Towards a Belief-Revision-Based Adaptive and Context-Sensitive Information Retrieval System

RAYMOND Y. K. LAU
City University of Hong Kong
PETER D. BRUZA
Queensland University of Technology
and
DAWEI SONG
Knowledge Media Institute, The Open University

8

In an adaptive information retrieval (IR) setting, the information seekers' beliefs about which terms are relevant or nonrelevant will naturally fluctuate. This article investigates how the theory of belief revision can be used to model adaptive IR. More specifically, belief revision logic provides a rich representation scheme to formalize retrieval contexts so as to disambiguate vague user queries. In addition, belief revision theory underpins the development of an effective mechanism to revise user profiles in accordance with information seekers' changing information needs. It is argued that information retrieval contexts can be extracted by means of the information-flow text mining method so as to realize a highly autonomous adaptive IR system. The extra bonus of a belief-based IR model is that its retrieval behavior is more predictable and explanatory. Our initial experiments show that the belief-based adaptive IR system is as effective as a classical adaptive IR system. To our best knowledge, this is the first successful implementation and evaluation of a logic-based adaptive IR model which can efficiently process large IR collections.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models; selection process*; H.3.4 [**Information Storage and

The work reported in this article has been funded in part by the Co-operative Research Centre for Enterprise Distributed Systems Technology (DSTC) through the Australian Federal Government's CRC Programme (Department of Education, Science and Training). This work is also funded in part by the UK's Engineering and Physical Sciences Research Council (EPSRC), Grant number EP/E002145/1 and Overseas Travel Grant number EP/E037402/1.
Authors' addresses: R. Y. K. Lau, Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR; email: raylau@cityu.edu.hk; P. D. Bruza, School of Information Systems, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia; email: p.bruza@qut.edu.au; D. Song, Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK; email: d.song@open.ac.uk.

Retrieval]: Systems and Software—*User profiles and alert services; performance evaluation (efficiency and effectiveness)*; H.3.m [**Information Storage and Retrieval**]: Miscellaneous

General Terms: Theory, Algorithms, Experimentation

Additional Key Words and Phrases: Belief revision, retrieval context, information flow, text mining, adaptive information retrieval

ACM Reference Format:

Lau, R. Y. K., Bruza, P. D., and Song, D. 2008. Towards a belief-revision-based adaptive and context-sensitive information retrieval system. ACM Trans. Inf. Syst. 26, 2, Article 8 (March 2008), 38 pages. DOI = 10.1145/1344411.1344414 http://doi.acm.org/10.1145/1344411.1344414

1. INTRODUCTION

Distributed computer-based information systems have undergone huge growth in recent years. The growing availability of large, dynamic, heterogeneous, and distributed sources of information via the World Wide Web (i.e., Web) has contributed greatly to "information overload" [Maes 1994; Levy 2005]. Accordingly, there is a growing demand for the development of highly autonomous and adaptive information retrieval (IR) systems which can automatically select relevant information items on behalf of their users.

Figure 1 highlights the main functional components of an adaptive IR system. An information seeker first translates her implicit information needs into queries. Recurring queries are often stored in a user profile. On the other hand, information objects from specific information sources such as the Web are characterized by a particular indexing scheme. These characterizations are also stored in the local cache of the adaptive IR system. The matching mechanism of the adaptive IR system tries to match the user's information needs with incoming information objects by comparing the corresponding queries and document characterizations. Information objects deemed relevant by the adaptive IR system are dispatched to the user in the form of a retrieval result set. After reviewing the information objects, the user can then provide relevance feedback to the adaptive IR system. The learning mechanism of the adaptive IR system will use this feedback information to revise and refine the initial user profile. As a user's information needs as well as the underlying retrieval context may change over time, the adaptive IR system should continuously revise its user profile based on the user's most recent relevance feedback so as to maintain the effectiveness of the information matching processes.

In general, information search is divided into information retrieval (IR) and information filtering (IF), although they share some common activities [Belkin and Croft 1992]. IR often refers to the situation that an information seeker takes an active role to specify his ad hoc queries, whereas IF is concerned with the removal of irrelevant information from an incoming stream of information, based on the information seeker's long-term and recurring retrieval goals stored in a user profile. As the proposed adaptive IR system can support both ad hoc interactive retrieval tasks and long-term recurring retrieval tasks, the more general term "IR" is applied to our prototype system.

IR involves uncertainty, both in terms of query and document representation. By way of illustration, given the query "Java", an IR system may return

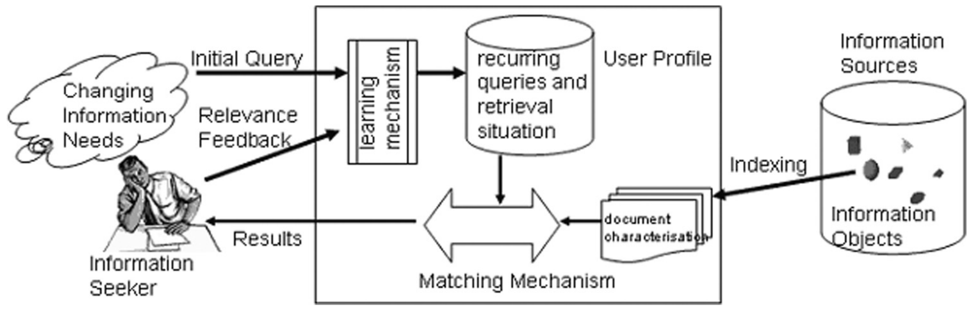

Fig. 1. An overview of an adaptive information retrieval system.

documents about "Computer Programming" or about "Merapi", a volcano on the island of Java. If the information seeker is a computer programmer (i.e., in the context of information technology), documents about Java programming are relevant. However, if the information seeker is a volcanologist (i.e., in the context of volcanology), documents about "Merapi" are more likely to be considered relevant. From the perspective of the IR system, there is uncertainty in determining which set of documents is relevant because the queries captured in the user profile are often incomplete and the implicit retrieval context may not be readily available. In fact, the issues of *partiality* and *uncertainty* are inherent in any IR processes [van Rijsbergen 1986; Lalmas and Bruza 1998], and we believe that these issues contribute significantly to information overload.

One of the justifications for the development of a belief-revision-based adaptive IR system is that the expressive power of logic is believed to be able to model most of the fundamental aspects in information retrieval [Chiaramella and Chevallet 1992; Lalmas and Bruza 1998; van Rijsbergen 1986; Sebastiani 1998]. Previous psychological study has shown that the postulates characterizing preferential logic are compatible with the characteristics of human reasoning [Neves et al. 2000]. It has also been reported that possibilistic, rather than probabilistic, reasoning is closer to the kind of approximate reasoning exercised by human experts [Raufaste and Neves 1998]. The AGM belief revision logic which underpins the proposed adaptive IR model has close connection with preferential logic and possibilistic logic, and hence shows promise in modeling the human approximate reasoning processes in IR.

The logical uncertainty principle [van Rijsbergen 1986], which is a generalization of the Ramsey test [Gärdenfors 1988], has spawned fruitful theoretical investigations into logic-based approaches for IR. The logical uncertainty principle states the following.

> "Given any two sentences x and y, a measure of the uncertainty of $x \rightarrow y$ relative to a given dataset is determined by the minimal extent to which we have to add information to the dataset to establish the truth of $x \rightarrow y$."

With a logic-based adaptive IR model, sentence x can be taken as the representation of an information seeker's needs, and sentence y is the characterization of a document [Nie 1986]. In addition, a dataset can be interpreted as a retrieval context which characterizes a particular information matching situation [Nie et al. 1995]. A retrieval context may consist of an information seeker's background, her long-term search goals, tasks at hand, knowledge about a retrieval domain, etc. The subject of context has received a great deal of attention in the field of information retrieval [Cool 2001; Cool and Spink 2002; Cohen and Singer 1996; Lawrence 2000; Nie et al. 1995]. Recently, there has been a series of ACM SIGIR workshops (IRiX) which examine the prominent features of contexts (e.g., task, document, environment, etc.) and applying these features to improve IR [Ingwersen and Järvelin 2005]. Belief revision can be taken as a means of directly implementing the logical uncertainty principle for adaptive IR because of its close connection with the Ramsey test [Gärdenfors 1988]. More specifically, a retrieval context is formally represented by a belief set K and the changing retrieval context is modeled by the corresponding belief revision function K^*_α [Alchourrón et al. 1985], where α is the logical representation of a user's relevance feedback. Then, a document d can be evaluated with respect to the refined retrieval context K and the query q. Such a matching process is underpinned by $q \mid\sim d$, where $\mid\sim$ is the expectation inference relation [Gärdenfors and Makinson 1994], a class of nonmonotonic inference relations. As a matter of fact, the idea of applying belief revision and nonmonotonic reasoning to practical applications has also been explored in other application domains [Bessant et al.1998].

One contribution of this article is the illustration of an effective way for discovering retrieval contexts, based on the computation of *information flow* through semantic space models [Song and Bruza 2003, 2001]. As retrieval contexts may evolve over time, this work also discusses a formal approach of revising queries and contexts based on the AGM belief revision framework. Furthermore, despite discussions of the benefits of logic-based IR models in the IR literature for the past two decades [Chiaramella and Chevallet 1992; Hunter 1997; Lalmas and Bruza 1998; Lau et al. 2001; Losada and Barreiro 2001; van Rijsbergen 1986; Sebastiani 1998], few successful empirical evaluations of logic-based IR models have emerged. One important contribution of this article is to report the evaluation of our belief-revision-logic-based adaptive IR system based on large IR benchmark collections. Last but not least, the merit of improved explanatory power of a logic-based adaptive IR model is confirmed, based on a usability study.

The remainder of the article is organized as follows. In Section 2, a review of logic-based IR models is presented, followed by a preliminary of the AGM belief revision logic highlighted in Section 3. Section 4 proposes an information-theoretic method for inducing an information seeker's interests, and Section 5 describes the information-flow-based text mining method for retrieval context discovery. Based on examples in context revision and information matching, Section 6 illustrates operational details of the belief-revision-based adaptive IR model. Section 7 reports our empirical evaluation of the belief-based adaptive IR model. Finally, we offer concluding remarks and describe future directions of this research.

2. LOGIC-BASED MODELS FOR IR

Investigations into logic-based IR have attempted to formalize the notion of "aboutness" by axiomatizing its properties in terms of a neutral, theoretical framework [Bruza and Huibers 1994; Huibers and Wondergem 1998; Bruza et al. 2000]. The motivation for this has been to study the aboutness relation from a theoretical stance in order to better understand what properties of this relation promote effective retrieval (as well as which properties do not). The neutral, underlying framework is important, as it allows aboutness to be studied independent of the idiosyncrasies of a given IR model. There is as yet no consensus regarding the property of aboutness except that it should be logic-based [Huibers and Wondergem 1998; Bruza et al. 2000]. The notion of aboutness in IR has been applied to examine the postulates characterizing the AGM belief revision logic to see if the belief revision framework is applicable in the context of adaptive IR [Lau et al. 1999]. Therefore, the adaptive IR model presented in this article has its roots in earlier theoretical work in logic-based IR.

Huibers and van Linder [1996] attempted to formalize intelligent information retrieval agents based on modal logic. Modal operators were introduced to address essential concepts such as aboutness, nonaboutness, and information preclusion [Bruza and Huibers 1994] in IR. For example, one kind of retrieval agent is defined based on the notion of aboutness $d \models_a q$ (i.e., a document d to be about a query q). Strictly speaking, $d \models_a q$ is established iff the agent knows that the query q is satisfied in at least one document model of d. It is believed that such a satisfiability relation should be developed based on nonclassical logic [Huibers and van Linder 1996]. Moreover, the retrieval agent considers a document d to be nonabout q, denoted $d \not\models_a q$, iff it knows that d implies the negation of q. For the belief-revision-based IR approach presented in this article, a document is about (or not about) a query to a certain degree and this gradated information matching is modeled by epistemic entrenchment [Gärdenfors and Makinson 1988]. One advantage of a formal approach for the development of information agents is that the agents' retrieval behavior is explanatory. This is a facet we intend to exploit, as typical IR systems are not in scrutable.

In statistical analysis, the relationships among key phrases are established by frequency ratios, whereas in semantic analysis, they are established by meaning. It is believed that semantic information is critical in matching a user's needs to information objects [Hunter 1996]. For automating the use of semantic information, it is necessary to specify when a particular specialization, generalization, or synonym relationship should be used. Accordingly, an expressive formal framework is required to capture and reason about the semantic information. [Hunter 1997, 1996] proposed to use nonmonotonic logics, particularly default logic, to process semantic information about terms, and hence to identify the semantic relationships between queries and documents. For example, given the default rule ($\frac{oil \wedge cooking \wedge \neg petroleum}{\neg petroleum}$) and a query ($olive \wedge oil \wedge cooking$), the original query will be refined to exclude any information items about petroleum. It was suggested that the default rules of term relationships could be manually elicited from domain experts by asking them to illustrate the synonym, polysemy, generalization, and specialization relationships [Hunter 1997]. One weakness of default logic is that it does not augur well for large-scale implementation. For instance, the computational complexity of model checking in default logic is NP-complete [Liberatore and Schaerf 1998]. Moreover, Hunter's default logic approach for IR assumes that default rules of term relationships are manually elicited from domain experts. This is the classical knowledge acquisition bottleneck. This article extends the approach of logic-based semantic information processing by employing the information-flow method to automatically discover the semantic relationships among terms.

Logical imaging has been applied to develop IR models [Crestani and van Rijsbergen 1995; Crestani 1998]. The goal is to evaluate the probability of the conditional $(d \to q)$, whereas d and q stand for a document and a query, respectively. Logical imaging has its root in nonclassical logic, but is also based on the kinematics of probability distributions over terms. When the probability $\Pr(d \to q)$ is evaluated, the formula d will be imaged on the closest world(s) t, where t is a term representing a world in the logical imaging IR model. Then, the formula q is evaluated in these closest world(s). To capture the uncertainty in IR processes, the worlds (i.e., terms) are characterized by a probability distribution. These prior probabilities can be induced based on the inverse document frequencies (IDFs) of terms in a collection. The IR logical imaging paradigm consists of several methods to deal with the kinematics of probabilities associated with the worlds. For instance, imaging on the d-world(s) is taken as transferring the priori probabilities from the non-d-world(s) to the closest d-world(s) according to a distance measure derived from the mutual information between pairs of terms. For standard imaging, the probability associated with a non-d-world is simply transferred to the closest d-world. Then, for each term appearing in a query, the posterior probability of the term is summed to derive the retrieval status value (RSV) of the document with regards to the query q. So, for standard imaging, the RSV is derived by $\Pr(d \to q) = \sum_t \Pr(t) \times \tau(t_d; q)$, where $\tau(t_d; q) = 1$ if a query term appears in a d world (i.e., d and q have overlapping terms); otherwise it is zero. For general imaging, standard imaging is generalized in the sense that there could be more than one closest world where d is true. For general logical imaging, the percentage of probability transferred from each non-d-term to a d-term can be defined separately via individual opinionated probability distribution. The main difference between the logical imaging IR model and the belief-revision-based IR model proposed here is that term weights representing a user's preferences are induced with respect to *epistemic entrenchment*, which satisfies possibilistic [Dubois and Prade 1991] rather than probabilistic, axioms in the belief-revision-based IR model. Above all, the entrenchment degrees of terms are derived according to a user's preferences over the underlying terms, and the kinematics of entrenchment degrees are also driven by the changes of a user's information preferences.

The notions of belief, desire, and intention have been applied to characterize an information seeker's (e.g., a librarian) high-level IR goals, and belief revision has been exploited to simulate the changes of mental state of an information seeker [Logan et al. 1994]. The natural language processing (NLP) technique was used to induce the system's beliefs about an information seeker's information needs based on the continuous interactions between the information seeker and the IR system. As IR can be taken as comprising many low-level subtasks, the corresponding IR system is designed as a multiagent system with each autonomous agent performing a particular IR subtask. As the authors have already indicated, computational complexity is the main obstacle against applying such a belief-based multiagent system to support real-world IR activities [Logan et al. 1994]. Our belief revision mechanism is implemented based on a computationally efficient transmutation method [Williams 1997].

Dalal's belief revision operator [Dalal 1988] was applied to document ranking in IR [Losada and Barreiro 2003a, 2001, 1999]. Dalal's revision makes use of the cardinality of the *symmetric difference* between two interpretations I and J as a measure of the distance between these interpretations[1] (i.e., dist(I, J)). For example, the semantic distance between the set of models of ψ (i.e., M(ψ)) and I is defined by $\text{dist}(\text{M}(\psi), I) = \min_{J \in M(\psi)} \text{dist}(J, I)$. Thereby, a faithful assignment of a total preorder $\leq_\psi$ is defined as $I \leq_\psi J$ iff $\text{dist}(\text{M}(\psi), I) \leq \text{dist}(\text{M}(\psi), J)$. In IR, if a user's information needs N and a document *Doc* are represented by formulae q and d, respectively, then the semantic similarity between N and *Doc* can be approximated by the symmetric distance of the corresponding models. For example, for each $m \in \text{M(d)}$, $\text{dist}(\text{M(q)}, m) = \min_{J \in M(q)} \text{dist}(J, m)$ is computed. An average measure can then be applied to compute the symmetric distance between M(q) and M(d) by $sim(D, N) = \frac{\sum_{m \in M(d)} dist(M(q), m)}{|M(d)|}$. However, it is extremely costly to compute the symmetric difference between sets of models, even with a moderate number of atoms [Losada and Barreiro 2001]. We employ a formula-based representation for our belief revision model so that the belief-based adaptive IR system is computationally tractable. Recently, Losada and Barreiro have also adopted a formula-based approach to implement their belief-revision-based IR matching function and demonstrated some successes in the TREC-3 routing task [Losada and Barreiro 2003b]. One main difference between our belief-revision-based adaptive IR model and Losada and Barreiro's IR model is that we apply the AGM belief revision framework to model an IR system's changing beliefs about retrieval situations, whereas Losada and Barreiro's work [Losada and Barreiro 2003a, 2001, 1999] focuses on a belief-based ranking function.

3. THE AGM BELIEF REVISION LOGIC

The AGM belief revision framework, one of the most influential works in belief revision theory, is coined after its founders Alchourrón, Gärdenfors, and Makinson [Alchourrón et al. 1985]. In this framework, belief revision processes are taken as the transitions among belief states. A belief state (set) K is represented by a theory of a classical language L. A belief is represented by a sentence of L, supplemented with an entrenchment degree indicating the degree of firmness of such a belief. Three principle types of belief state transition are

[1]An interpretation is a mapping function from the propositional symbols into the set {true, false}. A model of a logical expression is an interpretation that maps the logical expression into true.

identified and modeled by the corresponding belief function. These are follows.

—Expansion K_α^+ is the process of accepting a new belief α that does not contradict existing beliefs in a belief set K.
—Contraction K_α^- is the removal of a belief α and all other beliefs that logically imply α from a belief set K.
—Revision K_α^* is the incorporation of a belief α that may contradict existing beliefs in a belief set K; this operation corresponds to adding a new query α into an existing user profile K where not α was specified.

Belief expansion K_α^+ can be interpreted as adding the relevance feedback information α into a user profile which stores the representation of an existing retrieval situation K. Since the relevance feedback information α is consistent with the current retrieval situation K (e.g., existing queries), the new retrieval requirements (i.e., relevance feedback) can simply be added to the user profile without invoking complicated updating operations. Similarly, belief contraction K_α^- can be taken as removing outdated queries from the user profile based on the relevance feedback α. Belief revision K_α^* is the most common and also the most complicated belief-change operation. Since the new retrieval requirements α (e.g., requiring documents about Java) is contradictory to the existing retrieval situation K (e.g., not requiring documents about Java), some existing information stored in the user profile must be given up. This belief revision operation should be executed according to sound principles such that useful information stored in the user profile can still be maintained.

The AGM framework comprises sets of postulates to characterize the belief functions for *consistent* and *minimal* belief revision. In the context of IR, the belief revision function K_α^* can be applied to regulate (according to the principles of minimal and rational revision) the changes of a retrieval situation K, given an information seeker's relevance feedback α. For example, given an initial retrieval situation $K = \{\text{sculpture} \rightarrow \text{art}, \neg\text{sculpture}\}$, it is clear that the information seeker does not want items about "sculpture". The sentence (sculpture $\rightarrow$ art) means requiring items about "sculpture" implies (i.e., logically entails) requiring items about "art" for this particular information seeker. If the information seeker later informs the IR system that his preference has changed to "sculpture", the revised retrieval situation becomes K' = {sculpture $\rightarrow$ art, sculpture, art} via executing the belief revision function $K^*_{\text{sculpture}}$. Based on this example, it can be observed that one advantage of the belief-revision-based IR model is that an information seeker's shifting interest (e.g., documents about "art") can be automatically deduced by the IR system. Moreover, this change can be explained based on the logical axiom of *modus ponens* (i.e., sculpture $\wedge$ (sculpture $\rightarrow$ art) $|-$ art). The symbol $\wedge$ represents the logical AND, and the symbol $|-$ denotes the logical derivability relation. Above all, the change applied to a user profile is carried out according to sound AGM principles such as minimal belief change. For instance, there is no reason for removing the belief (sculpture $\rightarrow$ art), even though the belief ($\neg$sculpture), no longer holds.

The AGM framework also specifies the constructions of belief functions based on various mechanisms. One is epistemic entrenchment $\leq$ [Gärdenfors and

Makinson 1988]. This captures the notions of significance, firmness, or defeasibility of beliefs. Formally, an epistemic entrenchment ordering is a total preorder of the sentence (e.g., α, β, γ) in L, and is characterized by the postulates [Gärdenfors and Makinson 1988],

EE1: $\forall\alpha,\beta,\gamma \in \mathrm{K} : \alpha \leq \beta \leq \gamma$ implies $\alpha \leq \gamma$ (transitivity)
EE2: $\forall\alpha,\beta \in \mathrm{K} : \alpha \mid\!- \beta$ implies $\alpha \leq \beta$ (dominance)
EE3: $\forall\alpha,\beta \in \mathrm{K} : \alpha \leq \alpha \wedge \beta$ or $\beta \leq \alpha \wedge \beta$ (conjunctiveness)
EE4: If $\mathrm{K} \neq \mathrm{K}_{\perp}$, $\alpha \notin \mathrm{K}$ iff $\forall\beta \in \mathrm{K} : \alpha \leq \beta$ (minimality)
EE5: $\forall\beta \in \mathrm{K} : \beta \leq \alpha$ implies $\mid\!- \alpha$ (maximality)

whereas $\alpha \leq \beta$ means that β is at least as entrenched as α. The notation $\mathrm{K}_{\perp}$ in EE4 indicates a belief set with inconsistent beliefs, which is not a desirable state for a rational agent. Intuitively, epistemic entrenchment relations induce preference orderings of beliefs according to the importance of these beliefs in the face of change. If inconsistency arises during a belief revision operation, the least significant beliefs (i.e., beliefs with lowest entrenchment degree) are given up in order to restore consistency. The postulates of epistemic entrenchment are valid in the context of IR [Lau et al. 1999]. For a computer-based implementation of epistemic entrenchment and hence the AGM belief functions, Williams [1995] developed finite partial entrenchment rankings to represent epistemic entrenchment orderings.

Definition 1. A finite partial entrenchment ranking is a function $\boldsymbol{B}$ that maps a finite subset of sentences in L into the unit interval [0, 1] such that the following conditions are satisfied for all $\alpha \in \mathrm{dom}(\boldsymbol{B})$.

(PER1) $\{\beta \in \mathrm{dom}(\boldsymbol{B}){:}\boldsymbol{B}(\alpha) < \boldsymbol{B}(\beta)\} \mid\!\not- \alpha.$;
(PER2) *If* $\mid\!- \neg\alpha$ *then* $\boldsymbol{B}(\alpha) = 0$; and
(PER3) $\boldsymbol{B}(\alpha) = 1$ *if and only if* $\mid\!- \alpha$.

Essentially, PER1 states that the set of sentences ranked strictly higher than a sentence α cannot entail α. This property corresponds to the dominance property of epistemic entrenchment [Gärdenfors and Makinson 1988]. Specifically, $\boldsymbol{B}(\alpha)$ is referred to as the degree of entrenchment of an explicit belief α. The set of explicit beliefs of $\boldsymbol{B}$ is $\{\alpha \in \mathrm{dom}(\boldsymbol{B})\colon \boldsymbol{B}(\alpha) > 0\}$, and is denoted $\exp(\boldsymbol{B})$. The set of implicit beliefs $\mathrm{K} = \mathrm{Cn}(\exp(\boldsymbol{B}))$ is denoted content($\boldsymbol{B}$), where Cn is the classical consequence operator. For example, $\boldsymbol{B}$(java) = 0.6 represents the entrenchment degree (i.e., firmness) of the belief about information item "java". The finite partial entrenchment ranking $\boldsymbol{B}$ = {(computer $\wedge$ java, 0.6), (computer $\wedge$ java $\rightarrow$ programming, 0.5), (programming, 0.5)} satisfies the property PER1 to PER3 because {(computer $\wedge$ java), (computer $\wedge$ java $\rightarrow$ programming)} $\mid\!-$ programming, and $\boldsymbol{B}$(programming) = 0.5, which has the same entrenchment degree as the set of sentences entailing it. However, the ranking $\boldsymbol{B}$ = {(computer $\wedge$ java, 0.6), (computer $\wedge$ java $\rightarrow$ programming, 0.5), (programming, 0.4)} does not satisfy the properties of finite partial entrenchment ranking, since beliefs with higher entrenchment degree in the ranking entail ($\mid\!-$) a belief with lower entrenchment degree (i.e., violating PER1). In particular, the belief "programming" does not have the same firmness as the set of beliefs which logically entails it. In order to describe the epistemic entrenchment ordering $\leq_B$ generated from a finite partial entrenchment ranking $\boldsymbol{B}$, it is necessary to rank implicit beliefs.

Definition 2. Let $\alpha \in L$ be a contingent sentence. Let $\boldsymbol{B}$ be a finite partial entrenchment ranking and $\beta \in \exp(\boldsymbol{B})$. The degree of entrenchment of an implicit belief α is defined by

$$degree(B,\alpha) = \begin{cases} \sup(\{B(\beta) \in range(B) : cut_{\leq}(\beta) \mid\!- \alpha\}) & \text{if } \alpha \in content(B) \\ 0 & otherwise, \end{cases}$$

where the notation sup refers to supremum (i.e., least upper bound) of a possibility distribution [Dubois and Prade 1995]. The $\text{cut}_{\leq}(\beta)$ operation extracts a set of explicit beliefs which is at least as entrenched as β, according to a finite partial entrenchment ranking $\boldsymbol{B}$. Furthermore, $\mid\!-$ is the classical inference relation. More precisely, a cut operation is defined by $\text{cut}_{\leq}(\beta) = \{\gamma \in \text{dom}(\boldsymbol{B})\text{: } \boldsymbol{B}(\beta) \leq \boldsymbol{B}(\gamma)\}$. For example, given the belief set $\boldsymbol{B}$ = {(google, 0.8), (google→search-engine, 0.5)}, the operation $\text{cut}_{\leq}$(google) will return a single belief “google”. Moreover, degree($\boldsymbol{B}$, search-engine) = 0.5 is derived according to Definition 2 because the minimal entrenchment degree in the strongest cut of $\boldsymbol{B}$ that logically entails “search-engine” is 0.5.

In a belief-revision-based adaptive IR system, queries and query contexts are represented by a set of beliefs. When an information seeker's needs and the underlying retrieval context change, the entrenchment degrees of the corresponding beliefs are raised or lowered in the adaptive IR system's knowledge base. Raising or lowering the entrenchment degree of a belief is conducted via a belief revision operation $\boldsymbol{B}^*(\alpha,i)$, where α is a sentence and i is the new entrenchment degree. For example, if an information seeker is interested in documents about “volcano” now, the belief revision operation $\boldsymbol{B}^*$(volcano,0.8) can be invoked to revise such a belief into the adaptive IR system's knowledge base to represent her current interest. The entrenchment degree of 0.8 can be computed based on our preference induction mechanism, described in Section 4. The rapid anytime maxi-adjustment (RAM) method [Lau 2003] developed based on the Maxi-adjustment method [Williams 1997] is proposed to implement the belief revision operation $\boldsymbol{B}^*(\alpha,i)$

Definition 3. Let α be a contingent sentence, j = degree($\boldsymbol{B}$, α), and 0 ≤ i < 1. The (α,i) rapid maxi-adjustment of $\boldsymbol{B}$ is $\boldsymbol{B}^*(\alpha, i)$, defined by

$$B^*(\alpha,i) = \begin{cases} B^-(\alpha,i) & \text{if} \quad i < j \\ (B^-(\neg\alpha,0))^+(\alpha,i) & \text{if} \quad i > j \\ B^+(\alpha,i) & \text{if} \quad i = j \quad \text{and} \quad j > 0 \quad \text{and} \quad \alpha \notin \exp(B) \\ B & \quad\text{otherwise.} \end{cases}$$

For all $\beta \in \text{dom}(\boldsymbol{B})$, $\boldsymbol{B}^-(\alpha,i)$ is defined as follows.

(1) For β with $B(\beta) > j$, $B^-(\alpha,i)(\beta) = B(\beta)$.

(2) For β with $i < B(\beta) \le j$,

$$B^-(\alpha,i)(\beta) = \begin{cases} i & if\,\{\gamma : B^-(\alpha,i)(\gamma) > B(\beta)\} \cup \\ & \{\delta : B^-(\alpha,i)(\delta) = B(\beta) \wedge Seq(\delta) \le Seq(\beta)\} \mid - \;\; \alpha. \\ B(\beta) & \text{otherwise} \end{cases}$$

(3) For β with $B(\beta) \le i$, $B^-(\alpha,i)(\beta) = B(\beta)$.

For all $\beta \in \text{dom}(\boldsymbol{B}) \cup \{\alpha\}$, $\boldsymbol{B}^+(\alpha,i)$ *is defined as follows.*

(1) For β with $B(\beta) \ge i$, $B^+(\alpha,i)(\beta) = B(\beta)$.
(2) For β with $j \le B(\beta) < i$,

$$B^+(\alpha,i)(\beta) = \begin{cases} i & \text{if} \quad i < \text{degree}(B, \alpha \to \beta) \\ \text{degree}(B, \alpha \to \beta) & \text{otherwise.} \end{cases}$$

(3) For β with $B(\beta) < j$, $B^+(\alpha,i)(\beta) = B(\beta)$.

The intuition of the RAM method is that if the new entrenchment degree i of a sentence α is less than its existing degree j, the belief revision operation $\boldsymbol{B}^*(\alpha, i)$ invokes a contraction process, that is, $\boldsymbol{B}^-(\alpha, i)$. In other words, the entrenchment degree of α will be lowered. If the new degree i of α is higher than its existing degree j, an expansion operation $\boldsymbol{B}^+(\alpha, i)$ should be initiated. However, to ensure that the set of beliefs remains consistent in an agent's knowledge base, $\neg\alpha$ must first be assigned the lowest entrenchment degree (i.e., contracting it from the theory). Therefore, the contraction operation $\boldsymbol{B}^-(\neg\alpha, 0)$ must be executed before raising the degree of α to i (i.e., adding the belief α to the knowledge base). If the new degree i of α is equal to its existing degree j and α is not explicitly stored in the knowledge base (i.e., $\alpha \notin \exp(\boldsymbol{B})$), it will be added to the knowledge base by executing the belief expansion operation $\boldsymbol{B}^+(\alpha, i)$. Since α has already been an implicit belief in the knowledge base, inconsistency will not occur because of the belief expansion operation.

During raising or lowering the entrenchment degree of α, the degrees of other sentences (e.g., β) are adjusted in a minimal way such that the properties PER1, PER2, and PER3 are maintained. With reference to Definition 3, both the belief contraction operation $\boldsymbol{B}^-(\alpha, i)$ and belief expansion operation $\boldsymbol{B}^+(\alpha, i)$ are further defined by three suboperations. For belief extraction, the notation $\boldsymbol{B}^-(\alpha,i)(\beta)$ refers to the new entrenchment degree of β after contracting α from the finite partial entrenchment ranking $\boldsymbol{B}$. The suboperations 1 and 3 state that existing beliefs β with entrenchment degree strictly higher than the existing degree of α, or β with entrenchment degree less than or equal to the new entrenchment degree of α, will not be affected by the belief contraction operation. For other beliefs β, their entrenchment degree may either be lowered to i (if β together with other more entrenched beliefs in $\boldsymbol{B}$, entail, α) or remain unchanged (if β, together with other more entrenched beliefs in $\boldsymbol{B}$, does not entail α).

The suboperation 2 of belief contraction can be depicted by Figure 2, where the entrenchment degree of α is lowered to i. As can be seen from Figure 2, after such a belief change, the belief β (together with other beliefs more entrenched than β) will logically entail α, which leads to the violation of property PER1 (i.e., the dominance property) for finite partial entrenchment ranking. Therefore, the

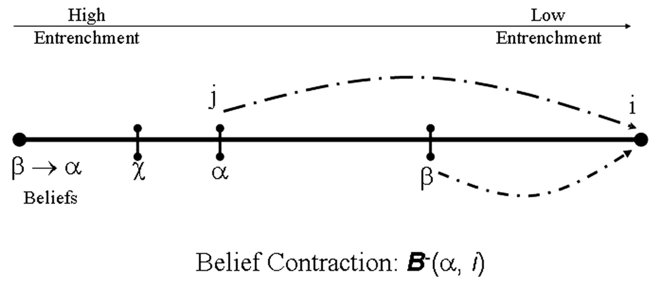

Fig. 2. The belief contraction operation.

entrenchment degree of β should also be lowered to i (the minimal change) in order to restore $\boldsymbol{B}$ to the normal state. With reference to Figure 2, an example in the context of IR is followed. Assuming that the formulae α, β, χ stand for "Web", "html", and "agent", respectively, the implication $\beta \rightarrow \alpha$ in Figure 2 indicates that if the information seeker is interested in documents about "html", she will be interested in documents about "Web" as well. In this example, the belief (html $\rightarrow$ web) is the most entrenched belief, although this may not necessarily be true in every world-world IR scenario. If the information seeker explicitly states that she is no longer interested in the "Web" (i.e., lowering the entrenchment degree of α from j to i in Figure 2), her degree of interest about "html" as modeled in an IR system's knowledge base should be lowered as well such that the property of finite partial entrenchment ranking (e.g., PER1) is maintained. The intuition is that if (html $\rightarrow$ web) is a very certain belief for this particular information seeker, and she is not interested in the Web anymore, it makes sense to drop the belief of "html"; otherwise the belief "Web" will still hold because the beliefs {(html $\rightarrow$ web), html} always imply the belief "Web". On the other hand, the belief "agent" is not affected by this belief contraction operation (i.e., its degree of significance should be the same as before).

The *Seq* function defined in suboperation 2 assigns a unique sequence number to a sentence β if there is more than one sentence with the same entrenchment degree (i.e., at the same entrenchment rank). In such a circumstance, it does not matter which sentence is contracted first (if violating PER1) because these sentences are equally preferred (or not) from an information seeker's point of view. This is the main difference between the maxi-adjustment method [Williams 1997] and the RAM method [Lau 2003] for implementing AGM belief revision. On the other hand, the belief χ will not be affected by the belief contraction operation for α because its entrenchment degree is higher than the original degree of α, and this rule is defined in suboperation 1 of $\boldsymbol{B}^{-}(\alpha, i)$ in Definition 3.

Similarly, for belief expansion $\boldsymbol{B}^{+}(\alpha, i)$, the suboperations 1 and 3 state that existing beliefs β with entrenchment degree higher than or equal to the new entrenchment degree of α, or β with entrenchment degree strictly less than the existing degree of α, will not be affected by the belief expansion operation. For other beliefs β, their entrenchment degree may either be raised to i or degree(B, $\alpha \rightarrow \beta$), dependent on which one is lower (suboperation 2).

If a finite partial entrenchment ranking $\boldsymbol{B}$ has x natural partitions, it only requires $\log_2 x$ classical satisfiability checks [Lang 1997]. Therefore, given the propositional Horn logic as the representation language, the transmutation-based RAM method for AGM belief revision only involves polynomial time complexity in the worst case. By means of the anytime approximation of an AGM belief revision operation and other optimization techniques, the belief-revision-based adaptive IR system can be scaled up to support large-scale IR. For example, since a belief revision operation is computationally expensive, the revision of a belief will be deferred until its cumulative change is larger than a predefined system threshold. We use SICStus Prolog, a commercially available Prolog system, to carry out theorem proving ($|-$). Our Java-based belief revision engine utilizes the Jasper Java interface of SICStus Prolog to communicate with its inference engine. Detailed examples of how our belief revision works within the IR context will be given later, in Section 6.

4. INDUCING USER PREFERENCES AND ENTRENCHMENT ORDERINGS

Conceptually, given a retrieval context, an IR system needs to establish a focus (i.e., a user's specific interest) over such a context. With reference to human information processing theory, this means passing a stimuli to the long-term memory to trigger the *spreading activation process* [Card et al. 1983]. Since a user may have difficulty in specifying her specific interest, a highly autonomous IR system should be able to automatically induce a user's interest based on the user's interactions with the system. For example, if the user has requested some documents recently, the retrieved documents will form the basis of supervised learning for the user's preferences. In terms of document representation, a document is preprocessed according to traditional IR techniques to extract a set of tokens (e.g., stems, n-grams, or phrases) as its characterization [Salton and McGill 1983]. At the symbolic level, each token$t$ is mapped to a positive literal of the classical propositional language L.

Conversely, a user's information need is induced based on a set of relevant documents D^+ and a set of nonrelevant documents D^-, directly or indirectly judged by the information seeker (e.g., based on archived documents or viewing time). Essentially, three types of token can be extracted: Positive tokens represent what items the information seeker would like to retrieve; negative tokens indicate what the information seeker does not want; neutral tokens are not good indicators of her information needs. The following preference induction method is used to extract various types of tokens and to induce the corresponding preference values. It is developed based on the information-theoretic *keyword classifier* which was successfully applied to adaptive information filtering [Kindo et al. 1997]. This is

$$\begin{aligned} pre(t) = \varepsilon \times \tanh\Bigg(&\frac{df(t)}{pos} \times \Pr(rel|t) \times \log_2 \frac{\Pr(rel|t)}{\Pr(rel)} \\ &- \frac{df(t)}{neg} \times \Pr(nrel|t) \times \log_2 \frac{\Pr(nrel|t)}{\Pr(nrel)}\Bigg), \end{aligned}$$

where $-1 < pre(t) < 1$ is the preference value for a term t. The proposed preference induction mechanism is similar to a linear classifier. If a term appears in many positive documents (i.e., $df(t) > pos$) and the presence of the term is

Table I. An Example of Training Documents

Document	Document Content	User Judgment
Doc1	java, technology, program	Relevant
Doc 2	java, technology, program	Relevant
Doc 3	java, volcano, program, technology	Relevant
Doc 4	java, volcano, program, technology	Relevant
Doc 5	java, volcano, program, technology	Relevant
Doc 6	program	Non-Relevant
Doc 7	computer, program, technology	Non-Relevant
Doc 8	computer, program, technology	Non-Relevant
Doc 9	computer, internet, program, technology	Non-Relevant
Doc 10	computer, internet, program, technology	Non-Relevant

likely a contributing factor (i.e., $\Pr(rel|t) > \Pr(rel)$) towards the relevance of the document, a high positive preference value will be derived. Similarly, the classifier can also take into account the presence of negative feedback. The relative weight of the positive and negative evidence is controlled by adjusting the parameters of *pos* and *neg*, respectively.

It should be noted that expected cross-entropy EH [Koller and Sahami 1997] bears much similarity with the previous preference induction formula. Expected cross-entropy is defined by $EH(t, C) = \Pr(t)\sum_{c\in C}\Pr(c|t) \times \log_2 \frac{\Pr(c|t)}{\Pr(c)}$, where the set C includes {relevant, nonrelevant}. The terms *pos* and *neg* are the learning thresholds for positive and negative terms, respectively. The function tanh is the hyperbolic tangent function. The adjustment factor ε ensures that the induced entrenchment degree is less than the maximal entrenchment degree because all the induced beliefs should be contractable (i.e., they are not tautologies). $\Pr(rel|t) = \frac{df(t_{rel})}{df(t)}$ is the estimated conditional probability (based on the training data) that a document is relevant, given that it contains a term t. It is expressed as the fraction of the number of relevant documents which contain the term t(i.e., $df(t_{rel})$) over the total number of documents which contain term t (i.e., $df(t)$). Similarly, $\Pr(nrel\,|t) = \frac{df(t_{nrel})}{df(t)}$ is the estimated conditional probability that a document is nonrelevant if it contains term t. The term $df(t_{nrel})$represents the number of nonrelevant documents which contain term t. In addition, $\Pr(rel) = \frac{|D^+|}{|D^+|+|D^-|}$ and $\Pr(nrel) = \frac{|D^-|}{|D^+|+|D^-|}$ are the priori probabilities that a document is relevant and nonrelevant respectively. A positive *pre(t)* indicates that the underlying term $t$is a positive token, whereas a negative preference value implies that t is a negative token. If the preference value of a token is below a threshold λ, the token is considered neutral. A positive token is mapped to a positive literal such as l, whereas a negative token is mapped to a negated literal such as $\neg l$. The entrenchment degree $\boldsymbol{B}(\alpha_t)$ of an explicit belief α_t is computed according to.

$$B(\alpha_t) = \begin{cases} \frac{|pre(t)|-\lambda}{1-\lambda} & if \quad |pre(t)| > \lambda \\ 0 & otherwise. \end{cases}$$

Table I depicts an example of ten training documents and their associated relevance judgments from an assumed user. Table II shows the result of applying the aforementioned preference induction method to the sample of training

Table II. An Example of Induced Epistemic Entrenchment

Term	df(t_{rel})	df(t_{nrel})	pre(t)	α_t	B(α_t)
java	5	0	0.724	java	0.605
computer	0	4	−0.631	¬computer	0.473
volcano	3	0	0.510	volcano	0.300
internet	0	2	−0.361	¬internet	0.087
technology	5	4	0.266	–	–
program	5	5	0	–	–

documents depicted in Table I. The last column in Table II shows the derived entrenchment degrees associated with beliefs representing some term preferences. In this example, the parameters $|D^+| = |D^-| = 5$, $\varepsilon = 0.95$, $\lambda = 0.3$, $pos = 5$, and $neg = 5$ are assumed in the preference induction process. Also, $|D^+|$ and $|D^-|$ represent the sizes of sets of known relevant documents and nonrelevant documents, respectively. The parameters ε, *pos*, and *neg* are applied to the preference induction formula to control the maximal preference value, weight of positive evidence, and weight of negative evidence, respectively. The parameter λ is used to filter out insignificant beliefs of an information seeker's preference. The parameters ε, λ, *pos*, and *neg* are estimated based on empirical evaluation. If a parameter value leads to satisfactory retrieval performance in the pilot runs, it will be adopted for the experiments related to that particular collection. As the contents of D^+ and D^- evolve according to an information seeker's changing preferences, the entrenchment degrees of the corresponding beliefs are raised or lowered in the user profile of the adaptive IR system. Changes applied to the epistemic entrenchment ordering of beliefs will then generate different nonmonotonic consequence relations, which underpin the adaptive IR system's decisions about document relevance at various points of time.

5. MINING CONTEXTUAL KNOWLEDGE

Contextual knowledge refers to the semantic relationships among concepts. This section illustrates the approaches we adopt to automatically derive two types of semantic relationship: information flow and information preclusion.

5.1 Information-Flow-Based Text Mining

Information-flow computations through a high-dimensional semantic space have been proposed as a means for computing term associations, both explicit and implicit. Song and Bruza [2003, 2001]. Information flow is motivated from the conceptual level of cognition [Gärdenfors 2000]. Encouraging results have been obtained with information-flow-based query expansion using a semantic space created by hyperspace analog to language (HAL) [Bruza and Song 2002]. In the belief-revision-based adaptive IR system, the information-flow method is applied to discover initial contextual information.

Hyperspace analog to language is an exemplar of models emerging from the cognitive science generally referred to as "semantic space" [Lowe 2001]. These models run over a corpus of text and build representations of words in a (reduced) high-dimensional space. The appeal of these models to IR-related applications is that they have an encouraging track record of replicating human information processing, for example, semantic word association norms. IR has a long history of exploring term associations. Typically, the underlying basis is probabilistic or more specifically, information-theoretic, such as EMIM. Granted there has been notable pragmatic success computing term associations in this way, the fundamental question remains as to whether such associations actually accord with those the user would make. For this reason, we prefer to use semantic space models due to their cognitive compatibility, especially in relation to human word association norms. Two prominent semantic space models are hyperspace analog to language (HAL) [Burgess et al. 1998] and latent semantic analysis (LSA) [Landauer et al. 1998]. We prefer to use HAL, as it does not involve the additional computational cost of dimensionally reducing the semantic space. Moreover, in the research cited earlier which employed HAL to underpin query expansion, the performance differential over two prominent probabilistic approaches was marked. Although the experiments reported do not settle the question of whether the computation of term associations should be cognitively or probabilistically motivated, there was more than enough evidence giving encouragement to pursue further investigations with HAL in an IR setting.

HAL produces representations of words in a high-dimensional vector space; these vector representations seem to correlate with the equivalent human representations. For example, word associations computed on the basis of HAL vectors seem to mimic human word association judgments [Burgess et al. 1998]. HAL is "a model that acquires representations of meaning by capitalizing on large-scale co-occurrence information inherent in the input stream of language" [Burgess et al., ibig]. The space comprises high-dimensional vector representations for each term in the vocabulary. Briefly, given an n-word vocabulary, the HAL space is a $n \times n$ matrix constructed by moving a window of length l over a corpus by one-word increments, ignoring punctuation, sentence, and paragraph boundaries. All words within the window are considered as cooccurring with each other, with strengths inversely proportional to the distance between them. After traversing the corpus, an accumulated cooccurrence matrix for all the words in a target vocabulary is produced. The cooccurrence matrix is added to its transpose to result in a symmetric matrix, the rows of which are termed HAL vectors.

As an example of a HAL vector derived from a large corpus, consider part of the normalized HAL vector for "superconductors" computed from a corpus of Associated Press news.

> superconductors = $<$ U.S.:0.11 american:0.07 basic:0.11 bulk:0.13 called:0.15 capacity:0.08 carry:0.15 ceramic:0.11 commercial:0.15 consortium:0.18 cooled:0.06 current:0.10 develop:0.12 dover:0.06 electricity:0.18 energy:0.07 field:0.06 goal:0.06 high:0.34 higher:0.06 improved:0.06 japan:0.14 loss:0.13 low:0.06 make:0.07 materials:0.25 new:0.24 require:0.09 research:0.12 researching:0.13 resistance:0.13 retain:0.06 scientists:0.11 semiconductors:0.10 states:0.11 switzerland:0.06 technology:0.06 temperature:0.48 theory:0.06 united:0.10 university:0.06$>$

This example demonstrates how a word is represented as a weighted vector whose dimensions comprise other words. The weights represent the strengths of association between “superconductors” and other words seen in the context of the sliding window: The higher the weight of a word, the more it has lexically cooccurred with “superconductors” in the same context(s). The quality of HAL vectors is influenced by window size; the longer the window, the higher the chance of representing spurious associations between terms. Burgess et al. [ibid.] use a window size of eight or ten in their studies.

More formally, a concept[2] c_i is a vector representation $c_i = \langle w_{c_i p_1}, w_{c_i p_2}, \ldots w_{c_i p_n} \rangle$, where $p_1, p_2, \ldots, p_n$ correspond to words in the vocabulary and are called *dimensions*, n is the dimensionality of the HAL space, and $w_{c_i p_i}$ denotes the weight of p_i in the vector of c_i. A dimension is termed a *property* of concept c_i if and only if its weight is greater than zero. A property p_i of a concept c_i is termed a quality property iff $w_{c_i p_i} > \partial$, where ∂ is a nonzero threshold value. Let $QP_\partial(c)$ denote the set of quality properties of concept c. Also, $QP_\mu(c)$ denotes the set of quality properties of concept c with above mean positive weight. (Mean positive weight is calculated as the mean of all dimensions greater than zero.) For notational convenience, $QP(c)$ will be used to denote $QP_0(c)$. The latter notation simply denotes the set of dimensions of concept c with positive weight.

Concept combination is an important issue, as combinations of words may represent a single underlying concept, for example, *space program*. An important intuition in concept combination is that one concept can dominate the other. For example, the term “space” can be considered to dominate the term “program” because it carries more of the information in the phrase. Given two concepts $c_1 = \langle w_{c_1 p_1}, w_{c_1 p_2}, \ldots w_{c_1 p_n} \rangle$ and $c_2 = \langle w_{c_2 p_1}, w_{c_2 p_2}, \ldots w_{c_2 p_n} \rangle$, the vector representation of the combined concept is denoted $c_1 \oplus c_2$. Dominance is assumed proportional to the inverse document frequency (IDF) of the concept in question. For example, “space” is deemed dominant over “program” as its IDF is higher than that of “program”. The combination of concepts represented by HAL vectors can be computed by a heuristic form of vector addition [Song and Bruza 2003]. For the purposes of this article it is sufficient to bear in mind that the result of concept combination is a vector representation.

A HAL vector can be considered to represent the information “state” of a particular concept (or combination of concepts) with respect to a given corpus of text. The degree of information flow between “space program” and “satellites”, say, is directly related to the degree of inclusion between the respective information states represented by HAL vectors. Total inclusion leads to maximum information flow. Assuming a vector space, information flow is computed by a function, the domain of which is vector pairs (u, v) and with range [0, 1]. The vector u is denoted as the *source* of the information flow and vector v as the target. Intuitively, the function tries to express how much a source concept implies a target. For example, one would expect significant information flow between “space program” and “satellites”, denoted $space \oplus program \vdash satellites$.

[2]The word “concept” is used somewhat loosely; it can be envisaged as “term” in the traditional IR sense. The word “concept” employed for the semantic space computed by HAL can be viewed as a computational approximation, albeit rather primitive, of human conceptual space. See Gärdenfors [2000] for more details.

When information flow is above a certain threshold ω, then an information flow is established between the source and target concepts, as

$$t_1, \ldots, t_m \mid\!- t_j \text{ iff } \mathrm{flow}\left(\oplus_{i=1}^{m} c_i, c_j\right) > \omega$$

where c_i denotes the HAL vector of the concept t_i. (For ease of exposition, $\oplus_{i=1}^{m} c_i$ will be simplified to c_i because combinations of concepts are also concepts). The threshold ω is set empirically. The degree of flow is computed in terms of the ratio of intersecting quality properties of c_i and c_j to the number of quality properties in the source c_i

$$\mathrm{flow}(c_i, c_j) = \frac{\sum\limits_{p_l \in (QP_\mu(c_i) \wedge QP(c_j))} w_{c_i p_l}}{\sum\limits_{p_k \in QP_\mu(c_i)} w_{c_i p_k}}$$

An information flow $t_1, \ldots, t_m \mid\!- t_j$ is converted to the corresponding belief $(t_1 \wedge t_2 \wedge, \ldots, \wedge t_m \rightarrow t_j, i)$. The entrenchment degree i is computed by multiplying $\mathit{flow}(\oplus c_i, c_j)$ with an adjustment factor. By way of illustration, after applying information-flow-based text mining to the Reuters-21578 collection, it is found that the concept "NEC" exhibits strong information flow to other concepts such as (computer, 0.9415), (electronics, 0.8355), (Japan, 0.7623), etc. In this way, contextual rules such as (NEC → Japan, 0.7623) are established.

5.2 Inducing Information Preclusion Relations

The other important semantic relationship of *information preclusion* [Bruza and Huibers 1994] can be acquired from a corpus through supervised learning. An information preclusion relation such as $t_i \perp t_j$ indicates that a token t_i precludes another token t_j driven by an information seeker's information needs. For example, *car* ⊥ *boat* may hold if a user interested in documents about "car" is not interested in documents about "boat" with respect to a particular retrieval task. An information preclusion relation is represented by a rule $t_i \rightarrow \neg t_j$. For a term t from the set of positive tokens (e.g., by looking up the system table as depicted in Table II), if $df(t_{rel}) > \gamma$ and $df(t_{nrel}) = 0$, the term t is added to the antecedent set L. Similarly, for a term t satisfying $df(t_{nrel}) > \gamma$ and $df(t_{rel}) = 0$, it is added to the consequent set R. Then, for each term $t_i \in L$, generate a rule $t_i \rightarrow \neg t_j$ for each $t_j \in R$. The entrenchment degree of such a rule is derived by $Pr(t_i) \times Pr(t_j) \times \delta$. The adjustment factor δ is estimated based on empirical tests during the pilot runs of a document collection. In practice, the computations of both information flow and information preclusion are conducted offline in order to maintain good online information retrieval performance. As information flow is an unsupervised learning method (i.e., the class label for a document is not required), it can be done before interactive information retrieval takes place. However, the mining of information preclusion associations can only be conducted when a certain number of training documents is available.

Table III. The First Retrieval Situation RS_1

Belief (α)	Before Revision	After Revision
science	0.900	0.900
volcanology $\rightarrow \neg$computer	0.830	0.830
java $\wedge$ computer $\rightarrow$ programming	0.713	0.713
java $\wedge$ volcano $\rightarrow$ merapi	0.713	0.713
science $\wedge$ volcanology $\rightarrow$ volcano	0.695	0.695
science $\rightarrow$ computer	0.427	0.427
java	**0.000**	**0.900**
computer	0.427	0.427
$\neg$volcanology	0.427	0.427
programming	**0.000**	**0.427**
$\neg$computer	0.000	0.000
merapi	0.000	0.000
volcano	0.000	0.000

6. PROFILE REVISION AND SEMANTIC-BASED DOCUMENT MATCHING

Based on an information seeker's explicit and implicit feedback, the belief-revision-based adaptive IR system can induce a set of beliefs representing the information seeker's current information needs, as described in Section 4. In addition, contextual knowledge is discovered via information-flow-based text mining. The second step of learning a (possibly changing) retrieval situation is to revise the corresponding beliefs into the system's user profile (i.e., knowledge base) via belief revision processes. In general, the belief-based adaptive IR system will maintain a separate user profile for each individual information seeker who is interested in a specific topic. Furthermore, these profiles can be aggregated to form a generic profile for a group of information seekers who share similar interests (e.g., a default profile for new system users). The following examples demonstrate how user profile revision (i.e., learning) and information matching are conducted in the belief-based adaptive IR system.

6.1 Learning in Retrieval Situation RS_1

Table III illustrates the changes applied to the adaptive IR system's user profile (i.e., knowledge base K) before and after a query is received. This example describes a retrieval situation that an information seeker with general interest in science (e.g., perhaps a student with major in science subjects) issues the query "java". Initial contextual knowledge is acquired before the query is received. The initial belief about the information seeker's background for example, (science, 0.9), is obtained via the dialog between the user and the system, and a default entrenchment degree is assigned to this relatively certain information. In addition, an inference rule such as (java $\wedge$ computer $\rightarrow$ programming) is obtained by executing the information-flow-based text mining method. An information preclusion relationship such as (volcanology $\rightarrow \neg$computer) is induced according to the method described in Section 5.2. In this example, it is assumed that the information seeker has reviewed some documents before, and therefore a training set is available. As illustrated in this example, a retrieval context is not simply a set of terms as viewed by some researchers [Lawrence 2000; Cohen and Singer 1996], but has a richer representation which includes the

Table IV. The Second Retrieval Situation RS_2

Belief (α)	Before Revision	After Revision
science	0.900	0.900
java	0.900	0.900
volcanology $\rightarrow \neg$computer	0.830	0.830
java $\wedge$ computer $\rightarrow$ programming	0.713	0.713
java $\wedge$ volcano $\rightarrow$ merapi	0.713	0.713
science $\wedge$ volcanology $\rightarrow$ volcano	0.695	0.695
science $\rightarrow$ computer	**0.427**	**0.000**
volcanology	**0.000**	**0.900**
$\neg$computer	**0.000**	**0.830**
merapi	**0.000**	**0.695**
volcano	**0.000**	**0.695**
$\neg$volcanology	**0.427**	**0.000**
programming	**0.427**	**0.000**
computer	**0.427**	**0.000**

relationships among terms triggered by particular retrieval tasks [Hirst and St-Onge 1998]. One distinct advantage of the belief-based adaptive IR system is that the expressive power of the belief revision logic allows the semantics of a retrieval context to be represented, thereby opening the door to more effective information matching.

To incorporate the user's current interest (i.e., a query) into the adaptive IR system's knowledge base, the rapid anytime maxi-adjustment (RAM) operation $\boldsymbol{B}^*$(java, 0.9) is invoked. After such a learning process, the system can automatically deduce that the user may be interested in documents about "computer programming". The revised user profile will be used to match against incoming documents. In the next subsection, we propose an entrenchment-based similarity measure as the matching function. By viewing a recommended document, the user may eventually find that she is interested or not interested in the document. The adaptive IR system can then further revise its beliefs about the user's preferences, which will be illustrated by the Table IV later, based on the user's relevance feedback and the preference induction mechanism described in Section 4. The upper sections (above the double lines) of Tables III and IV list all the explicit beliefs, and the lower sections show all the implicit beliefs deduced based on the explicit beliefs stored in the system's knowledge base. The second column in Tables III and IV indicates the entrenchment degrees of corresponding beliefs before belief revision takes place, and the third column shows the entrenchment degrees of those beliefs after belief revision pertaining to a retrieval situation is conducted. The entrenchment degrees of the implicit beliefs are computed according to Definition 2 of Section 3. Only implicit beliefs relevant for our discussion are shown in the tables. A belief with zero entrenchment degree is contracted from the knowledge base K. Beliefs with changing entrenchment degrees are highlighted in these tables.

6.2 Semantic Similarity Matching

Similarity measures are often used to induce document rankings for IR [Losada and Barreiro 1999; Salton and McGill 1983]. An entrenchment-based similarity measure *Sim(RS, doc)* is developed to approximate the *semantic correspondence* between a retrieval situation *RS* and an information object *doc*.

$$Sim(RS,\text{doc}) \approx Sim(B,d) = \frac{\sum_{l\in d}[\text{degree}(B,l) - \text{degree}(B,\neg l)]}{|S|}$$

A retrieval situation refers to a user query and associated retrieval context. It should be noted that the proposed similarity measure *Sim(RS, doc)* not only considers the syntactic aspects in information matching, but also takes into account the semantics among information items, by means of nonmonotonic inference conducted according to the degree function $degree(\boldsymbol{B}, \alpha)$ defined in Definition 2 of Section 3. The aforesaid similarity measure combines the advantages of quantitative ranking and symbolic reasoning in a single formulation. It is not a simple overlapping model, since the function $degree(\boldsymbol{B}, l)$ invokes nonmonotonic inference about the relevance of a document characterization d with respect to the knowledge base $content(\boldsymbol{B})$ which represents a retrieval situation *RS*. The basic idea is that a document *doc* is characterized by a set of positive literals $d = \{l_1, l_2, \ldots .l_n\}$. If the system's knowledge base $K = content(\boldsymbol{B})$ logically entails an atom l_i, a positive contribution is made to the overall similarity score because of the partial semantic correspondence between *RS* and *doc*. This kind of logical entailment is nonclassical and underpinned by the expectation inference relation $|\!\sim$ [Gärdenfors and Makinson 1994]. On the other hand, if K implies the negation of a literal $l_i \in d$, it indicates certain semantic distance between *RS* and *doc*. Therefore, the similarity value is reduced by a certain degree. The set S of active information carriers partially characterizing the document *doc* is defined by $\text{S} = \{l_i \in d : \text{degree}(\boldsymbol{B}, l_i) > 0 \vee \text{degree}(\boldsymbol{B}, \neg l_i) > 0\}$. At the implementation level, both the queries and associated retrieval context are represented by the beliefs stored in adaptive IR system's knowledge base K. With reference to the first retrieval situation RS_1, if the following three documents are evaluated by the adaptive IR system, the result of document matching will be

$d_1 = \{\text{computer, programming}\}$

$d_2 = \{\text{volcanology, computer, programming}\}$

$d_3 = \{\text{merapi, volcano}\}$

$\text{Sim}(\boldsymbol{B}, d_1) = 0.427$

$\text{Sim}(\boldsymbol{B}, d_2) = 0.142$

$\text{Sim}(\boldsymbol{B}, d_3) = 0$

$\therefore \text{doc}_3 \le \text{doc}_2 \le \text{doc}_1$

where $\text{doc}_i \le \text{doc}_j$ means doc_j is at least as preferable as doc_i with respect to the retrieval situation. Such a ranking corresponds to our intuition about document preference with respect to the retrieval situation RS_1.

6.3 Learning in Retrieval Situation RS_2

If the retrieval context is changed because the information seeker is actually a science student specializing in "volcanology", the system's knowledge base K

before and after incorporating such a contextual change is depicted in Table IV. The new information about the user's background is revised into K via the belief revision operation $\boldsymbol{B}^*$(volcanology, 0.9). In this case, the entrenchment degree is a default value applied to the background information entered by the user.

According to the RAM method, $(\boldsymbol{B}^-(\neg$ volcanology, $0))^+$ (volcanology, 0.9) is executed. As the belief (¬volcanology, 0.427) is implied by the knowledge base K, the belief revision operation must first lower the entrenchment degree of (¬volcanology) to zero, before adding the explicit belief (volcanology, 0.9) into K such that the representation of the retrieval situation remains consistent and coherent. The belief (volcanology, 0.9) then invokes a number of expansion operations leading to the insertions of implicit beliefs (Merapi, 0.695), (¬ computer, 0.830), and (volcano, 0.695). Before (¬ computer, 0.830) is inserted, $\boldsymbol{B}^-$ (computer, 0) first has to be executed. In doing so, the explicit belief (science → computer, 0.427) (i.e., the least entrenched belief) is contracted from the theory base exp($\boldsymbol{B}$). If a dispatch threshold *disp* is used, the system can make a binary decision of document selection. A document d will be selected by the system if $\text{Sim}(\boldsymbol{B}, d) > disp$ is established. For instance, if $disp = 0.1$ is chosen, the system will select doc_1 and doc_2 for the user in the first retrieval situation, but doc_3 in the second situation. The document ranking in retrieval situation RS_2 is as follows.

$$\text{Sim}(\boldsymbol{B}, d_1) = -0.830$$
$$\text{Sim}(\boldsymbol{B}, d_2) = 0.035$$
$$\text{Sim}(\boldsymbol{B}, d_3) = 0.695$$

$$\therefore \text{doc}_1 \leq \text{doc}_2 \leq \text{doc}_3$$

6.4 Explaining the System's IR Decision

Figure 3 is a screen shot of the belief-based adaptive IR system. It demonstrates how this adaptive IR system facilitates the generation of human-comprehensible explanations to justify the system's document selection decisions. In this example, the content of the system's knowledge base (theory base) is shown at the lower right corner. This represents the system's beliefs about the current retrieval situation (i.e., query and associated retrieval context). In particular, the hypothetical user's interest is about "Internet", and the contextual information includes {(internet→softbot, 0.85), (softbot→spider, 0.85), (spider→crawler, 0.85), (crawler→ ¬music, 0.023)}. On the other hand, a document characterized by d = {internet, spider, music, mp3}, and the system's justifications for selecting that document are depicted at the upper left window. As illustrated by the explanation window, the IR system selects this document because the token "internet" is contained in the document and this is explicitly requested by the user. Moreover, the token "spider" appears in the document, and such a token is associated with the token "internet"; this relationship is dynamically discovered by applying the information-flow mining to a set of documents recently browsed by the user (or against a document collection, as conducted in our experiments). Accordingly, the token "spider" contributes a positive value to the overall document similarity score *Sim(RS, doc)*.

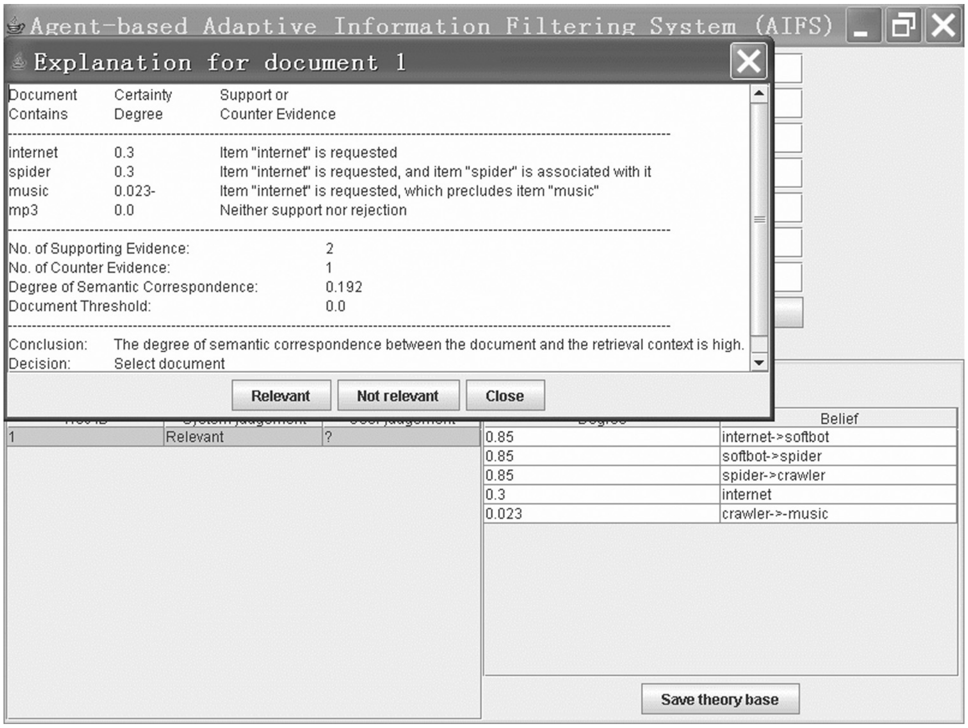

Fig. 3. Explaining the system's document selection decision.

On the other hand, the token "music" from the document contributes a negative value (-0.023) to the overall document score because there is a small semantic gap between the user's information need and part of the document, as evidenced by the contextual information $\{$(internet$\rightarrow$softbot, 0.85), (softbot$\rightarrow$spider, 0.85), (spider$\rightarrow$crawler, 0.85), (crawler$\rightarrow$ $\neg$music, 0.023)$\}$. After reviewing the system's judgment, the user can provide relevance feedback by clicking the "Relevant" or the "Not Relevant" button at the bottom of the document selection/explanation window. This action will trigger some belief revision operations to raise or lower the entrenchment degrees of the corresponding beliefs in the system's knowledge base K (i.e., user profile). Periodically, the training documents (documents judged by the user) will also be used to discover information preclusion relationships.

7. SYSTEM EVALUATION

7.1 General Experimental Procedures

The evaluation procedure of our belief-revision-based adaptive IR system (BR) is based on the adaptive information filtering benchmark task used in the 7th *Text REtrieval Conference (TREC-7)* [Hull 1998]. A set of TREC topics was used to represent the diverse initial information needs of a hypothetical information seeker, and a relevance feedback file is employed to simulate human relevance feedback for documents. The TREC relevance feedback file consists of three fields, the first field is a TREC topic number, the second contains the TREC-AP document ID, and the last indicates whether a particular document is relevant (constant “1”) or not (constant “0”) with respect to a particular TREC topic. A stream of incoming documents (e.g., newswires) is presented to a filtering system; the system needs to make a binary decision on document disposal (e.g., accept or reject) immediately whenever a document arrives. In this sense, the adaptive filtering task is quite different from traditional routing task where a batch mode of ranking operation is conducted [Robertson 2002]. If the filtering system decides to retrieve a document, the relevance judgment information associated with the document is available to the system to revise a user profile. Our experimental procedure differs from the TREC-7 adaptive filtering procedure in that the relevance judgment information of each document is available to our adaptive IR system to revise a user profile after the document dissemination phase, whereas only the relevance judgment information of a retrieved document is available to a filtering system to revise a user profile in TREC-7. In addition, our belief-based adaptive IR system can make use of information flow to develop the appropriate retrieval context before adaptive filtering begins. In each experimental run, documents are filtered with respect to the specific interest (a single topic) of a hypothetical information seeker. We used the same performance measures (F1 and F3 utilities) as adopted in the TREC-7 adaptive filtering task to evaluate the system. Specifically,

$$\begin{aligned} \mathrm{F1} &= 3 \times |\mathrm{Ret\_Rel}| - 2 \times |\mathrm{Ret\_Nrel}| \\ \mathrm{F3} &= 4 \times |\mathrm{Ret\_Rel}| - |\mathrm{Ret\_Nrel}| \end{aligned}$$

where |Ret_Rel| and |Ret_Nrel| refer to the number of relevant and nonrelevant documents retrieved by an IR system, respectively. The larger the F1 or F3 score, the better an IR system performs. In terms of the document collections, we have tested our system based on both the TREC-AP collection [Hull 1998] and the Reuters-21578 collection.

A baseline adaptive IR system (VS) was also developed based on the vector space model [Salton and McGill 1983] and on the Rocchio learning method [Rocchio 1971]. In the VS system, a term weighting formula was used to compute the weight w_t of a term t[Salton 1991], as

$$\mathrm{w_t} = \frac{\left(0.5 + 0.5\frac{tf_t}{\max\ tf}\right) \cdot \log_2 \frac{N}{N_t}}{\sqrt{\sum_{k \in d} \left(\left(0.5 + 0.5\frac{tf_k}{\max\ tf}\right) \cdot \log_2 \frac{N}{N_k}\right)^2}},$$

where tf_t is the occurrence frequency of term t in a document d, and N_t and N represent the number of documents containing term t and the total number of documents in a collection, respectively. The Rocchio learning method [Rocchio 1971] was used in the VS system. Specifically,

$$\vec{Q}_{i+1} = \alpha \times \vec{Q}_i + \frac{\beta}{|rel|} \times \sum_{d \in rel} \frac{\vec{d}}{\|\vec{d}\,\|} - \frac{\gamma}{|nrel|} \times \sum_{d \in nrel} \frac{\vec{d}}{\|\vec{d}\,\|},$$

where $\vec{Q}_{i+1}$ is the new prototype vector generated at time point $i+1$, and $\vec{Q}_i$ represents the original vector before learning takes place. The set of relevant (*rel*) and nonrelevant (*nrel*) documents are those documents filtered by the VS system during a training cycle. The simulated human relevance judgment for each document (AP or Reuters-21578 document) is obtained from the relevance feedback file. The notation $\vec{d}$ represents a document vector of TFIDF weights and $\|\vec{d}\|$ is Euclidian vector length. The parameters $\alpha = 1$, $\beta = 0.75$, and $\gamma = 0.25$ were applied in this experiment.

Information-flow mining was applied to the subset of a document collection (e.g., AP90 of the entire TREC-AP collection) to develop the initial retrieval context. For each topic, only 20 significant information-flow relations were loaded into the BR system's knowledge base before an adaptive filtering task began. No information preclusion mining was applied to the experiments reported in this article. To develop a comparable setting for the VS system, WordNet Miller et al. [1990] was used to expand each initial query (e.g., terms extracted from the title field of a TREC topic or from the topic description file of the Reuters-21578 collection). In particular, a maximum of 20 terms were added to the initial query-based on the synonym sets (synsets) found in WordNet. When WordNet-based semantic query expansion was performed, the most-common sense (as defined by WordNet) was consulted first, followed by the less-common senses. For instance, the title "acquisitions" of TREC topic 002 was expanded with terms such as "contracting", "getting", "transferred possession", "learning", "accomplishment", "attainment", etc. An expanded query was then loaded to VS system, before the adaptive filtering task was performed. All experimental runs were based on the configuration of a single Pentium III 800MHz CPU with 256MB main memory. Both the BR and VS systems were developed using Java SDK 1.4. For the experiments reported in this work, the parameters $\varepsilon = 0.95$, $\lambda = 0.3$, $pos = 5$, and $neg = 1000$ were applied to the BR system. These parameters were estimated based on empirical evaluation. For example, several topics from each collection were randomly selected, and then various parameter values were tried to obtain a reasonable filtering performance. The best combination of parameter values in these pilot runs was then applied to the BR system in subsequent runs.

Last but not least, document dissemination thresholding is also a challenging task in adaptive information filtering. Nevertheless, an optimal thresholding strategy often depends on the particular performance measure [Arampatzis and van Hameren 2001; Zhang and Callan 2001]. In order to retain the computational efficiency of our logic-based adaptive IR system, we adopted the simple thresholding strategy by tracking the average document scores between the set of relevant documents and the set of nonrelevant documents, and then set the threshold somewhere between these two points at each learning cycle [Callan 1998]. More specifically, simple heuristic rules were used to calibrate our dissemination threshold. For the typical situation where the average document score of relevant documents (Avg^+) is higher than that of nonrelevant documents (Avg^-), the dissemination threshold is estimated according to $\text{Avg}^+ + [(\text{Avg}^+ - \text{Avg}^-) \times \theta]$, where θ is an adjustment factor and is set to -0.25 for our experiments. On the other hand, if the average document score of the nonrelevant documents is greater than or equal to average document score of relevant documents at a particular learning cycle, our system will take a conservative decision by not altering the threshold at all. According to our observation, this situation could happen during the early stages of filtering. The initial dissemination threshold for a topic was established based on empirical testing. The same thresholding strategy was applied to both the BR and VS systems, respectively. We believe that employing a more sophisticated, dynamic thresholding method may improve the performance of our belief-based adaptive IR system. However, we will leave this task as part of our future work.

7.2 Evaluation Based on the TREC-AP Collection

The TREC-AP comprises the *Associated Press* (AP) newswires covering the period from 1988 to 1990, with a total number of 242,918 documents (728MB). As with the TREC-7 adaptive filtering task, a set of 50 topics (from topics 001 to 050) was used to represent the diverse initial information needs of a hypothetical user. An example of the TREC topic 008 is shown next.

<top>

<head> Tipster Topic Description

<num> Number: 008

<dom> Domain: International Economics

<title> Topic: Economic Projections

<desc> Description:

Document will contain quantitative projections of the future value of some economic indicator for countries other than the U.S.

</top>

A TREC topic contains several fields, each marked up by the corresponding tag. The TREC topic-number field is marked up by the <num> tag. Our experiments only used the title field marked up by the <title> tag to represent the initial query of a particular filtering run. Exactly the same relevance feedback file of TREC-7 was used in our experiment. Tables V(a) and V(b) show the results of our runs for TREC topics 001–050. The first column in these tables identifies the TREC topic number and the second column shows the number of true relevant documents judged by the TREC forum; the remaining three columns show the BR system's performance. The last three columns depict the performance difference between the BR and VS systems. A positive figure in column 6 or 7 indicates that the BR system out-performs the VS system. The last column shows how many additional seconds (a positive number) are consumed by the BR system. As either information-flow mining or WordNet-based query expansion is conducted before the adaptive filtering processes take place, their computational time is not included in Tables V(a) and V(b).

The mean and standard deviation of F1 scores achieved by the BR system over the 50 TREC topics are −19.7 and 84.2, respectively. The mean and standard deviation of the F3 scores achieved by the BR system are 38.7 and 133.9, respectively. On the other hand, the mean and standard deviation of the F1

Table V(a). BR versus VS Based on the TREC-AP Collection (Topics 1–25)

BR's Performance					BR vs. VS		
Topic	Rel.Doc.	F1	F3	Time	ΔF1	ΔF3	ΔTime
1	344	−4	18	50452	−66	−153	32899
2	459	−227	14	45361	−10	−130	24216
3	220	−94	−2	50083	−60	−35	35252
4	94	−39	−7	35012	74	−98	22390
5	68	7	21	24879	−39	−62	12249
6	270	−96	−13	38635	−335	−530	23927
7	311	−5	0	19572	192	−54	2490
8	66	−32	−16	28064	176	66	15324
9	107	−48	−19	31424	−22	−91	18444
10	265	370	660	71335	586	403	55939
11	427	−23	151	46962	858	−161	30299
12	624	−6	22	27370	−103	−149	2097
13	86	51	88	24760	−96	−138	12208
14	116	−1	12	23274	22	−39	9742
15	107	−87	−41	35985	146	−52	22615
16	142	−69	23	36756	−112	−46	23285
17	224	−73	151	67759	−248	−89	52111
18	130	−41	−13	26162	1581	1550	13222
19	269	−3	6	25886	25	−50	9829
20	158	−53	6	37530	−119	−192	23851
21	29	−109	−27	34365	−49	−79	22519
22	1238	−38	496	75793	−16	132	42735
23	237	250	445	37725	−72	9	22628
24	300	−111	92	48147	−107	−11	30979
25	65	−43	−9	32608	60	−65	20133

scores achieved by the VS system over the 50 TREC topics are −113.5 and 544.2, respectively. The mean and standard deviation of F3 scores achieved by the VS system are 49.9 and 270.8, respectively. As a matter of fact, each TREC topic exhibits unique semantic features such as *clarity* [Cronen-Townsend et al. 2002], and therefore retrieval effectiveness may vary across topics. The clarity score measures the degree of ambiguity of a query with respect to a collection of documents, and is computed based on the relative entropy between the query and collection language models [Cronen-Townsend et al., ibid.]. It seems that a topic-by-topic comparison between the two systems will lead to a more meaningful analysis. There are 24 out of the 50 TREC topics in which the BR system performs better than or equal to the VS system in terms of F1 score. However, in terms of the F3 score, there are only 16 topics in which the BR system performs better than or equal to the VS system. By testing the hypotheses H_{null}: $\mu_{BR} - \mu_{VS} = 0$ and $H_{alternative}$: $\mu_{BR} - \mu_{VS} > 0$ with a paired one-tail t-test on the F1 scores, the null hypothesis is rejected at the 11% level of significance ($t(49) = 1.218$, $p = .11$). However, in terms of the F3 scores, a significant statistical difference between the BR and VS systems cannot be established. Therefore, a conclusive statement cannot be made if the filtering performance of the BR system is better than that of the VS system, based on the TREC-AP collection. However, their filtering performance is comparable.

Table V(b). BR versus VS Based on the TREC-AP Collection (Topics 26–50)

BR's Performance					BR vs. VS		
Topic	Rel.Doc.	F1	F3	Time	ΔF1	ΔF3	ΔTime
26	61	−25	0	25315	49	−13	13205
27	20	−5	5	40551	−8	1	28876
28	66	−22	4	37944	−8	1	25685
29	7	−4	−2	23646	12	6	12248
30	1	0	0	18845	2	1	7576
31	1	−22	−11	23449	8	4	12166
32	6	−2	−1	22858	−2	−1	11496
33	13	0	0	22917	−1	−3	11328
34	2	−8	−4	24293	−5	−6	12964
35	1	−2	−1	20848	5	0	9578
36	10	−7	4	27193	−7	−6	15757
37	7	11	18	19665	36	23	8282
38	276	−12	4	25000	−13	−4	7703
39	4	−4	−2	20103	19	1	8808
40	118	−29	−12	26093	309	−243	11092
41	30	−3	1	25459	0	0	13015
42	151	−33	−14	38163	62	11	22969
43	102	−133	−64	29452	−146	−83	16153
44	152	−4	−2	21682	7	−59	8422
45	52	−135	−50	46152	21	−7	33236
46	74	−7	4	25081	−110	−175	12715
47	89	0	0	23135	−10	−25	10058
48	30	4	7	21602	49	27	9686
49	55	−8	−4	19266	12	−59	6931
50	6	−11	−3	30170	−11	−3	18635

The reason why the BR system outperforms the VS for some TREC topics is that inference about a retrieval context can be conducted by the BR systems even though only a short query is available to the system initially (e.g., only a few terms extracted from the TREC title field are made available). For example, the following information flows for TREC topic 008 (Economic Projections) are obtained via text mining. These information flows are revised into the BR system's knowledge base before adaptive filtering takes place.

> economic⊕projections |− (administration:0.96 growth:0.90 percent:0.87 budget:0.80 forecast:0.78 deficit:0.76 inflation:0.73 national:0.70 analysts:0.68 spending:0.67 drought:0.65 sprinkel:0.65 prices:0.63 outlook:0.63 rates:0.61 interest:0.61 optimistic:0.60 federal:0.59 estimates:0.59 exports:0.56)

The notation |−. indicates the flow of information from left to right. As explained in Section 5.1, the aforementioned information flows are translated to explicit beliefs, such as ((economic $\wedge$ projections) $\rightarrow$ growth, 0.9), ((economic $\wedge$ projections) $\rightarrow$ forecast, 0.78), ((economic $\wedge$ projections) $\rightarrow$ inflation, 0.73), etc. In fact, all terms are stemmed by our systems. Nevertheless, for readability reasons, the nonstemmed terms are shown as examples in this article. As shown in this example, only 20 significant information flows of a topic are applied to the BR system.

A concrete example of how belief-based informational inference takes place in the BR system is follows: The TREC-AP document (AP880319-0174—New Financial Help Given Argentina), which is judged relevant for TREC topic 008 by the TREC experts, is retrieved by the BR system even though the terms "economic projections" do not appear in the document. The reason is that the positive beliefs about the information seeker's needs, such as (budget, 0.8), (forecast, 0.78), (deficit, 0.76), (inflation, 0.73), (interest, 0.61), etc., are inferred by the BR system based on the given information flows, and these beliefs just happen to be the characterization of the particular document (i.e., the corresponding terms are found in the document). On the other hand, the same document cannot be retrieved by the VS system because there is no overlapping between query terms ("economic projections") and the terms characterizing the document.

It should be noted that the VS system can also learn relevant terms such as "budget", "forecast", etc., and can revise these terms into its user profile after a period of learning. However, the VS system cannot work as effectively as the BR during the initial period of filtering because the former cannot conduct context-sensitive query expansion. One may wonder why query expansion based on WordNet [Miller et al. 1990] does not help in this case. For instance, the term "projection" has nine senses in WordNet (version 1.6). The first synonym set (the most common sense) is "prediction, anticipation, foresight", the second synonym set is "visual communication", and the third synonym set is "plan, program, programme", etc. Nevertheless, none of these senses leads to an expanded query containing terms, such as "forecast", "budget", "inflation", which are relevant with respect to the financial domain. In general, WordNet is a generic lexicon and it is very difficult to perform *context-sensitive* query expansion based on it. The proposed belief-based adaptive IR system (BR) is able to conduct context-sensitive informational inference to enhance retrieval effectiveness. However, the current information-flow mechanism is not perfect. As can be seen, the information flow ((economic $\wedge$ projections) $\rightarrow$ percent) is a very general implication, and may lead to some nonrelevant documents being retrieved. Such an information flow may be regarded as noise generated from the text mining process. This weakness partly explains why there is still large variation in terms of performance of the BR system across TREC topics.

Another interesting observation is about TREC topic 018 (Japanese Stock Market Trends). Although there is a significant boost in retrieval effectiveness, the negative F1 and F3 scores achieved by the BR system are not really outstanding. After careful examination of retrieval results of the VS system, we find that quite a number of nonrelevant documents which are about stock markets in countries other than Japan or stock information rather than market trends are retrieved by the VS system. As a result, its utility scores are much lower than those of the BR system. Although a very general term such as "market" will probably contribute zero to the overall document score in the VS system, other query terms (terms contained in the TREC title-field) such as "stock" contribute a positive document score. Since a relatively low startup dissemination threshold and a primitive threshold updating strategy are used in our systems, a document just about "U.S. stock market" will be assigned by the VS system a document score slightly higher than the dissemination threshold term. On the other hand, a noisy term such as “market” will make a negative contribution to the overall document score in the BR system according to our semantic information matching function, defined in Section 6.2. Therefore, a document will not be selected by the BR system unless it contains quite a high number of positive terms such as “Nikkei”, “Japan”, etc. As a result, fewer mistakes are made by the BR system for this relatively noisy topic.

As a whole, the performance of the BR system is encouraging, although a conclusive statement about its retrieval effectiveness cannot be made due to the high variances of the differences between the BR and VS systems across TREC topics. In general, the BR system is likely to perform well in a TREC topic if the information flows can accurately characterize the particular retrieval context of that topic. However, more research is required to identify the exact conditions under which the BR system will outperform the VS system. The BR system spent 32,695.6 seconds to process a TREC topic and 0.135 seconds for a document on average. The VS system spent 14,096.3 seconds to process a TREC topic and 0.058 seconds for a document on average. Although the VS system is faster than the BR, the belief-revision-based IR system is still remarkably efficient even for processing such a large document collection. According to the system log, the most computationally expensive operations of the BR system are the belief revision operations. Basically, each belief revision operation invokes the underlying theorem prover (SICStus Prolog) to check if the dominance property (EE2) is maintained in the BR system’s knowledge base. Several optimization techniques were applied to the BR system to make it scalable. For instance, only relatively entrenched beliefs are selected for the belief revision processes, since each belief revision operation is computationally expensive. Moreover, a belief revision operation will not be invoked unless the change of entrenchment degree of a belief is greater than a predefined system threshold. The frequency of invoking a learning cycle where term preference value computation and belief revision take place is also under tight control. Finally, by using the “anytime” algorithm for the implementation of belief revision the elapsed time of each learning cycle will not exceed a predefined time limit.

7.3 Evaluation Based on the Reuters-21578 Collection

The BR system was also evaluated based on the Reuters-21578 collection with the Lewis-Split subset, which contains 19,813 documents (13MB). An example of a Reuters-21578 document is shown in the following.

<REUTERS TOPICS=”YES” LEWISSPLIT=”TRAIN” CGISPLIT=”TRAINING-SET” OLDID=”5544” NEWID=”1”>
<DATE>26-FEB-1987 15:01:01.79</DATE>
<TOPICS><D>cocoa</D></TOPICS>
<TEXT>
<TITLE>BAHIA COCOA REVIEW</TITLE>
<BODY>Showers continued throughout the week in
the Bahia cocoa zone, alleviating the drought since early
. . . .</BODY></TEXT>
</REUTERS>

A Reuters-21578 document contains many fields, each marked up by a pair of tags. For instance, the main body of the news is delimited by <BODY> and </BODY> tags, and the Topic field is delimited by the <TOPICS> and </TOPICS> tags. To apply the TREC adaptive filtering task to the Reuters-21578 collection, a collection preprocessing procedure is required. For instance, we need to construct a relevance feedback file similar to the one used in TREC-7. The topic field of each newswire document is parsed to extract the topic codes representing particular Reuters-21578 topics. For example, if the commodity code "cocoa" is found in the <Topics> field, a relevance feedback record will be created. The relevance feedback record consists of the Reuters-21578 record ID, the system-generated topic ID corresponding to the topic code, and a constant "1" indicating that the document is relevant for the particular topic. Such a file format is exactly the same as the relevance feedback file used in TREC-7. Since the <Topics> field may contain more than one topic code, multiple relevance feedback records could be generated based on one Reuters-21578 document. If a combination of record ID and topic ID cannot be found from the relevance feedback file, the corresponding Reuters document is considered nonrelevant for the particular topic.

The Reuters-21578 topic file (part of the Reuters-21578 collection), which consists of 135 predefined topics, was used to construct the initial queries. Each topic code retrieved from the topic file is used to extract the corresponding topic description from the Reuters-21578 category description file. For instance, the corresponding topic description for the corporate code "acq" is "acquisitions". This topic description is used to develop the initial query (user profile) for that particular Reuters-21578 topic. Usually, a topic description consists of one or two terms. Only the first 20 topics (e.g., from "acq" to "corn") were used in our experiment. Similar to the TREC-AP-based experiments, an initial query in the BR system was contextualized by the information flows extracted before the adaptive filtering task took place. As for the VS system, each initial query was expanded with terms found from the synsets of WordNet. After collection preprocessing, the topic field of each Reuters-21578 document is removed, and so the topic codes are not available to our adaptive IR systems during the adaptive filtering processes. The F1 and F3 scores achieved by the BR system and the comparison between the BR and VS systems are depicted in Table VI. There are 14 out of the 20 Reuters topics (i.e., more than a half) in which the BR system performs better than or equal to the VS in terms of F1 or F3 scores.

The means and standard deviations of the F1 and F3 scores achieved by the BR system over the 20 Reuters topics are ($\mu = 161.8, \sigma = 471.5$) and ($\mu = 287.1, \sigma = 823.0$), respectively. By contrast, the means and standard deviations of F1 and F3 scores achieved by the VS system are ($\mu = 115.5, \sigma = 284.8$) and ($\mu = 192.0, \sigma = 489.1$), respectively. By testing the hypotheses $\mathrm{H}_{\mathrm{null}} : \mu_{\mathrm{BR}} - \mu_{\mathrm{VS}} = 0$ and $\mathrm{H}_{\mathrm{alternative}} : \mu_{\mathrm{BR}} - \mu_{\mathrm{VS}} > 0$ with a paired one-tail t-test on F3 scores, the null hypothesis is rejected at a 11% level of significance ($t(19) = 1.26$, $p = .11$). However, the statistical difference between the BR and VS systems in terms of F1 scores can only be established at a 15% level of significance ($t(19) = 1.07$, $p = .15$). Although a conclusive statement cannot be made as to whether the filtering performance of the BR system is better than

Table VI. BR versus VS Based on the Reuters-21578 Collection

BR's Performance					BR vs. VS		
Topic	Rel.Doc.	F1	F3	Time	ΔF1	ΔF3	ΔTime
acq	2366	2096	3673	21142	848	1489	20109
alum	57	24	52	734	−24	−22	405
austdlr	4	0	0	408	0	0	142
austral	0	−6	−3	397	2	1	147
barley	51	120	190	411	43	44	100
bfr	0	0	0	362	0	0	112
bop	105	0	16	612	−50	−134	244
can	3	−29	−12	385	−17	−6	119
carcass	68	26	58	918	17	41	546
castor-meal	0	0	0	386	0	0	134
castor-oil	2	0	0	399	0	0	143
castorseed	1	3	4	2251	0	0	1987
citruspulp	1	3	4	2369	2	1	2110
cocoa	73	174	262	726	7	31	378
coconut	6	−22	4	433	−34	−17	172
coconut-oil	7	0	0	447	11	−2	178
coffee	139	267	476	3853	−40	60	3424
copper	65	111	218	734	23	84	387
copra-cake	3	−4	−2	439	−4	−2	175
corn	237	472	801	1351	141	333	903

that of the VS system, their filtering performance is once again comparable. The BR system spent 1,937.9 seconds to process a Reuters-21578 topic and 0.098 seconds for a document on average. The VS system spent 342.1 seconds to process a topic and 0.017 seconds for a document on average.

Similar to the situation when adaptive filtering was applied to the TREC-AP collection, context-sensitive information inference is the main reason why the BR system outperforms the VS for some Reuters-21578 topics. For example, the following information flows for the topic "acq" (acquisitions) were extracted from the training set of the Reuters-21578 collection.

> acquisitions |− (company:0.871 corp:0.845 dlrs:0.842 offer:0.714 shares:0.709 stock:0.684 merger:0.625 bank:0.614 shareholders:0.585 agreement:0.583 buy:0.577 tender:0.577 board:0.571 purchase:0.558 international:0.552 proposed:0.549 cash:0.541 completed:0.540 debt:0.538 bid:0.523)

The BR system can make use of beliefs such as (acquisitions → company, 0.871), (acquisitions → offer, 0.714), (acquisitions → shares, 0.709), (acquisitions → stock, 0.684), (acquisitions → merger, 0.625), etc., to contextualize the initial query "acquisitions". For the Reuters newswire document (96 – Investment Firms Cut Cyclops Stake), neither the term "acquisitions" nor the synonyms extracted from WordNet appear in the document. However, the document is characterized by terms such as "shares" and "stock". Therefore, the BR system was able to retrieve the document. On the other hand, the VS system was not able to retrieve the same document because there was no overlapping between its query vector and the document vector during the initial period of adaptive filtering. Similarly, document (128 – Liebert Corp Approves Merger) is characterized by terms such as “merger”, “shareholders”, “stock”, “shares”, etc., rather than the term “acquisitions”. The BR system was able to retrieve this document because the positive beliefs (shares, 0.709), (merger, 0.625), (shareholders, 0.585), etc., were inferred by the system as representing the current retrieval situation. On the other hand, the same document was rejected by the VS system, since there was no overlap between its query vector and the document vector.

7.4 Usability Study for the Belief-Based Adaptive IR System

In order to evaluate the explanatory power offered by the belief-based adaptive IR system, an experiment was conducted to compare the explanation capability of the BR system and that of the VS system. The explanatory power of an IR system refers to the ability of the system to provide some cues about its IR decisions. Traditional quantitative IR systems only provide document scores (e.g., retrieval status values, or RSVs) as a means of justifying their IR decisions. However, this black-box IR approach may not be sufficient to support mission-critical IR activities (e.g., retrieving information about possible terrorist attacks) because information seekers would like to see more justifications for the IR results to avoid paying the high cost of a false alarm.

A group of 20 college students taking the elective course of web intelligence participated in this experiment, which is part of the usability study for our belief-based adaptive IR system. These students attended six weeks of lectures in IR theory and techniques before participating in the usability experiment. One group of students used the BR system (i.e., experimental group) to retrieve documents from a small collection, and the other group (i.e., control group) used the VS system to conduct the same task. At the beginning of the experiment, 10 subjects were randomly chosen and assigned to the experimental group, and the remaining subjects were assigned to the control group. The test collection comprised 50 ACM technology news articles (*TechNews*) covering various topics in IT; standard document preprocessing procedures such as stop word removal and stemming were applied. *ACM TechNews* is delivered to its subscribers three times a week. An example from *TechNews* is shown next.

> “Softbots Stride Forward - The rapid progress of intelligent agent, or ”softbot,” technology has scientists convinced that people who do not have time to participate in conferences or perform business activities will be able to employ computer-generated avatars that act on their behalf within a decade. . . . ”

User profiles of both the BR and VS systems were initialized with the topic “intelligent agents”. A training set of *TechNews*, which was different from the test set, was used to develop contextual knowledge about the query. When the adaptive filtering task began, each subject used her/his assigned IR system to retrieve documents, and reviewed the system’s justifications for document selection. For the BR system, the user interface consists of an explanation window as depicted in Figure 3. The user interface of the VS system is similar to that of the BR system, except that its explanation window only displays a single-document selection score. After reviewing all 50 documents, a questionnaire was distributed to each participant to collect information about his/her perceived usability of the IR systems. From among the 10 response items appearing on the questionnaire, 3 are related to the dependent variable of "explanatory power" of an IR system, and the variable is measured using a 5-point semantic differential scale of strongly agree (5), agree (4), neutral (3), disagree (2), and strongly disagree (1). The three question items are: (a) "The IR system provides the cue why a document is retrieved"; (b) "The IR system provides an explanation of document selection"; and (c) "The IR system can justify why a document is selected." The response rate of the study was 100%. The overall mean scores and standard deviations of the three question items are ($\mu_{\text{BR}} = 4.03$, $\sigma_{\text{BR}} = 0.21$) and ($\mu_{\text{VS}} = 2.17$, $\sigma_{\text{VS}} = 0.35$) for the BR and VS systems respectively. By testing the hypotheses $\text{H}_{\text{null}} : \mu_{\text{BR}} - \mu_{\text{VS}} = 0$ and $\text{H}_{\text{alternative}} : \mu_{\text{BR}} - \mu_{VS} > 0$ with a paired one-tail t-test, the null hypothesis was rejected ($\text{t}(2) = 8.54$, $\text{p} < .01$). Therefore, we conclude that the explanatory power of the BR system is higher than that of the VS system according to this usability study.

8. CONCLUSIONS AND FUTURE WORK

The AGM belief revision logic provides a sound and rigorous framework to develop adaptive IR systems. In particular, the logic offers sufficient expressive power to represent IR contexts, and also provides a sound inference mechanism to model the nonmonotonicity of information matching arising in changing retrieval contexts. On the other hand, information-flow-based text mining allows IR systems to discover contextual IR knowledge autonomously. The induction power brought by information-flow-based text mining is complementary to the nonmonotonic reasoning capability offered by the belief revision system. One distinct advantage of the belief-based adaptive IR system is that its learning and matching behavior can be understood based on the axioms characterizing the AGM logic. Our experiments show that the belief-based adaptive IR system achieves comparable performance to a classical adaptive IR system for the TREC-AP and Reusters-21578 collections. The belief-based adaptive IR system is efficient enough to deal with large and complex IR tasks, and its explanatory power is confirmed by our usability study. To our best knowledge, this is the first comprehensive empirical evaluation of a logic-based adaptive IR model based on large IR benchmark collections, and yet encouraging performance is observed. Future work includes selectively (e.g., based on the clarity score) applying text mining to enrich a retrieval context so that overall retrieval effectiveness can be improved. Moreover, more sophisticated thresholding techniques will be exploited to bootstrap the performance of the belief-based adaptive IR system.

ACKNOWLEDGMENTS

The authors would like to thank the three anonymous reviewers for valuable comments. Also thanks to Professor Mary-Anne Williams for her explanation of the Maxi-adjustment transmutation method on several occasions.

REFERENCES

ALCHOURRÓN, C. E., GÄRDENFORS, P., AND MAKINSON, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *J. Symbol. Logic 50*, 510–530.

AMATI, G. AND GEORGATOS, K. 1996. Relevance as deduction: A logical view of information retrieval. In *Proceedings of the 2nd Workshop on Information Retrieval, Uncertainty and Logic*, 21–26.

ARAMPATZIS, A. AND VAN HAMEREN, A. 2001. The score-distributional threshold optimization for adaptive binary classification tasks. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 285–293.

BELKIN, N. AND CROFT, W. 1992. Information filtering and information retrieval: Two sides of the same coin? *Commun. ACM 35*, 12, 29–38.

BESSANT, B., GREGOIRE, E., MARQUIS, P., AND SADDOTIS, L. 1998. Combining nonmonotonic reasoning and belief revision: A practical approach. In *Proceedings of the 8th International Conference on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA)*. Lecture Notes in Artificial Intelligence, vol. 1480. Springer, 115–128.

BRUZA, P. D. AND SONG, D. 2002. Inferring query models by information flow analysis. In *Proceedings of the 11th International ACM Conference on Information and Knowledge Managment (CIKM)*. 260–269.

BRUZA, P., SONG, D., AND WONG, K. 2000. Aboutness from a commonsense perspective. *J. Amer. Soc. Inf. Sci. 51*, 12, 1090–1105.

BRUZA, P. AND VAN LINDER, B. 1998. Preferential models of query by navigation. In *Information Retrieval: Uncertainty and Logics*, M. Lalmas et al., eds. Kluwer Academic, Norwell, MA, 73–96.

BRUZA, P. AND HUIBERS, T. 1994. Investigating aboutness axioms using information fields. In *Proceedings of the 17th Annual International Conference on Research and Development in Information Retrieval (SIGIR)*, 112–121.

BURGESS, C., LIVESAY, K., AND LUND, K. 1998. Explorations in context space: Words, sentences, discourse. *Discourse Process. 25*, 211–257.

CALLAN, J. 1998. Learning while filtering documents. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 224–231.

CARD, S., MORAN, T., AND NEWELL, A. 1983. *The Psychology of Human Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ.

CHIARAMELLA, Y. AND CHEVALLET, J. 1992. About retrieval models and logic. *Comput. J. 35*, 3, 233–242.

COHEN, W. AND SINGER, Y. 1996. Context-Sensitive learning methods for text categorization. In *Proceedings of the 19th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR)*, 307–315.

COOL, C. 2001. The concept of situation in information science. *Annual Rev. Inf. Sci. Technol. 35*, 5–42.

COOL, C. AND SPINK, A. 2002. Issues of context in information retrieval (IR): An introduction to the special issue. *Inf. Process. Manage. 38*, 5, 605–611.

CRESTANI, F. AND VAN RIJSBERGEN, C. 1995. Information retrieval by logical imaging. *J. Documentat. 51*, 1, 3–17.

CRESTANI, F. 1998. Logical imaging and probabilistic information retrieval. In *Information Retrieval: Uncertainty and Logics*, M. Lalmas et al., eds. Kluwer Academic, Norwell, MA, 247–279.

CRONEN-TOWNSEND, S., ZHOU, Y., AND CROFT, W. 2002. Predicting query performance. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 299–306.

DALAL, M. 1988. Investigations into a theory of knowledge base revision: Preliminary report. In *Proceedings of the 7th Natonal Conference on Artificial Intelligence*, 475–479.

DRETSKE, F. 1981. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.

DUBOIS, D. AND PRADE, H. 1995. Possibility theory as a basis for qualitative decision theory. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1924–1932.

DUBOIS, D. AND PRADE, H. 1991. Epistemic entrenchment and possibilistic logic. *Artif. Intell. 50*, 2, 223–239.

GÄRDENFORS, P. 2000. *Conceptual Spaces: The Geometry of Thought.* MIT Press, Cambridge, MA.

GÄRDENFORS, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, MA.

GÄRDENFORS, P. AND MAKINSON, D. 1994. Nonmonotonic inference based on expectations. *Artif. Intelli. 65*, 2, 197–245.

Gärdenfors, P. and Makinson, D. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning and Knowledge (TARK)*, 83–95.

Hirst, G. and St-Onge, D. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In *WordNet: An Electronic Lexical Database*, C. Fellbaum, ed. MIT Press, Cambridge, MA, 305–332.

Huibers, T. and van Linder, B. 1996. Formalising intelligent information retrieval agents. In *Proceedings of the 18th British Computer Society Annual Information Retrieval Colloqium*, 125–143.

Huibers, T. and Wondergem, B. 1998. Towards an axiomatic aboutness theory for information retrieval. In *Information Retrieval: Uncertainty and Logics*, M. Lalmas et al., eds. Kluwer Academic, Norwell, MA, 297–318.

Hull, D. 1998. The TREC-7 filtering track: Description and analysis. In *Proceedings of the 7th Text Retrieval Conference*, 33–56.

Hunter, A. 1996. Intelligent text handling using default logic. In *Proceedings of the IEEE Conference on Tools with Artificial Intelligence (TAI)*, 34–40.

Hunter, A. 1997. Using default logic for lexical knowledge. In *Proceedings of the 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning*. Lecture Notes in Computer Science, vol. 1244. Springer, 322–335.

Ingwersen, P. and Järvelin, K. 2005. SIGIR workshop reports: Information retrieval in context: IRiX. *ACM SIGIR Forum 39*, 2, 31–39.

Kindo, T., Yoshida, H., Morimoto, T., and Watanabe, T. 1997. Adaptive personal information filtering system that organises personal profiles automatically. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, 716–721.

Koller, D. and Sahami, M. 1997. Hierarchically classifying documents using very few words. In *Proceedings of the 14th International Conference on Machine Learning (ICML)*, 170–178.

Lalmas, M. and Bruza, P. 1998. The use of logic in information retrieval modeling. *Knowl. Eng. Rev. 13*, 3, 263–295.

Lalmas, M. and van Rijsbergen, C. (1992. A logical model of information retrieval based on situation theory. In *Proceedings of the 14th British Computer Society Annual Information Retrieval Colloqium*, 1–13.

Landauer, T., Foltz, P., and Laham, D. 1998. An introduction to latent semantic analysis. *Discourse Process. 25*, 259–284.

Lang, J. 1997. Possibilistic logic: Algorithms and complexity. In *Handbook of Algorithms for Uncerainty and Defeasible Reasoning*. J. Kohlas and S. Moral, eds., Kluwer Academic, Norwell, MA.

Lau, R., ter Hofstede, A., and Bruza, P. 1999. A study of belief revision in the context of adaptive information filtering. In *Proceedings of the 5th International Computer Science Conference on Internet Applications*. L. C. K. Hui and D. L. Lee, eds. Lecture Notes in Computer Science, vol. 1749. Springer, 1–10.

Lau, R., ter Hofstede, A., and Bruza, P. 2001. Maxi-Adjustment and possibilistic deduction for adaptive information agents. *J. Appl. Non-Classic. Logics 11*, 1-2, 169–201.

Lau, R. 2002. The state of the art in adaptive information agents. *Int. J. Artif. Intelli. Tools 11*, 1, 19–61.

Lau, R. 2003. Context sensitive text mining and belief revision for adaptive information retrieval. In *Proceedings of the 2nd IEEE/WIC International Conference on Web Intelligence (WI)*, Halifax, Canada. IEEE Press, 256–262.

Lau, R., Bruza, P., and Song, D. 2004. Belief revision for adaptive information retrieval. In *Proceedings of the 27th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR)*, Sheffield, UK, 130–137.

Lawrence, S. 2000. Context in Web search. *IEEE Data Eng. Bull. 23*, 3, 25–32.

Levy, D. 2005. Users and interaction track: Memex and hypertext: To grow in wisdom: Vannevar Bush, information overload, and the life of leisure. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, 281–286.

LIBERATORE, P. AND SCHAERF, M. 1998. The complexity of model checking for propositional default logics. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI)*, 18–22.

LOGAN, B., REECE, S., AND JONES, K. 1994. Modelling information retrieval agents with belief revision. In *Proceedings of the 17th Annual International Conference on Research and Development in Information Retrieval* , 91–100.

LOSADA, D. AND BARREIRO, A. 1999. Using a belief revision operator for document ranking in extended boolean models. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 66–72.

LOSADA, D. AND BARREIRO, A. 2001. A logical model for information retrieval based on propositional logic and belief revision. *Comput. J. 44*, 5, 410–424.

LOSADA, D. AND BARREIRO, A. 2003a. Embedding term similarity and inverse document frequency into a logical model of information retrieval. *J. Amer. Soc. Inf. Sci. Technol. 54*, 4, 285–301.

LOSADA, D. AND BARREIRO, A. 2003b. Propositional logic representations for documents and queries: A large-scale evaluation. In *Proceedings of the 25th European Conference on IR Research (ECIR)*. Lecture Notes in Computer Science, vol. 2633. Springer, 219–234.

LOWE, W. 2001. Towards a theory of semantic space. In *Proceedings of the 23rd Annual Conference of the Cognitive Society*, 576–581.

MAES, P. 1994. Agents that reduce work and information overload. *Commun. ACM 37*, 7, 31–40.

MEGHINI, M., SEBASTIANI, F., STRACCIA, U., AND THANOS, C. 1993. A model of information retrieval based on terminological logic. In *Proceedings of the 16th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR)*, 298–307.

MILLER, G. A., BECKWITH, R., FELLBAUM, C., GROSS, D., AND MILLER, K. 1990. Introduction to WordNet: An online lexical database. *J. Lexicograph. 3*, 4, 234–244.

NIE, J. 1989. An information retrieval model based on modal logic. *Inf. Process. Manage. 25*, 5, 477–491.

NIE, J. 1986. An outline of a general model for information retrieval systems. In *Proceedings of the 9th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR)*, 495–506.

NIE, J., BRISEBOIS, M., AND LEPAGE, F. 1995. Information retrieval as counterfactual. *Comput. J. 38*, 8, 643–657.

NEVES, R., BONNEFON, J., AND RAUFASTE, E. 2000. Rationality in human nonmonotonic inference. In *Proceedings of the 8th International Workshop on Nonmonotonic Reasoning (NMR)*. http://xxx.lanl.gov/html/cs.AI/0003073.

RAUFASTE, R. AND NEVES, R. 1998. Empirical evaluation of possibility theory in human radiological diagnosis. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI)*, 124–128.

ROBERTSON, S. 2002. Comparing the performance of adaptive filtering and ranked output systems. *Inf. Retriev. 5*, 257–268.

ROCCHIO, J. 1971. Relevance feedback in information retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*. G. Salton, ed. Prentice Hall, Upper Saddle River, NJ, 313–323.

SALTON, G. 1991. Developments in automatic text retrieval. *Sci. 253*, 974–980.

SALTON, G. AND MCGILL, M. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

SEBASTIANI, F. 1998. On the role of logic in information retrieval. *Inf. Process. Manage. 34*, 1, 1–18.

SONG, D. AND BRUZA, P. 2001. Discovering information flow using a high dimensional conceptual space. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 327–333.

SONG, D. AND BRUZA, P. 2003. Towards a theory of context sensitive informational inference. *J. Amer. Soc. Inf. Sci. Technol. 54*, 4, 326–339.

VAN RIJSBERGEN, C. 1986. A nonclassical logic for information retrieval. *Comput. J. 29*, 6, 481–485.

WILLIAMS, M. 1997. Anytime belief revision. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, 74–79.

WILLIAMS, M. 1995. Iterated theory base change: A computational model. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1541–1547.
WONG, K. F., SONG, D., BRUZA, P. D., CHENG, C. H. 2001. Application of aboutness to functional benchmarking in information retrieval. *ACM Trans. Inf. Syst. 19*, 4, 337–370.
ZHANG, Y. AND CALLAN, J. 2001. Maximum likelihood estimation for filtering thresholds. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 294–302.

Received December 2005; revised November 2006; accepted June 2007