbell, 1998), for an overview). The two aspects can also be found in most probabilistic models, formulated as $P(Q|D)$ and $P(D|Q)$. For example, Turtle and Croft (1991) propose to calculate $P(Q|D)$ based on a Bayesian network in order to estimate relevance. In the Binary Independent Retrieval Model, the two main elements that one tries to determine are the conditional probabilities $P(D|\mathrm{rel}_Q)$ and $P(D|n\mathrm{rel}_Q)$, i.e., the probability that D is part of the relevant and irrelevant documents for the query Q. In addition, D is further discomposed into a set of occurrences and no occurrences of independent terms, that we denote here by x_{ti} ($x_{ti} = 1$ if the term t_i occurs in the document, and $x_{ti} = 0$ if it does not). The calculation of $P(D|\mathrm{rel}_Q)$ is approximated by $\prod x_{ti \in D} P(x_{ti}|\mathrm{rel}_Q)$ by assuming that all the terms are orthogonal (independence assumption). The aim of the model is in fact to determine the most important feature terms for a particular query Q. This is the role of $P(x_{ti}|\mathrm{rel}_Q)$. Notice further that rel_Q can be equivalently denoted as Q. So globally, one tries to estimate $P(D|Q)$ and $P(D|\neg Q)$. We are very close to the logical formulation of relevance, i.e., "$Q \rightarrow D$".

2.2.3. The neural network approach

Some researchers have suggested to use a connectionist approach for implementing probabilistic models (Kwok, 1995). In this case the computations are the same as those for probabilistic models. They are simply presented as spreading activation from nodes representing queries, query terms to nodes representing documents terms and documents or spreading activation in the reverse direction. Other networks proposed by Belew (1989), Boughanem, Christment, and Soulé-Dupuy (1999) or Wilkinson and Hingston (1991) are still very close to traditional models. A back-propagation (which corresponds to the second implication) is studied and tested by Wilkinson and Hingston (1991).

2.2.4. Synthesis

The models we have mentioned incorporate both aspects of specificity and exhaustivity. In fact, these aspects are important in the very basic operation of indexing; we are always concerned with the specificity of a term to a document, and the exhaustivity of the terms to the document contents. Both aspects are taken into account in term weighting. For example, the commonly used tf $*$ idf schema (Baeza-Yates & Ribeiro-Neto, 1999, p. 29) contains exactly these two aspects: the term frequency (tf) is related to the exhaustivity, while the inversed document frequency (idf) concerns the specificity.

[4] The material implication $D \supset Q$ is evaluated as $P(D \supset Q) = P(\neg D \vee Q)$.

The large variation of calculation shows that there is no agreement on the way to calculate the uncertainty degree of the logical implications. It is premature to take one of the calculations as a desired one. However, this analysis of the previous studies of IR modeling shows that in the current stage, an appropriate IR model would be the one that takes into account both logical aspects "$D \rightarrow Q$" and "$Q \rightarrow D$", and where the calculation of their uncertainties can be done in any appropriate formalism.

2.3. Analogy between relevance and resonance

Although the two notions have been developed for different purposes, we can observe a strong similarity between relevance and resonance. Let us first consider the two layers of the ART neural network shown in Fig. 2 and assume an idealized internal state of relevance for each query. This state corresponds to an internal representation of the system, i.e., it is a node in F2 layer. In IF, the stimuli at F1 layer are the coming documents. In our implementation, each node of F1 represents a word. A document is represented by an activation pattern of F1 (the nodes which correspond to the document words are activated). The problem now is to estimate the relationship between nodes in F1 and nodes in F2.

Consider now the connections between the nodes in the two layers. As we mentioned before, it is possible to associate the connection weights with the entailment strength of an information B by another information A. B is associated with the outcoming node and A is associated with the incoming node. Thus, if we associate each node in F1 with a word and each node in F2 with an idealized internal state of relevance of a particular query, each connection from F1 to F2 represents a rule which means "if this word is present then the document is relevant to the query". The higher the weight, the stronger the rule, therefore, the stronger the activation level of the outcoming node. This shows that the spreading activation from F1 to F2 (i.e., the bottom-up process) corresponds to the evaluation of the implication "$D \rightarrow Q$". As to the connection from F2 to F1, each connection represents a rule which means "if the document is relevant then this word occurs". So, the back-propagation from F2 to F1 can be associated with a measure of "$Q \rightarrow D$" implication.

Now, let us consider that a document is represented by an activation pattern at layer F1. If this activation is strongly propagated toward the internal relevance representation of a query (i.e., a strong evaluation of "$D \rightarrow Q$"), then the document is considered to be a potentially relevant document. If the activation of the relevance node further strongly back-activates the document nodes (i.e., a strong evaluation of "$Q \rightarrow D$") we arrive at a resonant state. In this case, we can say that the relevance of the document is confirmed. It is equivalent to the fact that both aspects "$D \rightarrow Q$" and "$Q \rightarrow D$" are satisfied at a high level of certainty. As we can see, the notion of resonance is coherent with the formulation of relevance by "$D \rightarrow Q$" and "$Q \rightarrow D$", and it provides a natural and very simple mechanism for the estimation of relevance.

3. Implementation of RELIEFS

We have experimented the general principles of ART in a textual document filtering task. Given a flow of documents, this task consists in selecting those which are relevant to a particular query and ignore the others. We especially considered the adaptive filtering task in which the system can

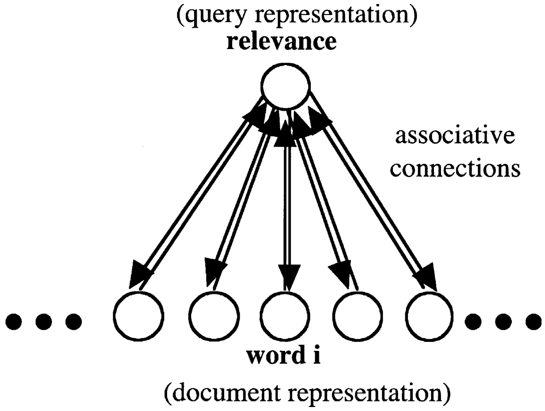

Fig. 3. Built and updated network.

learn from the user relevance judgments given for each selected document. Note that the system RELIEFS (which stands for RELevance Information Extraction Fuzzy System) is not exactly an implementation of ART and must not be considered as a neural network. The analogy with the neural process is much less developed. The accent is put on conceptual aspects.

3.1. General architecture

The ART principles for IR can directly be implemented as follows: An internal representation node is created in the system for each query, representing the perfect relevance state; each node in F1 represents a word; and a document corresponds to an activation pattern. The association between a relevance node and terms are weighted. If we isolate one particular query, the network that our system will construct and update is as shown in Fig. 3, where the relevance node is for this particular query, and the set of associated terms are those that are considered to be related to the requirement of the query (we will describe how these terms are selected later).

The utilization of such a network is as follows: When a document arrives, a set of term nodes corresponding to the words occurring in the document will be activated. Some of these words act as good indicators of relevance. A good indicator is the one such that it creates a good level of resonance with the relevance node. The global relevance level of the whole document will be calculated according to the resonance of all the terms with the relevance node. This is the utilization of the network for relevance estimation. On the other hand, there is also an important aspect of network updating. Once a relevance estimation is made, it is compared with the judgment of the user. A large change between them will ignite a large change in the connections between the terms and the relevance node in both directions. We will describe these two processes in the following subsections.

3.2. The estimation of document relevance

Relevance is estimated as resonance. The degree of resonance between a document and a query is calculated according to the level of resonance of the terms included in the document and the relevance node. The resonance between a term node i and the relevance node is calculated as the multiplication of the degree of implications in the two directions, i.e., $W_{iR} * W_{Ri}$ where W_{iR} is the weight (in the [0..1] range) of the connection from the term i to the relevance node, and W_{Ri} the weight (in the [0..1] range) of the connection in the opposite direction. This can be interpreted in terms of activation. If the word i is present, the activation of the node representing the word i is 1. The activation propagated to the relevance node considering the weight of the connection is obtained by multiplying the activation by the connection weight: $1 * W_{iR} = W_{iR}$. The back-propagated activation is obtained similarly. So we obtain $W_{iR} * W_{Ri}$. One can observe here that we assume the activation level of a word node to be always 1. This is a simplification. In fact, the activation level of a word node can vary from 0 to 1. However, the purpose of the present study is to test whether the principle of resonance applies to IF. Simplifications have been made for the auxiliary aspects. Note also that the choice of multiplication is also related to the nature of the relationship between the two implications. The underlying relation is a logical "and", i.e., we want both implications to be satisfied. In particular, we observe that if a resonant term is absent from a document, then the resonance level of that document with the relevance node will be largely affected. This reflects the intuition that the absence of an activation relay leads to the dispersion of the activation, and will not lead to the emergence of a resonant state.

The global resonance level between a document and the relevance node is determined by the sum of resonance levels between each word node and the relevance node, i.e., $\sum_i W_{iR} * W_{Ri}$ for every term i in the document. This value is normalized as follows:

$$R(d,q) = \frac{\sum_i W_{Ri} * W_{iR}}{\sum_j W_{Rj} * W_{jR}} \tag{1}$$

where i is a term present in the document, and j represents any term connected to the relevance node (including those absent in the document).

The product $W_{iR} * W_{Ri}$ has an intuitive interpretation. In fact, an important term (a term that affects significantly the relevance estimation when it occurs) is a term which allows predicting relevance (when the term is present, generally, the document is relevant) and which is also frequently present in relevant documents (i.e., when the document is relevant, generally, the term is present). Thus, among those terms with high W_{iR}, W_{Ri} selects those which do not appear randomly in relevant documents. These terms will likely be present in other relevant documents. So, they are useful for future judgments. When we use $R(d,q)$ in IF, we have also to determine a threshold T of $R(d,q)$ to decide whether there is a resonant state. This is similar to ART. In our system used for IF, a document is kept if and only if $R(d,q) > T$. The determination of T is an important issue in practice. We will discuss about it in the experimental part.

Here, we are far from the way that resonance is computed in ART. Indeed, we consider each term independently and not globally and no competition in F2 layer is considered.

3.3. Updating rule

Once a network is created, the connections are assigned with an initial weight which may not be the most appropriate (i.e., reflects well the user's relevance judgment). Therefore, we need to update the connection with the user's relevance feedback. The connections between the terms and the relevance node are updated when the user provides a relevance judgment. In what follows, we present our incremental method to update the connection from A to B. A and B correspond respectively to the relevance and a term, or the reverse.

Suppose that $W_{\mathrm{AB}(k-1)}$ and $W_{\mathrm{AB}(k)}$ are the weights associated to the connection from A to B after the $(k-1)$th and kth observations respectively. An observation here is a pair of document and relevance judgment. Let us denote it as O_k. Let $\mu_{\mathrm{A}}(O_k)$ denote the presence or absence of the object A in O_k (1 if present and 0 if absent). The learning rule we adopt is as follows:

$$W_{\mathrm{AB}(k)} = \frac{\alpha W_{\mathrm{AB}(k-1)} + \mu_{\mathrm{A}}(O_k) * \mu_{\mathrm{B}}(O_k)}{\alpha + \mu_{\mathrm{A}}(O_k)} \quad \text{where } \alpha = \sum_{i=1}^{k-1} \mu_{\mathrm{A}}(O_k) \tag{2}$$

This rule works according to the associative principle as follows:

First, let us consider that A is a term and B corresponds to relevance.

- If both A and B are present in the observation, i.e., the term occurs in a document judged relevant, the connection from A to B and that in the reverse direction are reinforced. This case corresponds to a term present in a document judged relevant.
- If A is present in the observation but B is not (i.e., A is a term which occurs in the non-relevant document) the weight of the connection from A to B is reduced, while the connection from B to A remains the same. This means that we will reduce the weight assigned to the term that appears in an irrelevant document.

Now, let us consider that A corresponds to relevance and B is a term

- If both A and B are present in the observation, it corresponds to the same case as above.
- If A is present in the observation but B is not, this means that we will reduce the weight of a term that does not appear in a relevant document.

The numerator of the rule corresponds precisely to the general form of Hebb rule (Hebb, 1949). This rule is based on the association principle: if two elements are simultaneously activated, then the connection between them is increased. The denominator is a normalization factor. It is easy to show that this particular normalization factor leads the obtained weight to be that of $P(\mathrm{B}|\mathrm{A})$, or the relative frequency of the presence of B among all the cases where A appears. In fact, suppose that there have been n observations, and the information A is present at least once in these observations (otherwise the connection does not exist), W_{AB_n} corresponds to the following relative frequency of B given A:

$$W_{\mathrm{AB}_n} = \frac{\sum_{i=1}^{n} \mu_{\mathrm{A}}(O_i) * \mu_{\mathrm{B}}(O_i)}{\sum_{i=1}^{n} \mu_{\mathrm{A}}(O_i)} \tag{3}$$

This result can be proven by induction (see Appendix A).

4. Experiments

The above approach has been used for information filtering task (Fig. 4).

IF aims to determine if an incoming document is relevant or not to a fixed information need (or profile). In an adaptive filtering task, the system can benefit from user's relevance judgments for

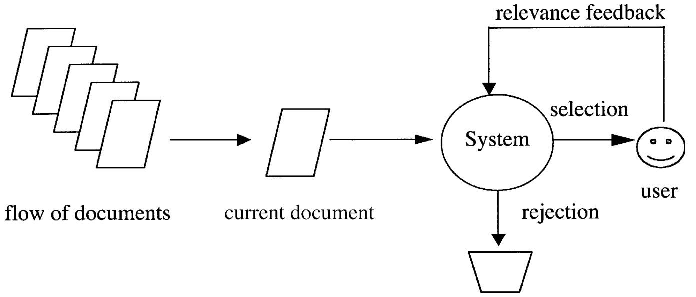

Fig. 4. The adaptive filtering task. The system receives a flow of documents. For each new document it has to decide to select the document or to reject it. In the first case (selection), a feedback is given by the user (yes this document is relevant or no this document is not relevant). This feedback can be used by the system to improve its future judgments.

the updating of weights. The adaptive filtering is the best framework to experiment our approach due to the existence of relevance judgments so that our updating process can take place. Note that this updating process changes the representation of the information need whereas the information need is supposed to be constant. The experiments are done in TREC-9 conference (Brouard & Nie, 2000). A small number of relevance judgments available for each information need (only two in our case) are used for an initial learning. The other judgments can be used only if the system decides to keep the document (i.e., judges it as relevant).

The documents used in these experiments are a subset (OHSUMED) of the MEDLINE collection in medicine. The documents for the tests are written in the period 1988–1991. There are about 300,000 documents in the test collection. Only two relevant documents written in 1987 are provided per query for the purpose of training of the system before the test. Sixty three topics (specifications of information need) are used. One topic was for example: "Are there adverse effects on lipids when progesterone is given with estrogen replacement?"

In order to evaluate the effectiveness of each system, we make use of the following utility measure: $T9U = 2 * R - N$, where R is the number of relevant documents selected, and N is the number of irrelevant documents selected. Two points are gained for each relevant document selected and one point penalty is applied for each irrelevant document selected. The goal of learning is to optimize the global utility (the average) for all the 63 queries.

4.1. Adapting RELIEFS to the filtering task

The principles described earlier correspond to the essence of the RELIEFS system. However, for this particular filtering task in TREC, several practical adjustments have to be made. In particular, we have to adjust the threshold of the resonance score to determine whether a document should be selected or not. Another module has to be added for the selection of important terms/words in the document.

4.1.1. Determination of the threshold

The queries are very different from each other. Therefore, a different dynamic threshold is determined for each query. The value of the threshold is determined empirically. The initial threshold is set according to the average of the scores of the two documents given for training (precisely, it is $0.7 * \text{average score}$). We used the following strategy to adjust the value:

- If a selected document is irrelevant, the system is considered to be too tolerant. The threshold is increased.
- If a document (that we do not know if it is relevant or not) is not selected, the system is considered to be too strict. Then the threshold is decreased. The increase value (we chose 0.1) is set to be higher than the decrease value (we chose 0.00001) because there are much more unselected documents than the selected ones. The decrease of threshold in the second case is due to the fact that we do not want the system to remain silent for a too long period. This allows to gradually correct an initial threshold that is fixed too high. In both cases, we considered different criteria for the modification of the threshold, including:
    - The number of irrelevant documents that are selected consecutively: The higher this number, the larger the increase value and the smaller the decrease value.
    - The number of consecutive relevant documents: The higher this number, the larger the decrease value. This change only concerns the decrease case.
    - The number of documents considered: The larger this number, the lower the change value. The intuition behind this criterion is that we would make larger changes at the beginning of the filtering. When a certain number of documents have been treated, the system should stabilize.

We also considered a more global criterion on the probability of relevance. Considering the utility measure, if the probability of relevance of a document is higher than 0.33, then the document should be selected; otherwise, it should not be. Indeed, we can see that the selected documents with a probability of relevance of 0.33 give an average score of zero with our specific utility measure. This selection criterion is optimal for the utility measure. The optimization process tries to update the threshold such that the frequency of relevant document is 0.33. For example if for a particular score S, the frequency of relevant documents is larger than 0.33, the documents with a score larger or equal to S are selected.

4.1.2. Selection of words in the documents

Usually, document contents are represented by a relatively large set of weighted keywords. The expected advantage of a large set of keywords is a good coverage of the contents. This seems to be true in IR, where the goal is to create an appropriate order among the documents in the response list. The relevance estimation can be based on a comparison of the contents of all the documents with query keywords chosen by a user. At the contrary, in IF, the keywords are automatically added and weighted. Then, the possible risk with a large set of keyword selected is that a keyword which is not related to the required topic could be associated with it by learning and so generates mistakes in future selections. Our selection allows us to keep only the ones which are likely to be related to the query. Therefore, we will only allow (as the most part of other adaptive filtering system) at most N terms to represent the incoming document. The selection of N terms is based on the resonance of the term with the relevance node. If less than N terms have been selected in this step, the remaining terms will be selected among the terms related to the query terms. This second selection makes use of a thesaurus that is constructed automatically (see "query expansion"

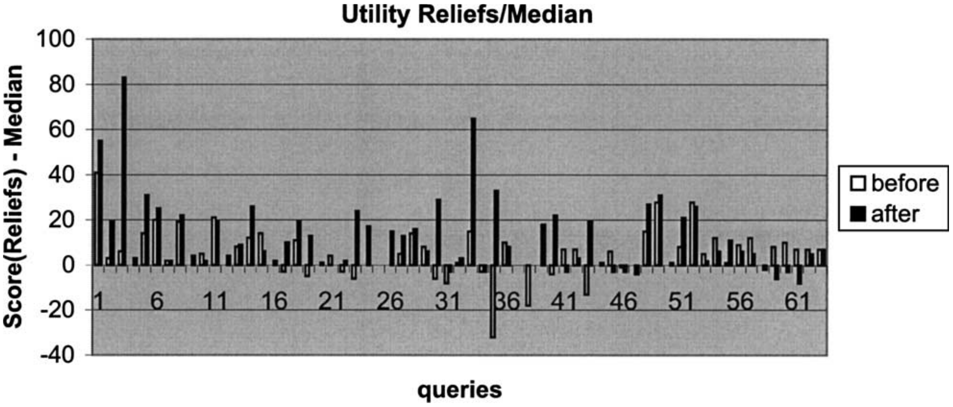

Fig. 5. Comparison of the two versions (before and after improvements) with the median utility value of TREC.

section) from the training set of documents (those in 1987). If the number of terms is still less than N, then the first terms/words read from the document are added to complete the N terms.

4.2. First experimental results

Our approach worked quite well for TREC-9. Two sets of results have been submitted to TREC for official evaluation, one set for each utility measure. In this series of experiments, we set $N = 20$ which was the value for which the results were good (even if we did not really optimize it). The result for the utility measure compares favorably to those of the other groups. In fact, the utility scores for about 80% of the queries obtained by our system are higher than the median score [5] (60% higher and 20% equal). In addition, our system is one of the five that resulted in positive value for the global utility scores for all the queries. The score of our system is +1.1.

Later on, we changed the denominator in formula (1) to consider only the 50 best terms, instead of all the terms connected to the relevance node. This is done in order to avoid considering the terms which are not really related to the query, although they do appear in relevant documents. Better results have been obtained: a utility score of +8.3 (with a precision of 0.29 and a recall of 0.24). This effectiveness is very close to the best system participating in the TREC. The two sets of results are shown in Fig. 5, in comparison with the median score. From these first tests, we see that the principle of resonance works well in practice.

4.3. Varying the importance of the implications

In our previous tests, both implications are considered with the same importance: they are multiplied. Our question is, should they play equal role in the estimation or should one factor be more important than another? In order to answer this question, we vary the importance of one of the implication by replacing the weight W_{Ri} associated to the link from term i to the relevance node

[5] The median score is the score with equal numbers of scores above and below it in the ordered list of scores.

Table 1
The impact of ρ on the utility

ρ	Utility	ρ	Utility	ρ	Utility	ρ	Utility
0.0	2.00	0.5	6.78	1.0	8.19	1.5	5.59
0.1	4.43	0.6	8.30	1.1	7.25	1.6	5.09
0.2	5.20	0.7	8.00	1.2	5.79	1.7	5.51
0.3	6.57	0.8	7.51	1.3	6.79	1.8	4.83
0.4	6.80	0.9	7.49	1.4	6.12	1.9	1.97

by $(W_{iR})^{\rho}$. The factor ρ changes the relative importance of this implication. The relevance estimate is defined as follows:

$$R(d,q) = \frac{\sum_i W_{Ri} * (W_{iR})^{\rho}}{\sum_j W_{Rj} * (W_{jR})^{\rho}}$$

The coefficient ρ has a direct impact on the stability and the plasticity of the system. The larger is ρ, the more adaptive is the system, since we favour the bottom-up process. If ρ equals 0, the score is only determined by W_{Ri}, i.e., the presence or absence of terms which occurred in a relevant documents even if they occurred in many irrelevant ones. The system does not take into account the relevance feedback given by the user if these terms occurred in irrelevant documents. This tends to make the system only look at the original terms found in the first relevant documents and this does not favour changes in the system.

If ρ tends to infinity, the score is only determined by W_{iR}, i.e., the presence or absence of terms which occurred very few time but in a relevant document. This tends to make the system too reactive to new relevance feedback. Other experiments will be done to test more precisely the impact of ρ on the system behavior. Nevertheless, in our case, it turns out that the best results obtain when the two aspects are balanced (ρ in [0.6, 1.1]) (Table 1). Similar results have been obtained with an other corpus (Brouard, 2002).

4.4. Query expansion

When less than N required representing terms have been selected for a query, an expansion process is carried out in order to complete the N terms. For this, we make use of a thesaurus to select the additional terms that are strongly associated to the original query terms.

The thesaurus is constructed automatically from the training documents written in 1987. This thesaurus is simply a network in which nodes represent words and oriented connections are the relative frequency of the outcoming term B given the incoming term A, i.e., W_{AB}. This thesaurus has been built using the learning rule (2). The construction of this thesaurus is similar to the approach used in Miyamoto (1990) and Radecki (1976) for fuzzy thesauri. The relationship between terms is also evaluated according to the resonance score between them. A term is considered to relate to another one if the implication relationships between them in both directions are strong. In our expansion process, instead of considering the resonance level between each term and one term in the query, we consider the resonance level between the term and all the query terms. This method allows us to select the terms that are strongly related to the whole query. This

Table 2
Results with and without thesaurus

ρ	With thesaurus	Without thesaurus
0.6	8.30	7.12
0.7	8.00	6.56
0.8	7.51	7.03
0.9	7.49	7.06
1.0	8.19	6.76
1.1	7.25	6.14

approach has been proven more effective than the one that consider pairwise relationships between terms (Qiu & Frei, 1993). Once queries have been expanded, we observe some improvements in the global utility, as we can see in the following table (the column "without thesaurus" means no expansion is carried out if less than N terms are associated to a query) (Table 2).

This series of experiments show that the RELIEFS system performs well in practice. This is an additional indication that our modeling approach is reasonable.

Another important fact to observe is that the principle of resonance can be implemented as it is, without heavy simplifications as this is usually the case for other models.

5. General discussion

In this paper, we suggested a new way to model relevance by using resonance proposed in ART. We showed that there is a very strong relation between relevance and resonance. Resonance has been developed in cognitive science in order to cope with the learning process of a person or a system, as well as the interactions between it and its environment. This idea seems to fit well to IR. In fact, each IR system has its own internal representation of relevance, even if this is not always explicitly expressed. In most IR systems once such an internal representation is created (e.g., a relevance estimation function is defined), it remains unchanged. The results in cognitive studies reject this approach. In fact, a person or a system is always in contact with his/its environment, and this constantly changes the internal representation of him/it.

Moreover, this approach opens new philosophical perspectives on the notion of relevance. Indeed, if we identify relevance with resonance, relevance becomes a way to select information which warranties to the system equilibrium between plasticity and stability and consequently a good evolution. This explicitly expresses the relation between relevance and interaction often mentioned. Our modeling approach is consistent with the previous approaches that draw a parallel between relevance and communication (Sperber & Wilson, 1995) since communication can be seen as a kind of interaction. For example, Saracevic (1975, p. 321) states: "In the most fundamental sense, relevance has to do with effectiveness of communication."

There have been several attempts to integrate interactions in IR process. For several years, TREC has included an interactive track to study the way that the system interacts with the user (Voorhees & Harman, 2000). Bruce (1994) is another example of such attempts. Denos (1998) proposed an IR system based on an iterative interaction between the user and the system. Our approach continues in the same direction.

An important point of our approach is that it integrates a neuronal mechanism (resonance) for which neurological and psychological datas confirm its plausibility as the principle of information selection in human cognitive system. In that sense, studies that tried to use semantic memory model based on spreading activation (Crestani, 1997) are similar. Although, we refer to low level notions such as activation, spreading activation, association and resonance, we draw a parallel between these notions and the higher level notions such as relevance, specificity, exhaustivity and adapting. In the previous studies of relevance, there was often a clear separation between cognitive/theoretical modeling and practical implementation. Few results from the former have been actually implemented without sever simplifications in a system. The present study tries to conciliate both theoretical and practical aspects. Our experiments showed that the principle of relevance as resonance can be directly implemented and it works well in practice.

We have also to point out that the current implementation and experiments are limited from several points of view. For example, only one updating rule has been examined. It would be possible to further analyze the rule and compare it with other possible forms. The limit of N representing terms is rather a practical setting. We do not know clearly the impact of this setting on the global effectiveness. Some further experiments are required. Despite these limits, our results show clearly that resonance corresponds well to the notion of relevance in IR, and it also results in a good effectiveness in practice. It seems a promising direction to pursue.

Appendix A. The proof of formula (3)

Consider the first observation O_k such that $\mu_A(O_k) \neq 0$. In this case, for all the observations prior to k, we have: $\forall i < k$, $\mu_A(O_i) = 0$. Therefore, $\alpha = 0$.

If we apply (2) we can see that the property is true for the first observation:

$$W_{AB(k)} = \frac{\mu_A(O_k) * \mu_B(O_k)}{\mu_A(O_k)} = \frac{\sum_{i=1}^{k} \mu_A(O_i) * \mu_B(O_i)}{\sum_{i=1}^{k} \mu_A(O_i)}$$

Now, we consider the recurrence hypothesis:

$$W_{AB(n)} = \frac{\sum_{i=1}^{n} \mu_A(O_i) * \mu_B(O_i)}{\sum_{i=1}^{n} \mu_A(O_i)}$$

We apply formula (2) to obtain:

$$W_{AB(n+1)} = \frac{\alpha * W_{AB_n} + \mu_A(O_{(n+1)}) * \mu_B(O_{(n+1)})}{\alpha + \mu_A(O_{(n+1)})}$$

which is equal to:

$$W_{AB(n+1)} = \frac{\sum_{i=1}^{n} \mu_A(O_i) * \frac{\sum_{i=1}^{n} \mu_A(O_i) * \mu_B(O_i)}{\sum_{i=1}^{n} \mu_A(O_i)} + \mu_A(O_{(n+1)}) * \mu_B(O_{(n+1)})}{\sum_{i=1}^{n} \mu_A(O_i) + \mu_A(O_{(n+1)})}$$

$$W_{AB(n+1)} = \frac{\sum_{i=1}^{n+1} \mu_A(O_i) * \mu_B(O_i)}{\sum_{i=1}^{n+1} \mu_A(O_i)}$$

References

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Wokingham, UK: Addison-Wesley.

Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science, 45*(3), 149–159.

Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing and Management, 34*(2/3), 219–236.

Belew, R. K. (1989). Adaptive Information Retrieval. In *Proceedings of the 12th international conference on reseacrch and development in information retrieval (SIGIR), Cambridge, Massachusetts* (pp. 11–20).

Boughanem, M., Christment, C., & Soulé-Dupuy, C. (1999). Query modification based on relevance back-propagation in an ad hoc environment. *Information Processing and Management, 35*(2), 121–139.

Brouard, C., & Nie, J.-Y. (2000). The system RELIEFS: a new approach for information filtering. In *Proceedings of the ninth text retrieval conference (TREC-9), Washington D.C.* (pp. 573–578).

Brouard, C. (2002). CLIPS at TREC11: experiments in filtering. In *Proceedings of the 11th text retrieval conference (TREC-11), Washington D.C.*

Bruce, H. B. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science, 45*(3), 142–148.

Clancey, W. (1997). *Situated cognition*. Cambridge: Cambridge University Press.

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

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review, 11*, 453–498.

Crestani, F., Lalmas, M., Rijsbergen, C. J., & Campbell, I. (1998). Is this document relevant?...Probably: a survey of probabilistic models in information retrieval. *ACM Computing Surveys, 30*(4), 528–552.

Crestani, F., & van Rijsbergen, C. J. (1995). Information retrieval by logical imaging. *Journal of Documentation, 51*(1), 1–15.

Denos, N. (1998). A user-oriented framework for multidimensional querying of an image retrieval system. In *Proceedings of the world automation congress (WAC'98-IFMIP'98), Alaska.*

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

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

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

Grossberg, S. (1976). Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*, 121–134.

Grossberg, S. (1999a). The link between brain learning, attention and consciousness. *Consciousness and Cognition, 8*, 1–44.

Grossberg, S. (1999b). Pitch-based streaming in auditory perception. In N. Griffith, & P. Todd (Eds.), *Musical networks: parallel distributed perception and performance* (pp. 117–140). Cambridge, MA: MIT Press.

Hebb, D. O. (1949). *The organisation of behaviour*. New York: John Wiley and Sons Inc.

JASIS (1994). Special issue on relevance research. *Journal of the American Society for Information Science, 45*(3), 124–217.

Kwok, K. L. (1995). A network approach to probabilistic information retrieval. *ACM Transactions on Information Systems, 13*(3), 324–353.

Lalmas, M. (1997). Logical models in information retrieval: Introduction and overview. *Information Processing and Management, 34*(1), 19–33.

Lalmas, M., & Ruthven, I. (1998). Representing and retrieving structured documents using the Dempster–Shafer theory of evidence: modelling and evaluation. *Journal of Documentation, 54*(5), 529–565.

Miyamoto, S. (1990). Information retrieval based on fuzzy associations. *Fuzzy Sets and Systems, 38*, 191–205.

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

Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting with Computers, 10*(3), 305–322.

Nie, J. Y. (1989). An information retrieval model based on modal logic. *Information Processing and Management, 25*(5), 477–491.

Nie, J. Y., & Brisebois, M. (1996). An inferential approach to information retrieval and its implementation using a manual thesaurus. *Artificial Intelligence Review, 10*, 409–439.

Qiu, Y., & Frei, P. H. (1993). Concept based query expansion. In *Proceedings of the 16th ACM international conference on research and development in information retrieval (SIGIR), Pittsburgh* (pp. 160–169).

Radecki, T. (1976). Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. *Information Processing and Management, 12*, 313–318.

Saracevic, T. (1975). Relevance: a review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science*, 321–343.

Saracevic, T. (1996). Relevance reconsidered. In *Proceedings of the second conference on conceptions of library and information science (CoLIS 2), Copenhagen* (pp. 201–218).

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

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing and Management, 26*(6), 755–776.

Smolensky, P. (1987). Connectionist IA, symbolic IA and the brain. *Artificial Intelligence Review, 1*, 95–109.

Sperber, D., & Wilson, D. (1995). *Relevance communication and cognition* (second ed.). Oxford: Blackwell.

Turtle, H., & Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems, 9*, 187–222.

van Rijsbergen, C. J. (1986). A non-classical logic for information retrieval. *The Computer Journal, 29*(6), 481–485.

Voorhees, E. M., & Harman, D. K. (2000). NIST Special Publication 500-249: The Ninth Text Retrieval Conference (TREC-9).

Wilkinson, R., & Hingston, P. (1991). Using the cosine measure in neural network for document retrieval. In *Proceedings of the 14th annual international conference on research and development in information retrieval (SIGIR)*, (pp. 202–210).