# Application of Aboutness to Functional Benchmarking in Information Retrieval

KAM-FAI WONG
The Chinese University of Hong Kong
DAWEI SONG, PETER BRUZA
University of Queensland
and
CHUN-HUNG CHENG
The Chinese University of Hong Kong

Experimental approaches are widely employed to benchmark the performance of an information retrieval (IR) system. Measurements in terms of recall and precision are computed as performance indicators. Although they are good at assessing the retrieval effectiveness of an IR system, they fail to explore deeper aspects such as its underlying functionality and explain why the system shows such performance. Recently, inductive (i.e., theoretical) evaluation of IR systems has been proposed to circumvent the controversies of the experimental methods. Several studies have adopted the inductive approach, but they mostly focus on theoretical modeling of IR properties by using some metalogic. In this article, we propose to use inductive evaluation for functional benchmarking of IR models as a complement of the traditional experiment-based performance benchmarking. We define a functional benchmark suite in two stages: the evaluation criteria based on the notion of "aboutness," and the formal evaluation methodology using the criteria. The proposed benchmark has been successfully applied to evaluate various well-known classical and logic-based IR models. The functional benchmarking results allow us to compare and analyze the functionality of the different IR models.

Categories and Subject Descriptors: H.1.1 [**Models and Principles**]: Systems and Information Theory; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*retrieval models*; *search process*; *selection process*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*performance evaluation* (*efficiency and effectiveness*); H.3.m [**Information Storage and Retrieval**]: Theoretical Study of Information Retrieval

General Terms: Measurement, Performance, Theory

Additional Key Words and Phrases: Aboutness, functional benchmarking, inductive evaluation, logic-based information retrieval

This project is partially supported by the Chinese University of Hong Kong's strategic grant (project ID: 44M5007), and by the Cooperative Research Centres Program through the Department of the Prime Minister and Cabinet of Australia.

Authors' addresses: K.-F. Wong, C.-H. Cheng, Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; email: {kfwong, ccheng}@se.cuhk.edu.hk; D. Song, P. Bruza, Distributed Systems Technology Centre, Building 78, Staff House Road, University of Queensland, Qld 4072 Australia; email: {dsong, bruza}@dstc.edu.au.

## 1. INTRODUCTION

The information retrieval (IR) problem can be described as a quest to find the set of relevant information objects (i.e., documents $D$) corresponding to a given information need, represented by a query $Q$. The assumption is that the query $Q$ is a good description of the information need $N$. An often used premise in IR is the following: if a given document $D$ is *about* the request $Q$, then there is a high likelihood that $D$ will be relevant with respect to the associated information need. Thus the information retrieval problem is reduced to deciding the aboutness relation between documents and queries.

Articles on aboutness have appeared sporadically in the literature for more than two decades. Hutchins [1977] provides a thoughtful early study of the topic. This account attempts to define a notion of aboutness in terms of a combination of linguistic and discourse analyses of a text. At a high level of information granularity, such as a sentence, Hutchins introduces *themes* and *rhemes* as the carriers of the thematic progression of a text. Roughly speaking, the theme states what the writer intends to express in the sentence (i.e., what it is about), and the rheme is the "new" information. Thematic elements of a sentence are typically bound textually to the preceding text, or assumed as given within the current context. Hutchins also considers how sequences of sentences combine to form textual elements of lower information granularity such as an episode. In other words, sentences are considered to be a part of the micro structure of the text, whereas an episode is considered to be an element of its macrostructure. Themes and rhemes can be generalized to the macro level. Hutchins asserts, "The thematic part of the text expresses what the text is 'about,' while the rheme expresses what the author has to say about it" [Hutchins 1977, p. 31].

Maron [1977] tackled aboutness by relating it to a probability of satisfaction. Three types of aboutness were characterized: S-about, O-about, and R-about. S-about (i.e., subjective about) is a relationship between a document and the resulting inner experience of the user. O-about (i.e., objective about) is a relationship between a document and a set of index terms. More specifically, a document $D$ is about a term set $T$ if user $X$ employs $T$ to search for $D$. R-about (i.e., retrieval about) purports to be a generalization of O-about to a specific user community (i.e., a class of users). Let $t_i$ be an index term and $D$ be a document; then "$D$ is R-about $t_i$" is defined as the ratio between the number of users satisfied with $D$ when using $t_i$ to formulate the request for information and the number of users satisfied by $D$. Using this as a point of departure, Maron further constructs a probabilistic model of R-aboutness. The advantage of this is that it leads to an operational definition of aboutness which can then be tested experimentally. However, once the step has been made into the probabilistic framework, it becomes difficult to study properties of aboutness; for example, how does R-about behave under conjunction? By way of illustration, assume document $D$ is characterized by the index terms $K_1, \ldots, K_n$. From a logical point of view, $D$ can be viewed as being represented by the conjunction $K_1 \wedge \cdots \wedge K_n$. Assume that $D$ is R-about term $t_i$. One can translate this relationship between a document and term into a relation between the document representation $K_1 \wedge \cdots \wedge K_n$ and term $t_i$ (now viewed as an atomic logical formula). What happens to the aboutness relationship if information, represented by the term $K_{n+1}$ is added to document $D$: is $K_1 \wedge \cdots \wedge K_n \wedge K_{n+1}$ about $t_i$? In other words, is aboutness monotonic with respect to the composition of information? Such questions cannot be answered within a probabilistic framework. The underlying problem relates to the fact that probabilistic independence lacks properties with respect to conjunction and disjunction. In other words, one's hands are largely tied when trying to express qualitative properties of aboutness within a probabilistic setting. (For this reason Dubois et al. [1997] developed a qualitative framework for relevance using possibility theory).

During the 1980s and early 1990s, the issue of aboutness remained hidden in the operational definitions of various retrieval models and their variations. For example, the vector space model represents both documents and queries as vectors in a high-dimensional space whereby the dimensions correspond to information-bearing terms. If the angle between the respective document and query vectors is above a certain threshold, the document is deemed to be about the query. This period also featured the emergence of sophisticated probabilistic retrieval models. Major effort was expended in producing ever more sophisticated matching functions between document and query representations. Such matching functions were evaluated by an experimental paradigm. The paradigm often has the following form. Given a set of test queries and a collection of documents, a set of relevant documents are a priori associated with each test query. In the actual experiment, a matching function produces a ranked list of documents descending on match score between a test query and a particular document representation. The performance of a matching function can be measured by studying the degree to which relevant documents are moved towards the top of the ranking produced by the matching function under observation. Statistical tests of significance can be applied to compare average performances of two ranking functions across the set of test queries, thus gaining some confidence that matching function A produces, on average, better rankings than matching function B. The experimental paradigm has long been one of the cornerstones of research into information retrieval, but it has long been debated as well. It is outside the scope of this article to descend into the controversies surrounding experimental information retrieval, but we illustrate one of its manifestations. Many of the more sophisticated matching functions rely on constants. The values of these constants can greatly influence the performance of the matching function. The specific values of the constants are not derived from theory, but are "tuned" according to a particular document collection and test query set.

The emergence of logic-based information retrieval in the mid-1980s allowed the matching function between document and query to be seen in a new light. In one of the founding papers Van Rijsbergen [1986a] states, "The single primitive operation to aid retrieval is one of uncertain implication." In other words, retrieval could be viewed as a process of plausibly inferring the query from the document. This view spawned a number of attempts at implementing logic-based retrieval systems (see Lalmas and Bruza [1998] for a survey and Crestani et al.[1998] for a compendium). Logic-based information retrieval also provided the framework to allow theoretical, rather than experimental, investigations in

IR [Sebastiani 1998]. It planted the seed for fundamental investigations of the nature of aboutness [Bruza and Huibers 1994, 1996; Hunter 1996; Nie et al. 1995] culminating in an axiomatic theory of information retrieval [Huibers 1996] and a characterization of aboutness in the terms of commonsense rules [Bruza et al. 2000a]. Aboutness theory has also recently appeared in the context of information discovery [Proper and Bruza 1999]. Broadly speaking, these works view information retrieval as a reasoning process, determining aboutness between two information carriers (e.g., a document about a query, or a document about a document). Work in this area attempted to symbolically characterize qualitative aspects of the matching function, which, up to that point, were normally hidden in the numeric expressions of these functions. In a broad sense an attempt was made to flesh out the assumptions underpinning matching functions, and more generally to provide a symbolic IR-centric account of "the most important relationship in IR—the one in which one object contains information *about* another" (italics ours) [Van Rijsbergen 1993]. An important consequence of logic-based information retrieval was that it allowed IR to be studied symbolically within a neutral framework; for example, researchers were free to posit questions such as: Is aboutness transitive, or is the aboutness relationship preserved under the composition of information? Once properties of aboutness are described by a set of postulates, they can be used to compare IR models depending on which aboutness postulates they support [Bruza and Huibers 1994; Huibers 1996; Bruza et al. 2000a]. This opens the door to an inductive, rather than experimental, theory of comparing matching functions. The development of an inductive theory of information retrieval evaluation parallels a similar development in the area of nonmonotonic reasoning. Throughout the 1980s a number of logics were proposed to model commonsense reasoning, for example, default logic, autoepistemic logic, circumscription and so on. At that time, there was no way to compare these different logics until the metatheory of nonmonotonic reasoning appeared [Kraus et al. 1990]. This theory embodied a suite if desired properties of nonmonotonic logic in terms of rules interpreted in a neutral framework (in this case, preferential models). By using this framework, the previously mentioned logics could be compared according to which properties they supported.

The theoretical analysis and comparison of information retrieval models need not take place within a logic-based framework. Losee [1997, 1998] provides an analytic theory. He states that a theory of the operation of text filtering and retrieval systems should describe current performance, predict future performance, and explain performance. The difference between Losee's analytical theory and the logic-based inductive theory is more in approach and scope rather than philosophical point of departure. Both aim to gain understanding of why particular IR systems perform the way they do. Losee's analytic theory is statistically based. Measures such as the average search length (ASL, expected position of a relevant document) are used to analyse the quality of a ranking of documents in the context of a hypothesized database. For example, ASL can be plotted against the probability that a given term is in a relevant document yielding a surface. It has been shown that when this probability increases, the ASL steadily and more strongly decreases due to the increase in discrimination power of the terms. This is reflected in the plots by pivoting of the surface away from the median (random) performance of ASL. In this way, the Boolean and probabilistic retrieval models have been scrutinized from a theoretical point of view [Losee 1997]. In contrast to Losee's analytical theory, the logic-based inductive theory focuses primarily on describing the aboutness properties embodied by a given matching function, and analyzing and comparing matching functions according to which aboutness postulates they support. "*Functional benchmarking*" is the general term coined for such analysis [Song et al. 1999].

The primary objective of this article is to propose a formal methodology for functional benchmarking and apply it to inductive evaluation and comparison of various typical IR models. Our evaluation targets in this article were deliberately chosen to review the practicality of the proposed benchmark. We have evaluated and compared the functionality of the more prominent classical and logical IR models—Boolean, naïve (i.e., zero-threshold and binary) vector space, threshold vector space (multivalued), probabilistic, situation theory-based, naïve (i.e., zero-threshold and binary) possible world-based and threshold possible world-based (multivalued) IR models. The advantages and disadvantages of the properties inherent to these models and how these properties affect effectiveness are analyzed. Furthermore, some important experimental results could be explained theoretically via the benchmarking. It is hoped this will shed light on existing IR models and help further research towards more effective IR models.

The rest of the article is organized as follows. In the next section (i.e., Section 2), the definition of the functional benchmark is outlined. The benchmark is based on the aboutness framework proposed by Bruza et al. [Bruza and Huibers 1994, 1996; Bruza et al. 2000a]. A formal functional benchmarking methodology is also proposed in this section. Sections 3 and 4 then present the evaluation of some classical [Van Rijsbergen 1979; Salton 1988, etc.] and logical IR models [Bruza and Lalmas 1996; Lalmas 1998; Lalmas and Bruza 1998], respectively, using the proposed benchmark. Finally, a conclusion including a summary of the evaluation results is given in Section 5.

## 2. DEFINING THE FUNCTIONAL BENCHMARK SUITE

Our approach in defining the functional benchmark suite is performed in stages. (a) We first identify a set of aboutness properties, which are used to analyze matching functions. They are used as the evaluation criteria for the functional benchmark. (b) We then define a formal methodology outlining the steps to perform inductive evaluation.

### 2.1 Property of Aboutness

Despite several research studies devoted to aboutness, there is still no consensus on the desirable properties of the aboutness relation. Nonetheless, a number of properties are commonly discussed in the literature, for example, reflexivity, transitivity, symmetry, simplification, supraclassicality, equivalence, and right weakening and left (right) monotonicity [Lalmas and Bruza 1998]. The primary reason for the lack of consensus is the fact that the logic-based framework chosen has some influence on the associated aboutness properties. One would think that reflexivity, that is, the assumption that an information carrier (such as a document) is about itself, would not generate any difference in opinion. However, reflexivity is a property *not* supported by Hunter's [1996] default logic-based aboutness framework, but *is* supported by Huibers' [1996] situation-theoretic framework. In addition, a substantial body of work on defining aboutness properties has been inspired by symbolic characterizations of the preferential entailment relation[1] found in nonmonotonic reasoning. This has slanted the corresponding characterizations of aboutness [Bruza and Huibers 1994, 1996; Amati and Georgatos 1996; Bruza and Van Linder 1998]. Recent work has argued that the aboutness relationship goes beyond the notion of preferential entailment [Bruza et al. 2000a].

The attempts in the literature to characterize the aboutness relationship have been useful to stimulate investigation into what "aboutness" really is without being burdened by the baggage of a particular retrieval model. An unfortunate consequence of this freedom has been a lack of connection with commonly accepted notions of IR performance. We argue that aboutness properties selected for the purposes of functional benchmarking should be able to be related to the traditional IR performance criteria: Precision[2] and Recall.[3] This allows theoretical insights provided by the inductive evaluation to be correlated with insights gleaned via experimental evaluation.

The inductive evaluation paradigm requires that the aboutness properties be expressed symbolically. This in turn requires that a conceptual framework be established which provides a sufficient diversity of concepts with which useful aboutness properties can be expressed. In this regard, Lalmas and Bruza [1998] have stated: "The framework should not be biased towards any given model, i.e., it should be neutral. Moreover, it should be sufficiently abstract to filter away unnecessary details of the various IR models. In such an abstract and neutral setting, IR models can be inductively compared."

In this article, we employ the framework proposed by Bruza et al. [Bruza and Huibers 1994, 1996; Bruza et al. 2000a]. This framework is abstract and not biased towards any given IR retrieval model, and is parsimonious with respect to the number of underlying concepts. Moreover, it is based on notions drawn from information-based logic. It would seem reasonable to build on research from this area if one accepts that determining whether a document is about a query, involves an information-based reasoning process.

In the framework, descriptors, documents, and queries share the same notion of information carriers. Given two information carriers $i$ and $j$, the aboutness between $i$ and $j$ (i.e., $i$ is about $j$) is denoted by a binary relation $\models$; that is, $i \models j$. On the other hand, $i \not\models j$ denotes "$i$ is not about $j$." For example, assuming an animal context, "penguin" is about "birds," but "penguin" is not about "flying."

[1] The term "migration" preferentially entails "salmon" if and only if all preferred documents on migration are also about salmon. That is, the user's information need is assumed to impose a preferential ordering on the set of underlying documents.

[2] Precision is defined as the ratio of relevant retrieved documents to retrieved documents.

[3] Recall is the ratio of relevant retrieved documents to relevant documents.

Information carriers can be composed. The composition of information is denoted by $i \oplus j$, which contains the information carried by both $i$ and $j$. It can be conceived of as a form of informational "meet." Viewed from a situation-theoretic perspective [Lalmas 1996], the information composition represents the intersection between the situations supporting $i$ and the situations supporting $j$. For example, *flying* $\oplus$ *bird* represents the intersection of "flying" situations and "bird" situations, that is, the situations that support the information, "A bird is flying."

Information carriers are ordered. For example, we can say, "*An information carrier i contains at least the same information that another carrier j does.*" In the literature, several authors have proposed that information can be ordered with respect to containment [Barwise and Etchemendy 1990; Landman 1986]. Information containment, denoted by $i \to j$, is a relation over the information carriers formalizing the intuition that information is fundamentally "nested" (see also Van Rijsbergen [1989]). This nesting may simply be a product of the syntax of the information carriers; for example, in a Boolean language, $i \wedge j \to i$. Information containment also embodies how information is sometimes implicitly nested. For example, the information conveyed by "salmon" also carries the information "fish." The former we refer to as *surface containment*, and the latter *deep containment*. In general, information containment (either *surface* or *deep*) is denoted by the symbol $\to$, whereby $\to$ is the union of the relation's surface containment ($\xrightarrow{s}$) and deep containment ($\xrightarrow{d}$). It is important to make this distinction as some IR models only support surface containment, whereas others support a notion approximating deep containment. Moreover, related to the information composition, there are $i \oplus j \to i$ and $i \oplus j \to j$.

Information carriers $i$ and $j$ are said to preclude each other, denoted $i \perp j$, if the information carried by $i$ clashes with, or contradicts, the information carried by $j$. It is acceptable to assume that an information carrier precludes its own negation. However, information preclusion is a more subtle notion than contradiction in logic. Information carriers may clash due to underlying natural language semantics, or convention. For example, *swimming* $\oplus$ *crocodile* is acceptable, but *flying* $\oplus$ *crocodile* is meaningless in most contexts. It has also been suggested that information preclusion arises in IR as a consequence of information needs [Bruza and Van Linder 1998]. For example, when searching for documents about *wind surfing*, terms such as *Internet*, *Web*, *net*, and so on may be precluded as the user is not interested in *Web surfing*. In some accounts, (e.g., Landman [1986] and Bruza and Huibers [1994]), the composition of clashing information is formalized as the "meaningless" information carrier, denoted by 0. It is attributed with properties similar to *falsum* in propositional logic; for example, $i \perp j \Leftrightarrow i \oplus j = 0$. The *meaningless information carrier* contains all the information carriers used in an application.

Furthermore, the concept of an information field is defined. It provides the necessary building blocks to express the properties of aboutness. An information field is a structure $(\mathfrak{I}, \to, \oplus, \perp, 0)$ where

—$\mathfrak{I}$ is a nonempty set of information carriers;

—$(\mathfrak{I}, \to)$ is a poset (partially ordered set);

—$0 \in \mathfrak{I}$ and for all $i \in \mathfrak{I}$, $0 \rightarrow i$;

—if $i, j \in \mathfrak{I}$ then $i \oplus j \in \mathfrak{I}$, where $i \oplus j$ is the largest information carrier such that $i \oplus j \rightarrow i$ and $i \oplus j \rightarrow j$; and

—$\perp \subseteq \mathfrak{I} \times \mathfrak{I}$.

A set of postulates[4] determining the aboutness properties is given in terms of concepts from the information field. IR models can be mapped to the aboutness framework. Based on these postulates, the properties they satisfy can be reflected. Moreover, different IR models can be compared according to the postulates they support.

Postulate 1: Reflexivity (R)

$$i \models i.$$

An information carrier is about itself.

Postulate 2: Containment (C)

$$\frac{i \rightarrow j}{i \models j}.$$

An information carrier is about the information it contains (surface or deep). Deep containment models the transformation of information. For example, assuming that "penguin" has the information "bird" nested within it (i.e., *penguin* $\rightarrow$ *bird*), then the Containment postulate permits the conclusion that "penguin" is about "bird(s)." As a consequence, a document about "penguin" is also about "bird." This postulate is recall-oriented.

On the other hand, exact match IR models, which attempt to promote precision, can be defined in terms of surface Containment: $D \models Q$ only if $D \overset{s}{\longrightarrow} Q$. In other words, document $D$ is not about query $Q$ if $D$ does not include $Q$ (completely). This can be modeled by the following postulate.

Postulate 3: Closed World Aboutness Assumption (CWAA)

$$\frac{i \not\rightarrow^{s} j}{i \not\models j}.$$

If an information carrier $i$ is present in another carrier $j$, we sometimes infer that $i$ is not about $j$. Exact match IR models, such as Boolean retrieval, are based on the CWAA. For example, if query $Q$ is not contained in a document $D$, it is assumed that $D$ is not about $Q$. CWAA helps improve precision but degrades the recall, because it ignores the partial matching and the possible information transformation, which could establish the aboutness relationship between $D$ and $Q$. The negative impact of the closed world assumption has been known for some time [Van Rijsbergen 1986b].

Postulate 4: Right Containment Monotonicity (RCM)

$$\frac{k \models i, i \rightarrow j}{k \models j}.$$

[4] The notion "postulate" is intended to characterize the assumptions inherent within a given retrieval mechanism with regard to aboutness.

This postulate allows transitivity of the aboutness relation with respect to information containment. More implicit aboutness relationships can be derived via this postulate. Thus it is recall-oriented. For example, given a document $d$ is about "penguin" and "penguin" contains the information "bird," we can conclude that $d$ is also about "bird(s)." From an IR perspective, RCM models term-based query expansion whereby the term $i$ is replaced by the broader term $j$.

Postulate 5: Left Compositional Monotonicity (LM)

$$\frac{i \models k}{i \oplus j \models k}.$$

Postulate 6: Right Compositional Monotonicity (RM)

$$\frac{i \models k}{i \models k \oplus j}.$$

LM and RM are used as an underlying assumption of some overlap-based IR models: aboutness is preserved under composition. Therefore, they are recall-oriented postulates and they could negatively affect the precision (see Bruza et al. [2000a] for an extended discussion on this topic). By way of illustration, consider a document $d$ about "emperor penguins" ($d \models \mathit{emperor} \oplus \mathit{penguin}$), so $d$ is also about "penguins" (via RCM: $d \models \mathit{penguin}$). Right Compositional Monotonicity allows us to compose arbitrary information on the right-hand side. Thus, $d \models \mathit{publisher} \oplus \mathit{penguin}$ would be permitted, which is an example of an unsound aboutness inference that would lead to a loss of precision in the retrieval mechanism. Query expansion is an example of an IR process that is not monotonic with respect to information composition. The terms selected to expand a query must be carefully chosen. This suggests that a conservatively monotonic process is involved.

The postulates LM and RM can be more clearly related to IR in the following way. LM models the case whereby aboutness is preserved when information $j$ is added to a document:

$$\frac{d \models q}{d \oplus j \models q}.$$

A retrieval function satisfying this property makes aboutness judgment insensitive to a document's length. In this way, the issue of document length normalization[5] can be characterized at the symbolic level.

RM, on the other hand, can be envisaged as query expansion, or any process that attempts to improve a query by composing information in it (e.g., pseudorelevance feedback [Xu and Croft 1996]). We have just shown that this is unsound:

$$\frac{d \models q}{d \models q \oplus j}.$$

Next, we give some conservative forms of mononicity to constrain how information is composed in various ways in order to promote precision.

[5] Document length normalization improves the effectiveness of retrieval; more sophisticated matching functions normalize according to document length.

Postulate 7: Mix (M)

$$\frac{i \models k, j \models k}{i \oplus j \models k}.$$

For example, from "penguin $\models$ bird" and "tweety $\models$ bird," we can derive "tweety $\oplus$ penguin $\models$ bird," meaning "penguin" is about "bird(s)," "tweety" is about a "bird," so "Tweety, the penguin" is about a "bird."

Postulate 8: Context-Free And (C-FA)

$$\frac{k \models i, k \models j}{k \models i \oplus j}.$$

Boolean retrieval is founded on this postulate. For example, if a document is about "computer software" and the same document is about "computer hardware," it is also about both "computer software and hardware."

Postulate 9: Guarded Left Compositional Monotonicity (GLM)

$$\frac{i \models k, i \not\perp j}{j \oplus j \models k}.$$

Postulate 10: Guarded Right Compositional Monotonicity (GRM)

$$\frac{i \models k, k \not\perp j}{i \models k \oplus j}.$$

GLM and GRM are conservative forms of LM and RM. An information carrier can only be composed in another one when no preclusion relationships are violated. For example, suppose "penguin" precludes "flying" and "penguin" is about "bird." According to GLM, "flying" cannot be composed in "penguin" so that "flying $\oplus$ penguin $\models$ bird" (flying penguin is about a bird) cannot be derived.

Postulate 11: Qualified Left Monotonicity (QLM)

$$\frac{i \models k, k \not\perp j}{i \oplus j \models k}.$$

Postulate 12: Qualified Right Monotonicity (QRM)

$$\frac{i \models k, i \not\perp j}{i \models k \oplus j}.$$

QLM and QRM are other conservative forms of LM and RM. LM allows " bird $\oplus$ tweety $\models$ flying" (Tweety, which is a bird, is about flying) to be inferred from "bird $\models$ flying" (A bird is about flying). QLM prevents this via the qualifying preclusion "tweety$\perp$ flying." QRM works in a similar way.

The next postulate expresses a principle based on the preservation of "non-aboutness."

Postulate 13: Negation Rational (NR)

$$\frac{i \not\models k}{i \not\models k \oplus j}.$$

If a document is not about *bird*, it is impossible to be about *flying bird*. This is the intuition behind the postulate NR. Thus it is precision-oriented.

The above postulates could be classified into *recall-oriented* and *precision-oriented* according on their effects on IR. Postulate R can be considered a starting point of aboutness inference. Postulates C (deep), RCM, LM, RM, and CWA are mainly recall-oriented because they tend to produce more aboutness relations than exact match. Postulates C-FA, M, GLM, GRM, QLM, QRM, and NR, on the other hand, intend to prevent undesirable aboutness inferences by employing some kinds of guarded conditions. This is closely related to the conservative monotonicity of IR, which is discussed later in Section 5. The Containment (surface) postulate characterizes exact match IR models, meaning the query must be fully contained in the document.

## 2.2 Formal Evaluation Methodology

The functional benchmark for IR is based on a formal methodology for inductive evaluation. It is conducted in the following steps.

*Step A.* For each IR model, perform the following.

(A.1) Define the background of the IR model to be evaluated.

(A.2) Map the IR model to the aboutness framework. This includes the representations of document, query, aboutness decision, containment, composition, and preclusion.

(A.3) Inductive evaluation. Determine which aboutness postulates the IR model supports. With respect to an aboutness postulate, the IR model could fall into one of the following four categories.

—It *fully supports* the postulate.

—It does *not support* the postulate.

—It *conditionally supports* the postulate: the model does not support the postulate in every situation. Under certain conditions, which are determined extraneously, however, it would be supported. In this article, “conditionally supporting” is applicable to models which involve settings or estimations outside the models themselves. For example, whether the threshold vector space model, threshold possible world-based model, and the classical probabilistic model support certain postulates depends on the threshold settings or the estimations. Note that the notion of “conditionally support” is inapplicable to IR models not involving extraneous factors.

—The postulate is *inapplicable* to the model: some operators involved in the postulate may be foreign (i.e., inapplicable) to the model. Thus we are unable to evaluate the model using that postulate. For example, the preclusion operator is foreign to the vector space model. This in turn implies that postulates involving the preclusion operator are inapplicable to the vector space model. Practically, this is the same as “not supported.” This category is separate in order to provide additional information on why a model fails to support the postulate.

*Step B.* Collect the evaluation results of the different IR models and compare their functionality.

In the following sections, we use the above-defined functional benchmarking suite to evaluate and analyze various classical and logical IR models. We only show the formal proofs of postulates Left Monotonicity (LM) and Right Monotonicity (RM) for illustration. The other postulates can be proven similarly (refer to Song [2000] for details).

## 3. INDUCTIVE EVALUATION OF CLASSICAL IR MODELS

The common classical IR models are the Boolean, vector space, and probabilistic models. In particular, the vector space model is divided into two types, zero-threshold (binary) and threshold (multivalued) vector space models. The former is referred to as the naïve vector space model.

### 3.1 Boolean Model

3.1.1 *Background.* The Boolean model is based on set theory and Boolean algebra. This model has been adopted by many early retrieval systems due to its simplicity. In Boolean retrieval, a document $D$ is represented by a set of characterization terms $X(D) = \{t_1, t_2, \dots, t_n\}$; a query $Q$ is expressed in terms of index terms combined by Boolean logical connectives AND, OR, and NOT. A document is retrieved if and only if the query $Q$ can be deduced from $X(D)$ according to the following set of inference rules.

Rule 1. If $t_i \in X(D)$ then $X(D) \vdash t_i$, where $\vdash$ denotes the logical consequence.
Rule 2. If $X(D) \vdash t_i$ and $X(D) \vdash t_j$, then $X(D) \vdash t_i \wedge t_j$.
Rule 3. If $X(D) \vdash t_i$ or $X(D) \vdash t_j$, then $X(D) \vdash t_i \vee t_j$.
Rule 4. If $X(D) \nvdash t_i$ then $X(D) \vdash \neg t_i$.

To generalize, Boolean expressions are assumed to be in CNF (conjunctive normal form) of DNFs (disjunctive normal form); for example, $(t_1 \vee t_2) \wedge (t_3 \vee t_4) \wedge (t_5 \vee t_6)$.

3.1.2 *Boolean Aboutness* ($\models_{BL}$). Let $U$ be the set of all documents, and $T$ be the set of index terms. Let $D$ be a document (i.e., $D \in U$) and $Q$ a query. Suppose $t_i \in T$, $X(D) = \{t_1, t_2, \dots, t_n\}$ denotes the set of characterization terms of $D$. Let $BL_{OR}$ be the Boolean language defined on $T$ in DNF of $t_i$ (or $\neg t_i$). Furthermore, let $Q = q_1 \wedge q_2 \wedge \cdots \wedge q_m$ be a formula in CNF, where $q_i \in BL_{OR}$; that is, $q_i = t_{i1} \vee t_{i2} \vee \cdots \vee t_{ik}$. Thus, aboutness in the Boolean model is characterized by the following definition.

—$D \models_{BL} Q$ if and only if $X(D) \vdash Q$ (Aboutness)
$X(D) \vdash Q$ if and only if $(\forall q_i)\,(X(D) \vdash q_i)$
$X(D) \vdash q_i$ if and only if $(\exists t_{ij})\,(X(D) \vdash t_{ij})$
—If $D \not\models_{BL} Q$ then $D \models_{BL} \neg Q$ (Close World Assumption)
—$D \xrightarrow{s} Q$ if and only if $X(D) \vdash Q$ (Surface Containment)
—Deep Containment is inapplicable.

- Let $Q1 = q_{11} \wedge q_{12} \wedge \cdots \wedge q_{1m}$ and $Q2 = q_{21} \wedge q_{22} \wedge \cdots \wedge q_{2l}$;
  $Q1 \rightarrow Q2$ if and only if $Cl(\{q_{11}, q_{12}, \ldots, q_{1m}\}) \supseteq \{q_{21}, q_{22}, \ldots, q_{2l}\}$ where $Cl(Q1)$ is defined as the set of DNF formulas that are logical consequences of $q_{11}, q_{12}, \ldots$, and $q_{1m}$.
- $Q1 \oplus Q2 \Leftrightarrow Q1 \wedge Q2$ (Query Composition)
- $D1 \oplus D2 \Leftrightarrow D1 \cup D2$ (Document Composition)
- Suppose $D$ is considered as formula $t_1 \wedge t_2 \wedge \cdots \wedge t_n$, then $D \perp Q \Leftrightarrow D = \neg Q$ (Preclusion)
- $Q \perp \neg Q$.

### 3.1.3 *Inductive Evaluation*

THEOREM 1. *The Boolean model supports the Postulates R, C (Surface), C-FA, RCM (Surface), LM, M, GLM, QLM, NR, and CWAA.*[6] *Deep Containment is inapplicable to this model.*

Proofs of LM and RM are shown as below.

- LM: Left Compositional Monotonicity is supported.
  Given $D1 \models_{BL} Q$
  $\Rightarrow X(D1) \vdash Q$
  $\Rightarrow X(D1 \oplus D2) = X(D1 \cup D2) \vdash Q$
  $\therefore D1 \oplus D2 \models_{BL} Q$.
- RM: Right Compositional Monotonicity is not supported.
  Given $D \models_{BL} Q1$ and $Q = Q1 \oplus Q2$
  $\Rightarrow X(D) \vdash Q1$ and $Q1 \oplus Q2 \Leftrightarrow Q1 \wedge Q2$
  But $X(D) \vdash Q1 \wedge Q2$ cannot be concluded
  $\therefore D \models_{BL} Q1 \oplus Q2$ cannot be concluded.

### 3.1.4 *Remarks*

- The Boolean model is an exact match IR model, thereby promoting precision.
- The Boolean model is left monotonic, rendering it insensitive to document length.
- The Boolean model supports the closed world assumption, which would negatively affect recall.
- RM is not supported by the Boolean model. Instead, a conservative form, C-FA, is supported. This would promote precision.

In general, the Boolean model supports a fair degree of precision and is weak in recall. Its insensitivity to document length makes it less effective than models whose matching functions support document length normalization.

## 3.2 Vector Space Model

3.2.1 *Background.* In the vector space model, both queries and documents are represented as a vector of weighted or binary index terms. Practically, each

[6] Note that postulates MIX, GLM, and QLM are trivially supported, as LM is supported.

index term is treated as an axis in an $n$-dimensional space. The documents are ranked by the similarity between the document $D$ and the query $Q$. There are a number of measures of vector similarity, such as inner product, dice coefficient, cosine coefficient, and so on. The commonly used form is the cosine function:

$$\mathrm{Cos}(D, Q) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \times \sum_i y_i^2}}$$

where

$$D = \{x_1, x_2, \ldots, x_n\}, \qquad Q = \{y_1, y_2, \ldots, y_n\}.$$

A threshold value is always employed to determine relevance. In the following discussions, we first consider the naïve and simplest case of the model. For this case, the aboutness between $D$ and $Q$ is equivalent to simple overlapping; that is, if $D$ and $Q$ share at least one index term, they are about each other. We then investigate the more general case of nonzero multivalued threshold. Note that the threshold value is extraneously controlled. To simplify, we just consider the case where index terms are unweighted. The case of weighted terms could be investigated similarly.

3.2.2 *Naïve Vector Space Aboutness* ($\models_{VS-NAIVE}$). Let $U$ be the set of all documents, and T be the set of index terms. Let $D \in U$ be a document, and $Q$ a query. Both $D$ and $Q$ are represented as vectors.

—$D = D^+ \cup D^-$
$D^+ = \{t_1^+, t_2^+, \ldots, t_f^+\}$
$D^- = \{t_1^-, t_2^-, \ldots, t_g^-\}$
$Q = Q^+ \cup Q^-$
$Q^+ = \{t_1^+, t_2^+, \ldots, t_k^+\}$
$Q^- = \{t_1^-, t_2^-, \ldots, t_h^-\}$
$f + g = k + h = n$ (dimension of the vector),
where $t_i \in T, t_i^+$ is the $i$th nonzero term in the vector, and $t_j^-$ is the $j$th zero term in the vector.

Based on the above $D$ and $Q$ vectors, the following definitions of naive vector space aboutness are defined.

—$D \models_{VS-NAIVE} Q$ if and only if $D^+ \cap Q^+ \neq \emptyset$ (Aboutness)
$D \not\models_{VS-NAIVE} Q$ if and only if $D^+ \cap Q^+ = \emptyset$

—$D \xrightarrow{s} Q$ if and only if $D^+ \supseteq Q^+$ (Surface Containment)
$Q1 \xrightarrow{s} Q2$ if and only if $Q1^+ \supseteq Q2^+$

—Deep Containment is inapplicable.

—$Q = Q1 \oplus Q2 \Leftrightarrow Q^+ = Q1^+ \cup Q2^+$ and $Q^- = (Q1^- - Q2^+) \cup (Q2^- - Q1^+)$ (Query Composition)

—$D = D1 \oplus D2 \Leftrightarrow D^+ = D1^+ \cup D2^+$ and $D^- = (D1^- - D2^+) \cup (D2^- - D1^+)$ (Document Composition)

—$\perp$ is inapplicable, as it is not supported in the naive vector space model.

### 3.2.3 *Inductive Evaluation*

THEOREM 2. *The naive vector space model supports R, C (surface), C-FA, LM, and RM.*[7] *Deep containment is inapplicable to this model. The postulates GLM, GRM, QLM, and QRM are inapplicable, as preclusion is inapplicable.*

Proofs of LM and RM are shown as below.

—LM: Left Compositional Monotonicity is supported.
Given $D1 \models_{VS-NAIVE} Q, D = D1 \oplus D2$
$\Rightarrow D1^+ \cap Q^+ \neq \emptyset, D = D1 \oplus D2$
$\Rightarrow (\exists t_i)(t_i \in D1^+ \wedge t_i \in Q^+)$, and
by the definition of composition, $D = D1 \oplus D2 \Leftrightarrow D^+ = D1^+ \cup D2^+$
$\Rightarrow t_i \in D^+$ and $t_i \in Q^+$
$\Rightarrow D^+ \cap Q^+ \neq \emptyset$
$\therefore D1 \oplus D2 \models_{VS-NAIVE} Q$.

—RM: Right Compositional Monotonicity is supported.
Given $D \models_{VS-NAIVE} Q1, Q = Q1 \oplus Q2$
$\Rightarrow D^+ \cap Q1 \neq \emptyset$ and $Q = Q1 \oplus Q2$
$\Rightarrow$ $(\exists t_i)(t_i \in D^+ \wedge t_i \in Q1^+)$, and by the definition of composition, that is, $Q = Q1 \oplus Q2 \Leftrightarrow Q^+ = Q1^+ \cup Q2^+$
$\Rightarrow t_i \in D^+$ and $t_i \in Q^+$
$\Rightarrow D^+ \cap Q^+ \neq \emptyset$
$\Rightarrow D \models_{VS-NAIVE} Q1 \oplus Q2$.

3.2.4 *Threshold Vector Space Aboutness* ($\models_{VS-T}$). Let $U$ be the set of all documents, and $T$ be the set of index terms. Let $D \in U$ be a document, and $Q$ a query. Both $D$ and $Q$ are represented as vectors. Based on these, the following definitions of threshold vector space aboutness are given.

—$D \models_{VS-T} Q$ if and only if $\cos(D, Q) \geq \partial$, where $\partial \in (0, 1]$. (Aboutness)
$D \not\models_{VS-T} Q$ if and only if $\cos(D, Q) < \partial$

—The mappings of containment, composition, and preclusion are the same as those in Section 3.2.2.

### 3.2.5 *Inductive Evalution*

THEOREM 3. *The threshold vector space model supports R, and conditionally supports C (surface), CWAA, RCM (surface), LM, RM, M, C-FA, and NR. Deep containment is inapplicable to this model. The postulates GLM, GRM, QLM, and QRM are inapplicable, as preclusion is inapplicable.*

The proofs of LM and RM are as follows.

—LM: Left Compositional Monotonicity is conditionally supported.

[7] Note that postulate MIX is trivially supported, as LM is supported. The postulate C-FA is trivially supported, as RM is supported.

Let $|D1^+| = f1, |D2^+| = f2, |Q^+| = k, |D1^+ \cap Q^+| = c1, |D2^+ \cap Q^+| = c2$ and $|D1^+ \cap D2^+| = l$.
Then there are $|Q^+ \cap (D1 \oplus D2)^+| = c1 + c2 - l$ and $|(D1 \oplus D2)^+| = f1 + f2 - l$. Given $D1 \models_{VS-T} Q, D = D1 \oplus D2$
$\Rightarrow \cos(D1, Q) = \frac{c1}{\sqrt{f1+k}} \geq \partial, D = D1 \oplus D2$.
This cannot imply $\cos(D1 \oplus D2, Q) \geq \partial$. Consider the case where $D2^+$ is much larger than $D1^+$. $\cos(D1 \oplus D2, Q)$ may be reduced to a very small value, even less than $\partial$.
$\therefore D1 \oplus D2 \models_{VS-T} Q$ cannot be guaranteed.
To ensure $D1 \oplus D2 \models_{VS-T} Q$, $\cos(D1 \oplus D2, Q) = \frac{c1+c2-l}{\sqrt{(f1+f2-l)+k}}$ must not be less than $\partial$.
Thus, given $D1 \models_{VS-T} Q$, that is, $\cos(D1, Q) = \frac{c1}{\sqrt{f1+k}} \geq \partial$, the LM postulate is supported only under the condition of $\partial \leq \frac{c1+c2-l}{\sqrt{(f1+f2-l)+k}}$.

—RM: Right Compositional Monotonicity is conditionally supported.
Let $|D^+ \models f, |Q1^+| = k1, |Q2^+| = k2, |D^+ \cap Q1^+| = c1, |D^+ \cap Q2^+| = c2$ and $|Q1^+ \cap Q2^+| = l$.
Then there are $|D^+ \cap (Q1^+ \oplus Q2^+)| = c1 + c2 - l$ and $|(Q1 \oplus Q2^+)| = k1 + k2 - l$. Following the similar proof for LM, we get the conclusion that, given $D \models_{VS-T} Q1$, that is, $\cos(D, Q1) = c1/(\sqrt{f+k1}) \geq \partial$, the RM postulate is supported only under the condition of $\partial \leq ((c1+c2-l)/(\sqrt{f+(k1+k2-l)}))$.

3.2.6 *Remarks.* The naïve vector space model is both left and right monotonic. As these properties degrade precision, this model would be imprecise in practice.

We argue that IR is *conservatively monotonic* in nature, rather than fully monotonic or nonmonotonic. Conservative monotonicity means that when new information is composed on either the left- or right-hand side, the aboutness relationship should be preserved only under certain guarding conditions. For example, consider the query expansion process. When a query is expanded using additional terms, the terms added are not arbitrary. They must be chosen carefully; that is, conservative monotonicity is at work here. In terms of aboutness, such models embody properties such as QLM, QRM, and the like without also supporting LM and RM.

The threshold vector space model only supports R. The monotonic properties such as LM and RM are conditionally supported depending on the threshold. This means that by adjusting the threshold value, users could adjust the degree of nonmonotonicity. In this way, the threshold vector space model mimics conservative monotonicity by conditionally supporting LM and RM. For example, the condition of the threshold vector space model supporting LM can be conceived in the following terms. Consider a set of terms $Q$ (the query) and the set of terms $D$ (the document). For reasons of clarity, assume that $Q \subset D$. The decision whether $D \models_{VS-T} Q$ holds can be analyzed in terms of LM: starting with $Q$, terms are composed in $Q$ until the set $D$ has been constructed. Observe that as the number of terms in $D$ increases, the cosine normalization will increase.

There will be a point where the cosine between $D$ and $Q$ will become less than the threshold value $\delta$. In other words, LM is more likely to be preserved for short documents, which in a practical sense means that the threshold vector space model will favor the retrieval of short documents. Observe that the nonmonotonicity of the threshold vector space model is not determined by the model itself, but by external settings. This is undesirable from a theoretical point of view.

## 3.3 Probabilistic Model

3.3.1 *Background.* In the probabilistic model, the probability of relevance of a document $D$ subjected to a query $Q$ is given by P(rel | $D$). To simplify, $D$ is assumed to be a vector-valued random variable $(t_1, t_2, \ldots, t_n)$, and $t_1, t_2, \ldots, t_n$ are assumed to be stochastically independent of each other. P(D) is then given by:

$P(D) = P(D \mid \mathrm{rel})P(\mathrm{rel}) + P(D \mid n\mathrm{rel})P(n\mathrm{rel})$.
$P(\mathrm{rel} \mid D)$ is computed as follows.
$P(\mathrm{rel} \mid D) = \dfrac{P(D \mid rel)P(\mathrm{rel})}{P(D)}$.
$P(n\mathrm{rel} \mid D) = \dfrac{P(D \mid \mathrm{rel})P(\mathrm{rel})}{P(D)}$.
$P(D \mid \mathrm{rel}) = \prod_{i=i}^{n} P(t_i \mid \mathrm{rel})^{t_i}$.
$P(D \mid n\mathrm{rel}) = \prod_{i=i}^{n} P(t_i \mid n\mathrm{rel})^{t_i}$.
$t_i = 0$ if and only if term $i$ is absent in $D$.
$t_i = 1$ if and only if term $i$ is present in $D$.
$P(\mathrm{rel})$ and $P(n\mathrm{rel})$ are the priori probabilities of relevance and irrelevance, respectively.

$P(t_i \mid \mathrm{rel})$ and $P(t_i \mid n\mathrm{rel})$ could be estimated if we have complete information about the relevant and irrelevant documents in the collection.

The Bayes' Decision Rule is used to make the decision for or against relevance: $D$ is relevant if and only if $P(\mathrm{rel} \mid D) > P(n\mathrm{rel} \mid D)$; that is, $P(D \mid \mathrm{rel})P(\mathrm{rel}) > P(D \mid n\mathrm{rel})P(n\mathrm{rel})$. This leads to a discriminant function:

$$g(D) = \frac{P(D \mid \mathrm{rel})P(\mathrm{rel})}{P(D \mid n\mathrm{rel})P(n\mathrm{rel})} = \frac{P(\mathrm{rel}) \prod_{i=i}^{n} P(t_i \mid \mathrm{rel})^{t_i}}{P(n\mathrm{rel}) \prod_{i=i}^{n} P(t_i \mid n\mathrm{rel})^{t_1}}.$$

The document D is retrieved if and only if $g(D) > 1$.

Note that P(rel)/P($n$rel) is constant for a given query and document base, and is independent of any particular document.

3.3.2 *Probabilistic Aboutness* ($\models_{PB}$). Let $U$ be the set of all documents, and $T$ be the set of index terms. Let $D \in U$ be a document, and $Q$ a query. $D$ is represented as a vector of index terms, as described in the last section. The representation of a query is not specified in the model. In this article we just assume the representation of $Q$ is the same as that of $D$. Based on these, the following definitions of probabilistic aboutness are given.

- The representations of $D$ and $Q$ are the same as those of the vector space model.

—$D \models_{PB} Q$ if and only if $g(D) > 1$. (Aboutness)
$D \not\models_{PB} Q$ if and only if $g(D) \leq 1$.

—The mappings of containment, composition, and preclusion are the same as those in Section 3.2.2.

### 3.3.3 *Inductive Evaluation*

THEOREM 4. *The probabilistic model conditionally supports R, C (surface), CWAA, RCM (surface), LM, RM, C-FA, M, and NR. Deep containment is inapplicable to this model. The postulates GLM, GRM, QLM, and QRM are inapplicable, as preclusion is inapplicable.*

The proofs of LM and RM are shown as follows.

—LM is conditionally supported.
Given $D1 \models_{PB} Q, D = D1 \oplus D2$
$\Rightarrow g(D1) = \frac{P(\text{rel})\prod_{i=1}^{n} P(t_i \,|\, \text{rel})^{t_i}}{P(n\text{rel})\prod_{t=1}^{n} P(t_i \,|\, n\text{rel})^{t_i}} > 1$ with respect to $Q1, D^+ = D1^+ \cup D2^+$.

Suppose the terms $\{t_j, \ldots, t_k\}$ in $D^+$ but not in $D^+$
$\Rightarrow g(D) = g(D1) \times \frac{\prod_{i=1}^{k} P(t_i \,|\, \text{rel})}{\prod_{i=1}^{k} P(t_i \,|\, n\text{rel})}$.

Whether $g(D) > 1$ depends on $((\prod_{i=j}^{k} P(t_i \,|\, \text{rel}))/(\prod_{i=j}^{k} P(t_i \,|\, n\text{rel}))$ . Only if the new composed terms from $D2$ have higher probability of occurring in the relevant set than the irrelevant set, is LM supported (i.e., $g(D) > 1$).

—RM is conditionally supported.
Given $D \models_{PB} Q1$
$\Rightarrow g(D) = \frac{P(\text{rel})\prod_{i=1}^{n} P(t_i \,|\, \text{rel})^{t_i}}{P(n\text{rel})\prod_{i=1}^{n} P(t_i \,|\, n\text{rel})^{t_i}} > 1$ with respect to $Q1$.
With respect to $Q1 \oplus Q2$, however, the above estimations may change. Thus $g(D) > 1$ could not be guaranteed any more.
Therefore, with respect to $Q1 \oplus Q2$, only when the estimations of the a priori probability of relevance and the probability of index terms in $D$ occurring in the relevant set are stronger than those of irrelevance, could $g(D) > 1$ obtained.

3.3.4 *Remarks.* The classical probabilistic model conditionally supports R, LM, and RM. This shows that it is fully nonmonotonic. The nonmonotonicity is achieved by the estimation of relevance and irrelevance and the probability of occurrence of index terms in the relevant and irrelevant sets via a training process. This leads to good performance for the probabilistic model in practice.

The properties supported by the threshold vector space and probabilistic models are almost the same. These models are generally most effective in practice. The key here is that LM and RM are conditionally supported (i.e., they mimic conservative monotonicity). For example, the condition of the probabilistic model supporting RM is that new terms composed in a document must have higher probability of occurrence in the relevant set than the irrelevant set. This is consistent with the nature of conservative monotonicity.

The advantage of the probabilistic model over the threshold vector space model is that its decision rule is included within the model, whereas the threshold value in the threshold vector space model is not determined by itself. However, the probabilistic model does not directly deal with the matching between documents and queries. Instead, as we have shown in the proofs of its properties, the estimations are conducted on the whole document set with respect to a query. Moreover, the model itself does not specify the criteria of the estimation. This means it may vary from one query to another. This explains why the probabilistic model does not fully support R (i.e., even if a document is identical to a query, the probabilistic model could not determine that they are relevant).

### 3.4 Discussion of Extended Boolean and Inference Network Models

A well-known alternative Boolean model is the extended Boolean model [Salton 1988], also called the *p-norm* model. On the other hand, the inference network model [Turtle and Croft 1992] is an alternative probabilistic model. Both of them can simulate from the conventional Boolean model to the inner-product vector space model by tuning certain parameters between their top and bottom margins (e.g., $1 \leq p \leq \infty$ for the extended Boolean model; $n \leq c \leq \infty$ for the inference network model, where $n$ is the number of parents at a given node in the inference network). It has been proven by Turtle and Croft [1992] that when the extended Boolean and inference network models are adjusted to simulate Boolean and inner-product vector space models, respectively, they produce the same results. They are similar to each other when they produce the intermediate systems between Boolean and inner-product vector space models for $1 < p < \infty$ and $n < c < \infty$, respectively. For this reason, we only give the detailed discussion on the extended Boolean model in this article. The inference network model can be analyzed similarly. Moreover, the treatment of this model is a bit different from the others. We focus on showing how the most important property, left and right monotonicity, of the extended Boolean model changes from Boolean to vector space models with the change of $p$-value.

The extended Boolean model [Salton 1988] provides term weighting and ranking of the answer set. The similarity between a document and a query is adjusted by a special parameter, namely, the $p$-value. Different $p$-values lead to different document output values. In this model, a query is the conjunction or disjunction of $n$ terms, and a document is represented as a vector $D = (t_1, t_2, \ldots, t_n)$. For the purpose of this article, we assume terms in the query are binary. The similarity between a document and a query is given by

$$\begin{aligned}\mathrm{Sim}(D, Q_{\mathrm{and}}) &= 1 - \left[\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}\right]^{1/p} \\ \mathrm{Sim}(D, Q_{\mathrm{or}}) &= \left[\frac{t_1^p + t_2^p + \cdots + t_n^p}{n}\right]^{1/p}, \quad \text{where } 1 \leq p \leq \infty.\end{aligned}$$

When $p = \infty$, the extended Boolean model simulates normal Boolean logic; that is, $\mathrm{sim}(D, Q_{\mathrm{and}}) = \min(t_i)$ and $\mathrm{sim}(D, Q_{\mathrm{or}}) = \max(t_i)$. For $p = 1$, it behaves like a simple normalized inner-product vector space model; that is, $\mathrm{sim}(D, Q_{\mathrm{and}}) = \mathrm{sim}(D, Q_{\mathrm{or}}) = \sum t_i / n$.

For intermediate $p$-values, this model generates “soft” Boolean systems whose properties are between the Boolean and vector space models. We then show this by analyzing how the monotonicity of the extended Boolean model changes from a Boolean to an inner-product vector model with respect to the $p$-value. We first define the extended Boolean aboutness ($\models_{EB}$) as below.

$$D \models_{EB} Q \text{ if and only if } \mathrm{sim}(D, Q) \geq \partial, \quad \text{where } \partial \in (0, 1].$$

We suppose the query is represented in conjunction normalized form (CNF). To simplify the analysis, we use the representation of sim ($D$, $Q_{\text{and}}$) for the computing of complex queries in CNF, since both sim($D$, $Q_{\text{or}}$) and $d$, are in the interval [0, 1]. Information composition ($\otimes$) between two queries is modeled as logical AND, whereas composition between two documents is modeled as $D = D1 \oplus D2 \Leftrightarrow D^+ = D1^+ \cup D2^+$. The left and right monotonicity of extended Boolean aboutness can then be analyzed.

—Left Monotonicity is supported.
Given $D1 \models_{EB} Q$
$\Rightarrow \mathrm{Sim}(D1, Q) = 1 - [\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}]^{1/p} \geq \partial$
$D = D1 \oplus D2 \Leftrightarrow D^+ = D_1^+ \cup D_2^+$; Suppose $D = (t'_1, t'_2, \ldots, t'_n)$
$\Rightarrow \mathrm{Sim}(D_1, Q) = 1 - [\frac{(1-t'_1)^p + (1-t'_2)^p + \cdots + (1-t'_n)^p}{n}]^{1/p} \geq \mathrm{Sim}\ (D1, Q) \geq \partial$
$\Rightarrow D \models_{EB} Q$.

The above proof shows that the extended Boolean model is left monotonic no matter what the $p$-value is. This is consistent with the conventional Boolean model (see Section 3.1). Compared with the threshold vector space model using the cosine function (see Section 3.2.5), which conditionally supports left monotonicity, the similarity function of the extended Boolean model is normalized using only the query terms, without considering the expansion of document space. Thus it is not as effective as the cosine vector space system with respect to left monotonicity. That is, it remains insensitive to document length.

—Right Monotonicity.
Given $D \models_{EB} Q1$
$\Rightarrow \mathrm{sim}(D, Q1) = 1 - [\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}]^{1/p} \geq \partial$.
Suppose Q2 is a conjunction of $k$ components.
$\Rightarrow \mathrm{sim}(D, Q1 \oplus Q2) = 1 - [\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p + \cdots + (1-t_{n+k})^p}{n+k}]^{1/p}$.
It is not necessary that sim($D$, $Q1 \oplus Q2$) $\geq \partial$. Thus RM is conditionally supported depending on the values of $p$ and $\partial$.

Now, let’s consider how the change of $p$ leads to the change of the degree of right monotonicity of the model. Suppose sim($D$, $Q1 \oplus Q2$) $< \partial$. $P$ being increased implies 1/p being decreased. Due to

$$\left[\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}\right] \leq 1,$$

$$\left[\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}\right]^{1/p}$$

would be increased and in turn

$$1 - \left[\frac{(1-t_1)^p + (1-t_2)^p + \cdots + (1-t_n)^p}{n}\right]^{1/p}$$

should be decreased. Thus, larger $p$ implies larger distance between $\mathrm{sim}(D, Q1 \oplus Q2)$ and $\partial$, that is, a higher degree of right nonmonotonicity. For $\mathrm{p} = \infty$ and binary document terms, the extended Boolean model reduces to the conventional Boolean model, which has the highest degree of nonmonotonicity (i.e., right monotonicity is not supported; see Section 3.1. Only if all the new terms composed in the query are true in the document, can the original aboutness relation be preserved. This condition is too strict; that is, many documents even with a high possibility of relevance could not be retrieved. For $p = [1, \infty)$, smooth decrease of $p$ means smooth decrease of the degree of nonmonotonicity. When $p$ is reduced to 1, the extended Boolean model becomes the inner-product vector space model, which has the most relaxed condition for conditionally supporting right monotonicity. As a consequence, this model would not be ideal for supporting query expansion or pseudorelevance feedback. Following this procedure, the other aboutness properties can be analyzed similarly.

### 3.5 Summary

In summary, the probabilistic model has potentially the highest degree of precision, followed by the threshold vector space model, then the Boolean model, and the naïve vector space model. This conclusion is consistent with the experimental results. The motivation for this judgment lies in the varying degrees to which they respectively support (or do not support) conservative monotonicity.

## 4. INDUCTIVE EVALUATION OF LOGICAL IR MODELS

In the past decade, a number of logic-based IR models have been proposed (see Bruza and Lalmas [1996], Lalmas [1998], and Lalmas and Bruza [1998] for detailed surveys). These models can be generally classified into three types: situation theory-based, possible world-based, and other types. In what follows, we investigate two well-known logic IR models.

In the following analyses, the fact of a document $D$ consisting of information carrier i is represented by $D \overset{\sim}{\rightarrow} i$. For example, guarded left compositional monotonicity (i.e., Postulate 7) means that if *a document consisting of i is about k* (*i.e.*, $i \models k$), under the guarded condition that *i doesn't preclude j* ($i \pm j$), we can conclude that *a document consisting of* $i \oplus j$ *is about* $k$ ($i \oplus j \models k$). In the following benchmarking exercise, we adopt this interpretation for logical IR models for reasons of simplicity. For the classical models, we treat the document and the query as information carriers directly, for there are no term semantic relationships involved in classical models.

### 4.1 Situation Theory-Based Model

4.1.1 *Background.* Van Rijsbergen and Lalmas developed a situation theory-based model [Lalmas 1996; Van Rijsbergen and Lalmas 1996]. In their model, a document and the information it contains are modeled as a situation and types. A situation $s$ supports the type $\varphi$, denoted by $s \models \varphi$, meaning that $\varphi$ is a part of the information content of the situation. The flow of information is modeled by constraints ($\rightarrow$). Here we assume $\varphi \rightarrow \varphi$. A query is one type (single type query) or a set of types (complex query); for example, a query $\phi = \{\varphi, \psi\}$.

For a situation $s$ and a set of types $\phi$, there are two methods to determine whether $d$ supports $\phi$. The first is that $d$ supports $\phi$ if and only if $s$ supports $\varphi$ for all types $\varphi \in \phi$ [Barwise 1989]. Later Lalmas [1996] relaxed the condition to represent partial relevance: any situation supports $\phi$ if it supports at least one type in $\phi$.

The IR system is to determine to which extent a document d supports the query $\phi$, denoted by $d \models \phi$. If $d \models \phi$ , then the document is relevant to the query with certainty. Otherwise, constraints from the knowledge set will be used to find the flow that leads to the information $\phi$. The uncertainty attached to this flow is used to compute the degree of relevance.

A channel is to link situations. The flow of information circulates in the channel, where the combination of constraints in sequence $(c_1; c_2)$ and in parallel $(c_1 \| c_2)$ can be represented. Given two situations $s1, s2, s1 \mid\xrightarrow{c} s2$ means that $s1$ contains the information about $s2$ due to the existence of the channel $c$. A channel $c$ supports constraint $\varphi \rightarrow \psi$, denoted $c \models \varphi \rightarrow \psi$, if and only if for all situations $s1$ and $s2$, if $s1 \models \varphi, s1 \mapsto s2$, and $\varphi \rightarrow \psi$, then $s2 \models \psi$. The notation $s1 \models \varphi \mid \xrightarrow{c} s2 \models \psi$ stands for $c \mid \varphi \rightarrow \psi$ and $s1 \mid\!\rightarrow s2$, which means that $s1 \models \varphi$ carries the information that $s2 \models \psi$, due to channel $c$. If $s1 \models \varphi \mid\xrightarrow{c} s2 \models \psi$ and $s1 = s2$, then $c$ is replaced by a special channel 1, and $\varphi$ logically entails $\psi$.

4.1.2 *Situation Theory Based Aboutness*($\models_{ST}$). Let $U$ be the set of documents, $S$ be the set of situations, $T$ be the set of types, and $C$ be the set of channels. Furthermore, let $D \in U$ be a document, and $Q$ a query. Then,

—*$D$ is modeled as a situation.*

—*$Q$ is modeled as a set of types*

—*Given two sets of types $\phi 1$ and $\phi 2$:*

—$D' \overset{\sim}{\rightarrow} \phi 1$ if and only if $(\forall \varphi \in \phi 1)(D \models \varphi)$.

—$\phi 1 \models_{ST} \phi 2$ if and only if $(\exists c \in C)\ (\forall D \mid D \overset{\sim}{\rightarrow} \phi 1)(\exists \varphi \in \phi 1)(\exists \psi \in \phi 2)(D \models \varphi \mid\xrightarrow{c} D' \models \psi)$. Note that $D'$ could be $D'$ itself; that is, $c = 1$. A more special case is $D \models \psi \mid\xrightarrow{1} D \models \psi$. (Aboutness)

—$\phi 1 | \not\models_{ST} \phi 2$ if and only if $(\nexists c \in C)(\forall D \mid D \overset{\sim}{\rightarrow} \phi 1)(\exists \psi \in \phi 1)(\exists \psi \in \phi 2)(D \models \varphi \mid\xrightarrow{c} D' \models \psi)$.

—$\phi 1 \xrightarrow{s} \phi 2$ if and only if $\phi 1 \supseteq \phi 2$ (Surface Containment)

—$\phi 1 \xrightarrow{d} \phi 2$ if and only if $(\exists \psi 1 \in \phi 1)(\exists \psi 2 \in \phi 2)(\varphi \rightarrow \psi)$. (Deep Containment)

—$\phi 1 \oplus \phi 2 \Leftrightarrow \phi 1 \cup \phi 2$ (Composition)

—A type precludes its negation; for example, $(s \mid s \models \ll \text{hit}, \text{john}, x; 1 \gg) \perp (s \mid s \models \ll \mathit{hit}, \text{john}, x; 0 \gg)$. (Prelusion)

—Suppose the negation of a set of types $Q$ is the set of the negations of every component type; then $Q \perp \neg Q$.

4.1.3 *Inductive Evaluation*

THEOREM 5. *The situation theory-based IR model supports R, C, LM, RM, M, C-FA, GLM, GRM, QLM, and QRM.*[8]

The proofs of LM and RM are provided as follows.

—LM: Left Compositional Monotonicity is supported.

Given $\phi 1 \models_{ST} \ \phi 2$

$\Rightarrow (\exists\, c1 \in C)(\forall D \overset{\sim}{\rightarrow} D \overset{\sim}{\rightarrow} \phi 1)(\exists \psi 1 \in \phi 1)(\exists \psi 2 \in \phi 2)(D \models \psi' 1 \mid\!\xrightarrow{c1} D' \models \psi 2)$,

$\phi 1 \oplus \phi 3 \Leftrightarrow \phi 1 \cup \phi 3$, and $\{\forall D \mid D \overset{\sim}{\rightarrow} \phi 1 \oplus \phi 3\} \subseteq \{\forall D \mid D \overset{\sim}{\rightarrow} \phi 1\}$

$\Rightarrow (\forall D \mid D \overset{\sim}{\rightarrow} \phi 1 \oplus \phi 3)(\exists \psi 1 \in \phi 1 \oplus \phi 3)(\exists \psi 2 \ \in \phi 2)(D \models \psi 1 \xrightarrow{c1} D' \models \psi 2)$,

$\therefore \phi 1 \oplus \phi 3 \models_{ST} \phi 2.$

—RM: Right Compositional Monotonicity is supported.

Given $\phi 1 \models_{ST} \ \phi 2$

$\Rightarrow (\exists\, c1 \in C)(\forall D \mid D \overset{\sim}{\rightarrow} \phi 1)(\exists \psi 1 \in \phi 1)(\exists \psi 2 \in \phi 2)(D \models \psi 1 \mid\!\xrightarrow{c1} D' \models \psi 2), \phi 1$

$\oplus\, \phi 3 \Leftrightarrow \phi 2 \cup \phi 3$, and $\{\forall D \mid D \overset{\sim}{\rightarrow} \phi 2 \oplus \phi 3\} \subseteq \{\forall D \mid D \overset{\sim}{\rightarrow} \phi 2\}$

$\Rightarrow (\forall D \mid D \overset{\sim}{\rightarrow} \phi 1)(\exists \psi 1 \in \phi 1)(\exists 2 \in \phi 2 \oplus \phi 3)(D \models \psi 1 \xrightarrow{c1} D' \models \psi 2)$,

$\therefore \phi 1 \models \ _{ST} \phi 2 \oplus \phi 3.$

## 4.2 Possible World-Based Model

4.2.1 *Background.* A number of possible world-based logical IR models have been proposed. As stated in Lalmas and Bruza [1998] , these systems are founded on a structure$\langle W, R\rangle$, where $W$ is the set of worlds and $R \subseteq W \times W$ is the accessibility relation. They can be classified according to the choice made for the worlds $w \in W$ and accessibility relation $R$. For example, $w$ can be a document (or its variation) and $R$ is the similarity between two documents $w1$ and $w2$ [Nie 1989, 1992], or $w$ is a term and $R$ is the similarity between two terms $w1$ and $w2$ [Crestani and van Rijsbergen 1995a,b, 1998], or $w$ is the "retrieval situation" and $R$ is the similarity between two situations $w1$ and $w2$ [Nie et al. 1995], and so on.

Most of these systems use a technique called imaging. To obtain $P(D \rightarrow Q)$, where the connective $\rightarrow$ represents conditional, we can move the probability from non-$D$-world to $D$-world by a shift from the original probability distribution $P$ of the world $w$ to a new probability distribution $P_D$ of its closest world $w_D$, where $D$ is true. This process is called deriving $P_D$ from $P$ by imaging on $D$. The truth of $D \rightarrow \ Q$ at $w$ will then be measured by the truth of $Q$ at $w_D$.

[8] Note that postulates MIX, GLM, and QLM are trivially supported, as LM is supported. Postulates C-FA, GRM, and QRM are trivially supported, as RM is supported.

To simplify the analysis, let's suppose that the truth of Q in a world is binary[9] and the closest world of a world $w$ is unique.[10]

$P(d \rightarrow q)$ can be computed as follows.

$$P(D \rightarrow Q) = \sum_{w \in W} P(w) w_D(Q) = \sum_{w \in W} P_D(w) w(Q) \tag{1}$$

$$\sum_{w} P(w) = 1 \tag{2}$$

$$w(Q) = \begin{cases} 1, & \text{if} \quad Q \text{ is true in } w \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$P_D(w) = \sum_{w \in W} P(w') I(w, w') \tag{4}$$

$$I(w, w') = \begin{cases} 1, & \text{if} \quad w = w_D^{'} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$w_D \text{ is the closest world of } w \text{ where } D \text{ is true.} \tag{6}$$

Now we study in detail Crestani and van Rijsbergen's model which models the terms as possible worlds to see some properties of the possible world-based approach. In this model, the term is considered as the vector of documents, while the document and query are vectors of terms. The accessibility relations between terms are estimated by the cooccurrence of terms. $P(D \rightarrow Q)$ can be computed as

$$P(D \rightarrow Q) = \sum_{t \in T} P(t) t_{\mathrm{D}}(Q) = \sum_{t \in T} P_D(t) t(Q) \tag{7}$$

$$\sum_{t \in T} P(t) = 1 \tag{8}$$

$$t(Q) = \begin{cases} 1, & \text{if} \quad t \text{ occurs in } Q \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$P_D(t) = \sum_{t \in T} P(t') I(t, t') \tag{10}$$

$$I(t, t') = \begin{cases} 1, & \text{if} \quad t = t_D^{'} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$t_D \text{ is the closest term of } t \text{ where } d \text{ is true } (t_D \text{ occurs in } D). \tag{12}$$

Generally, $D$ is deemed relevant to $Q$ when $P(D \rightarrow Q)$ is greater than a threshold value, for example, a positive real number $\partial$. Similar to the vector space model (see Section 3.3.2), the simplest case is where at least one term occurs in both $D$ and $Q$, or it is also the closest term of some other terms occurring in $D$ and $Q$. This case is referred to as the naïve possible world-based model and the general case as the threshold possible world-based model.

[9] Actually, it can be multivalued in an interval.

[10] There is also an approach called general logical imaging that does not rely on this assumption.

4.2.2 *Naïve Possible World Aboutness Based on Crestani and Van Rijsbergen's Model* ($\models_{NAIVE-PW-CV}$). Let $U$ be the set of all the documents and T be the set of all the index terms. Furthermore, let $D \in U$ be a document, $Q$ be a query, and t be a term. The aboutness in the naïve possible world-based models is defined as follows.

- $D$ and $Q$ are sets of terms
- $D \models_{NAIVE-PW-CV} Q$ if and only if $P(D \rightarrow Q) > 0$ (Aboutness)
- $D \not\models_{NAIVE-PW-CV} Q$ if and only if $P(D \rightarrow Q) = 0$ (Surface containment)
- $D \rightarrow Q$ if and only if $D \supseteq Q$
- $Q1 \rightarrow Q2$ if and only if $Q1 \supseteq Q2$
- $t1 \rightarrow t2$ if and only if $t1$ is the closest term of $t2$ (Deep containment)
- $D1 \oplus D2 \Leftrightarrow D1 \cup D2$ (Composition)
- $Q1 \oplus Q2 \Leftrightarrow Q1 \cup Q2$
- Preclusion is foreign to this model.

4.2.3 *Inductive Evaluation*

Theorem 6. *The naïve possible world-based model supports $R$, $C$ (surface), $LM$, $RM$, $M$, and C-FA.*[11] *Postulates GLM, GRM, QLM, and QRM are inapplicable as preclusion is inapplicable.*

Proofs of LM and RM are given as follows.

—LM: Left Compositional Monotonicity is supported.

$$\begin{aligned}
&\text{Given } D1 \models_{NAIVE-PW-CV} Q, \text{ and } D = D1 \oplus D2 \\
&\Rightarrow P(D1 \rightarrow Q) - \textstyle\sum_t P_{D1}(t)t(Q) > 0, D1 \oplus D2 = D1 \cup D2 \\
&\Rightarrow \text{ At least one term } t_i \text{ is the closest term of some terms where} \\
&\quad D1 \text{ is true and } t_i \in Q, \text{ and } D1 \oplus D2 = D1 \cup D2 \\
&\Rightarrow t_i \text{ is also true in } D1 \oplus D2, \text{ and } t_i \in Q \\
&\Rightarrow P(D1 \oplus D2 \rightarrow Q) = \textstyle\sum_t P_{D1 \oplus D2}(t)t(Q) > 0, \\
&\therefore D1 \oplus D2 \models_{NAIVE-PW-CV} Q.
\end{aligned}$$

—RM: Right Compositional Monotonicity is supported.

$$\begin{aligned}
&\text{Given } D \models_{NAIVE-PW-CV} Q1, \text{ and } Q = Q1 \oplus Q2 \\
&\Rightarrow P(D \rightarrow Q1) - \textstyle\sum_t P_{\mathrm{D}}(t)t(Q1) > 0, \text{ and } Q = Q1 \oplus Q2 = Q1 \cup Q2, \\
&\Rightarrow (\exists t_i \in Q1)(\exists t_t^{'} \in T)(\mathrm{I}(t_i, t_t^{'} = 1) \text{ and } t_i \in Q \\
&\Rightarrow P(D \rightarrow Q1 \oplus Q2) = \textstyle\sum_t P_D(t)t(Q1 \oplus Q2) > 0, \\
&\therefore D \models_{NAIVE-PW-CV} Q1 \oplus Q2.
\end{aligned}$$

[11] Note that postulates MIX is trivially supported, as LM is supported. Postulate C- FA is trivially supported, as RM is supported.

#### 4.2.4 *Threshold Possible World Aboutness Based on Crestani and van Rijsbergen's Model* ($\models_{T-PW-CV}$).

Let $U$ be the set of all documents and $T$ be the set of all index terms. Furthermore, let $D \in U$ be a document, $Q$ be a query, and t be a term. The aboutness in this model is then defined as follows.

—$D$ and $Q$ are sets of terms

—$D \models_{T-PW-CV} Q$ if and only if $P(D \to Q) \geq \partial$,
where $\partial$ is a positive real number in the interval (0, 1] (Aboutness)

—$D \not\models_{T-PW-CV} Q$ if and only if $P(D \to Q) < \partial$

—The mappings of containment, composition, and preclusion are same as those in Section 4.3.2.

#### 4.2.5 *Inductive Evaluation*

THEOREM 7. *The threshold possible world-based model supports R, LM, RM, M, C-FA, and conditionally supports C, CWAA, RCM, and NR. Postulates GLM, GRM, QLM, and QRM are inapplicable as preclusion is inapplicable.*

Proofs of LM and RM are given as follows.

—LM: Left Compositional Monotonicity is supported.

Given $D1 \models_{T-PW-CV} Q$, and $D = D1 \oplus D2$

$\Rightarrow P(D1 \to Q) = \sum_t P_{D1}(t)t(Q) \geq \partial, D1 \oplus D2 = D1 \cup D2$

$\Rightarrow$ The number of index terms that are the closest terms of certain terms where $D1 \oplus D2$ is true must be not less than that of index terms that are the closest terms of certain terms where $D1$ is true. This implies that $P_{D1\oplus D2}(t) \geq P_{D1}(t)$ .

$\Rightarrow P(D1 \oplus D2 \to Q) = \sum_t P_{D1\oplus D2}(t)t(Q) \geq \sum_t P_{D1}(t)t(Q) \geq \partial,$

$\therefore D1 \oplus D2 \models_{T-PW-CV} Q.$

—RM: Right Compositional Monotonicity is supported.

Given $D \models_{T-PW-CV} Q1$, and $Q = Q1 \oplus Q2$

$\Rightarrow P(D \to Q1) = \sum_t P_D(t)t(Q1) \geq \partial$ and $Q = Q1 \oplus Q2 = Q1 \cup Q2$

(i.e.,$Q1 \subseteq Q$ and $Q2 \subseteq Q$),

$\Rightarrow P(D \to Q) = \sum_t P_D(t)t(Q1 \oplus Q2) \geq \sum_t P_D(t)t(Q1)$

$\Rightarrow P(D \to Q1 \oplus Q2) = \sum_t P_D(t)t(Q1 \oplus Q2) \geq \partial,$

$\therefore D \models_{T-PW-CV} Q1 \oplus Q2.$

### 4.3 Discussion

Deep containment is irrelevant to classical models, unless they are augmented by thesauri and the like from which deep containment relationships like *penguin* → *bird* can be extracted. Logical models, by their very nature, can directly handle deep containment relationships. This means logical models support information transformation, for example, logical imaging in the possible world models. This is a major advantage of logical models. Moreover, they provide stronger expressive power; for example, concepts such as situation, type and channel, and so on in a situation theory-based model make it more flexible.

The properties of an IR model are largely determined by the matching function it supports. Two classes of matching functions are widely used: *exact match* and *overlapping (naïve and nonzero threshold)*. The Boolean model is an example of an exact match model, which requires that all the information of the query must be contained in or can be transformed to the information of the document. The naïve vector space model and naïve possible world-based model have similar properties (except that deep containment is applicable to the possible world-based model only) due to their simple overlapping retrieval mechanism (i.e., a document is judged to be relevant if it shares at least one term with the query). Compared with the Boolean model, the naïve vector space and the naïve possible world-based models support Left and Right Compositional Monotonicity, which causes imprecision. The Boolean model supports Right Containment Monotonicity, which promotes recall, at the expense of precision. They also support the Negation Rationale, which can improve precision. For the naïve vector space- and possible world-based models, Right Containment Monotonicity and Negation Rationale are not supported. In summary, it is evident that the Boolean model is more effective than the naïve vector space- and the naïve possible world-based models.

The naïve possible world model uses imaging (i.e., imaging from non-$D$ world to $D$-world) besides simple overlapping. Even though there may exist a containment relation between a term $t1$ in the document and another term $t2$ in the query, if $t1$ is not shared by the document and the query, then this transformation from $t2$ to $t1$ is ineffective to establish the relevance. This explains why the naïve possible world model does not support Containment (deep). The mechanics of imaging is dependent on a notion of similarity between worlds. Experimental evidence shows a relation between retrieval performance and the way in which the relationship between worlds is defined [Crestani and Van Rijsbergen 1998]. As the underlying framework for inductive evaluation presented in this article does not explicitly support a concept of similarity, the mapping of the possible world-based model into the inductive framework is incomplete. More is said about this point in the conclusions.

The threshold possible world model is both left and right monotonic. As a consequence there are some grounds to conclude that this model would be imprecise in practice, and also be insensitive to document length. As mentioned in the previous paragraph, retrieval performance depends on how the similarity between worlds is defined. As both LM and RM are supported, it can be hypothesized that the baseline performance for the threshold possible world model would be similar to the naïve overlap model. More sophisticated similarity metrics between worlds would improve performance above this baseline. Crestani and Van Rijsbergen allude to this point as follows: "... it is possible to obtain higher levels of retrieval effectiveness by taking into consideration the similarity between the objects involved in the transfer of probability. However, the similarity information should not be used too drastically since similarity is often based on cooccurrence and such a source of similarity information is itself

Table I. Summary of Evaluation Results[a]

| Model Postulates | Boolean | Naïve Vector Space | Threshold Vector Space | Probabilistic Model | Situation Theory-Based | Naïve Possible World | Threshold Possible World |
|---|---|---|---|---|---|---|---|
| R | ✓ | ✓ | ✓ | CS | ✓ | ✓ | ✓ |
| C (Surface) | ✓ | ✓ | CS | CS | ✓ | ✓ | CS |
| C (Deep) | NA | NA | NA | NA | ✓ | × | CS |
| RCM (Surface) | ✓ | × | CS | CS | × | × | CS |
| RCM (Deep) | NA | NA | NA | NA | × | × | CS |
| CWAA | ✓ | × | CS | CS | × | × | CS |
| LM | ✓ | ✓ | CS | CS | ✓ | ✓ | ✓ |
| RM | × | ✓ | CS | CS | ✓ | ✓ | ✓ |
| M | ✓ | ✓ | CS | CS | ✓ | ✓ | ✓ |
| C-FA | ✓ | ✓ | CS | CS | ✓ | ✓ | ✓ |
| GLM | ✓ | NA | NA | NA | ✓ | NA | NA |
| GRM | × | NA | NA | NA | ✓ | NA | NA |
| QLM | ✓ | NA | NA | NA | ✓ | NA | NA |
| QRM | × | NA | NA | NA | ✓ | NA | NA |
| NR | ✓ | × | CS | CS | × | × | CS |

[a] NA means *not applicable*, CS means *conditionally support*, ✓ means support, and × means not supported.

uncertain" [Crestani and Van Rijsbergen 1998]. When the threshold possible world model judges a document $D$ relevant to the query Q, this implies that $D$ shares a number of terms with Q or a number of terms can be transformed to the shared terms so that $P(D \to Q)$ is not less than the threshold $\partial$. The expansion of $D$ or Q can only increase $P(D \to Q)$. This judgment is not true for the threshold vector space model, for after the expansion of $D$ (or Q), the increase of the space of $D$ (or Q) (i.e., the number of terms in $D$ and Q) may be much more than the increase of the shared terms. Thus the degree of overlapping may be decreased.

The threshold possible world model and situation theory using Lalmas' relaxed condition support LM and RM. This suggests that these models would be less precise than probabilistic and threshold vector space models. This in turn reflects the likely possibility that despite their previously mentioned expressive power, this power does not necessarily translate into precision. The scant experimental evidence available bears this out [Crestani et al. 1995].

## 5. RESULTS SUMMARY AND CONCLUSIONS

### 5.1 Result Summary

Table I presents the results.

### 5.2 Conclusion

The functional benchmarking exercise presented in this article indicates that functional benchmarking is both feasible and useful. It has been used to analyze and compare the functionality of various classical and logical IR models.

Through functional benchmarking, some phenomena encountered in experimental IR research can be explained from a theoretical point of view using a symbolic perspective. The theoretical analysis could in turn help us better understand IR and provide guidelines to investigate more effective IR models.

A major point to be drawn is that IR is conservatively monotonic in nature. It is important that conservatively monotonic models be studied and developed, as these would help achieve a better understanding of the tradeoff between precision and recall. The postulates GLM, GRM, QLM, QRM, and so on guarantee the conservatively monotonic properties, but they are foreign to some models. Even in those models, which support some of the conservatively monotonic properties, preclusion is only based on the assumption that an information carrier precludes its negation. Moreover, GLM, QLM, and MIX are the special cases of LM, and GRM, QRM, and C-FA are the special cases of RM. As such, if a model supports LM, GLM is vacuously supported. Therefore a model supporting conservative monotonicity should embody conservatively monotonic properties without supporting LM and RM. The probabilistic model and threshold vector space model show good performance in practice as they mimic conservative monotonicity.

Current logical IR models have the advantage of modeling information transformation and their expressive power. However, they are still insufficient to model conservative monotonicity. A primary reason is that important concepts, such as (deep and surface) containment, information preclusion, and the like, upon which conservative monotonicity is based, are not sufficiently modeled. For example, semantics of information preclusion is not explicitly defined in current logical models. We just simply assume that an information carrier precludes its negation during the benchmarking. It is interesting to show that if we add some kind of semantics of preclusion to the logical IR models, the conservative monotonicity could be partially realized. For example, we could add the following definition to the model.

*Preclusion*:
Given two types $\varphi 1$ and $\varphi 2$, $\varphi 1 \perp \varphi 2, s1 \models \varphi 1$ and $s2 \models \varphi 2$, there does not exist any channel between $s1$ and $s2$.
The Left composition monotonicity (LM) is no longer supported:
Given $\phi 1 \models_{ST} \phi 2$

$\Rightarrow (\exists c1 \in C)(\forall D \mid D \overset{\sim}{\rightarrow} \phi 1)(\exists \psi 1 \in \phi 1)(\exists \psi 2 \in \phi 2)\ (D \models \psi 1 \overset{c1}{\longmapsto} D' \models \psi 2)$,

$\phi 1 \oplus \phi 3 \Leftrightarrow \phi 1 \cup \phi 3$.

Assume LM is supported; that is, $(\forall D \mid D \overset{\sim}{\rightarrow} \phi 1 \oplus \phi 3)\ (\exists \psi 1 \in \phi 1 \oplus \phi 3)\ (\exists \psi 2 \in \phi 2)\ (D \models \psi 1 \overset{c1}{\longmapsto} D' \models \psi 2)$.
Consider the case of $\phi 2 \perp \phi 3$. This implies for $D \models \phi 3$ and $D' \models \phi 2$, there does not exist a channel between $D$ and $D'$. This contradicts the above assumption, because $\{\forall D \mid D \overset{\sim}{\rightarrow} \phi 1 \oplus \phi 3\} \subseteq \{\forall D \mid D \models \phi 3\}$,

$\therefore$ it is not necessary that $\phi 1 \oplus \phi 2 \models_{ST} \phi 2$.

On the other hand, RM is not supported for the similar reason of LM. However, by applying the conservative forms of monotonicity, QLM and QRM, with the qualifying nonpreclusion conditions, the above-like counterexample will no longer exist.

The above definition of preclusion is simply for the purposes of illustration. It is true that current IR systems are not explicitly defined in terms of concepts such as preclusion, information containment, and so on. However, such informational concepts are in the background. Preclusion relationships can be derived via relevance feedback [Amati and Georgatos 1996; Bruza and Van Linder 1998]. For restricted domains, information containment relationships can be derived from ontologies and the like. For example, we have been investigating the automatic extraction of deep containment relationships based on Barwise and Seligman's [1997] theory of information flow [Bruza and Song 2001; Song and Bruza 2001]. When language processing tools have advanced further, the concepts under the aboutness theory could be applied to IR more easily and more directly. More sensitive IR systems would then result, in particular those that are conservatively monotonic with respect to composition. Therefore more investigations about how to achieve conservative monotonicity in current logical IR models are necessary.

Finally, we reflect on the strengths and weaknesses of the inductive theory of information retrieval evaluation. The strengths are summarized below.

*Enhanced Perspective.* Matching functions can be characterized qualitatively in terms of aboutness properties that are, or are not, implied by the matching function in question. It may not be obvious what the implications are of a given numeric formulation of a matching function. The inductive analysis allows the teasing out of some of these implications. By way of illustration, models based on overlap may imply monotonicity (left or right), which is precision degrading. In addition, inductive analysis allows one to compute under what conditions a particular aboutness property is supported. It has been argued that a conservatively monotonic aboutness relationship promotes effective retrieval. The analysis in this article revealed that although both the threshold vector space and probabilistic models mimic conservative monotonicity, the fundaments of this support are very different: the thresholded vector space model support for conservative monotonicity depends on overlap between document and query terms modulo the size of the document. Support for conservative monotonicity in the probabilistic model depends on whether the terms being added have a high enough probability of occurring in relevant documents. From an intuitive point of view, the latter condition would seem a more sound basis for support because it is directly tied to relevance.

*Transparency.* One may disagree with a given functional benchmark (as represented by a set of aboutness properties), or with how a given matching function has been mapped into the inductive framework; however, the assumptions made have been explicitly stated. This differs from some experimental studies where the underlying assumptions (e.g., the import of certain constants) are not, or are insufficiently, motivated.

*New Insights.* The use of an abstract framework allows new insights to be gleaned. Inductive evaluation has highlighted the import of monotonicity in retrieval functions, and its effect on retrieval performance. Designers of new matching functions should provide functions that are conservatively monotonic with respect to the composition of information. More sensitive IR systems would then result. The lack of such systems currently can be attributed in part to the inability to effectively "operationalize" information preclusion. Most common IR models are either monotonic or nonmonotonic; another class of IR models, namely, those that are explicitly conservatively monotonic is missing. For this reason, the inductive analyses reported in this article revealed no distinctions based on conservatively monotonic rules such as MIX and CF-A. Conservatively monotonic models are interesting for purposes of producing a symbolic inference foundation to query expansion and perhaps even relevance feedback.

The weaknesses of an inductive theory for evaluation are as follows.

*Difficulty in Dealing with Weights.* Much of the subtlety of IR models remains buried in different weighting schemes. Due to its symbolic nature, the inductive approach can abstract "too much," thereby losing sensitivity in the final analysis. For example, the nuances of document length normalization [Singhal et al. 1996], term independence assumptions, and probabilistic weighting schemes are difficult, if not impossible, to map faithfully into a symbolic inductive framework.

*Difficulties with Mapping.* For an arbitrary model, it may not be obvious how to map the model into an inductive framework. This is particularly true for heavily numeric models such as probabilistic ones. It is often the case that such models do not support many symbolic properties—they are like black holes defying analysis [Bruza et al. 2000a]. However, analyzing the conditions under which given properties are supported allows us to "peek at the edges of the black hole."

*Incompleteness of Framework.* In order to pursue functional benchmarking, a sufficiently expressive framework is necessary in order to represent salient aspects of the model in question. This is an issue of completeness. In the inductive analysis of the possible world-based models presented in this article, we have seen that the notion of similarity inherent to these models cannot be directly translated into the underlying inductive framework. This suggests that the framework presented in this article should be extended. One could also argue that not all salient aspects of aboutness have been captured by the properties used for the benchmark. These are not criticisms of inductive evaluation, but of the expressiveness of the underlying informational framework, in this case information fields.

It is noteworthy that conventional experimental IR evaluation approaches are reasonably solid but sometimes fail to address deeper issues. Functional benchmarking is a framework and methodology that can help fill this gap. It is not intended to replace the former, but to complement it.

# APPENDIX

## A. List of Notations

Information carrier (IC)
Information composition ($\oplus$)
Information containment ($\rightarrow$)
Surface containment ($\xrightarrow{s}$)
Deep containment ($\xrightarrow{d}$)
Information preclusion ($\perp$)
Aboutness ($\models$)
Nonaboutness ($\not\models$)
A document $D$ (or a query $Q$) consisting of information carrier $i(D \overset{\sim}{\rightarrow} i$ or $Q \overset{\sim}{\rightarrow} i)$

## ACKNOWLEDGMENTS

P. Bruza would like to thank Gianni Amati, Theo Huibers, Bernd van Linder, and Keith van Rijsbergen for the interesting discussions in the topic area of this article. The authors thank the anonymous reviewers for their helpful comments.

## REFERENCES

AGOSTI, M. AND SMEATON, A. F., EDS. 1996. *Information Retrieval and Hypertext*. Kluwer, Hingham, Mass.

AMATI, G. AND GEORGATOS, K. 1996. Relevance as deduction: A logical view of information retrieval. In *Proceedings of the Second Workshop on Information Retrieval*, *Uncertainty and Logic* (Glasgow).

BAEZA-YATES, R. AND RIBEIRO-NETO, B. 1999. *Modern Information Retrieval*. ACM Press and Addison-Wesley, New York.

BARWISE, J. 1989. *The Situation in Logic*. In *CLSI Lecture Notes* 17, Stanford, Calif.

BARWISE, J. AND ETCHEMENDY, J. 1990. Information, Infons and Inference. In *Situation Theory and its Applications*, R. Cooper, et al. Eds., *CLSI Lecture Notes* 1, 33–78.

BARWISE, J. AND SELIGMAN, J. 1997. *Information Flow—The Logic of Distributed Systems*. Cambridge University Press, Cambridge, MA.

BRUZA, P. D. AND HUIBERS, T. W. C. 1996. A study of aboutness in information retrieval. *Artif. Intell. Rev*. 10, 1–27.

BRUZA, P. D. AND HUIBERS, T. W. C. 1994. Investigating aboutness axioms using information fields. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin), 112–121.

BRUZA, P. D. AND LALMAS, M. 1996. Logic based information retrieval: Is it really worth it? In *Proceedings of WIRUL 96*, *the Second Workshop on Information Retrieval*, *Uncertainty and Logic* (Glasgow).

BRUZA, P. D. AND VAN LINDER, B. 1998. Preferential models of query by navigation. In *Information Retrieval*, *Logic and Uncertainty*, F. Crestani, M. Lalmas, and C. J. van Rijsbergen, Eds. Springer-Verlag, New York.

BRUZA, P. D. AND SONG, D. 2001. Informational inference via information flow. In *Proceedings of the Twelfth International Workshop on Database and Expert Systems Applications* (Munich, September 3–7), 327–341.

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

BRUZA, P. D., SONG, D., WONG, K. F., AND CHENG, C. H. 2000b. Commonsense aboutness for information retrieval. In *Proceedings of the International Conference on Advances in Intelligent*

*Systems*: *Theory and Applications* (*AISTA 2000*) (Canberra, Australia, February 2–4), 317–324.

Bruza, P. D., Song, D., and Wong, K. F. 1999. Fundamental properties of aboutness. In *Proceedings of the Twenty-Second Annual International ACM–SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR'99*) (Berkeley, Calif., August 15–19).

Crestani, F., Ruthven, I., Sanderson, M., and van Rijsbergen, C. J. 1995. The troubles with using a logical model of IR on a large collection of documents. In *Proceedings of the Fourth Text Retrieval Conference* (*TREC-4*), 509–526.

Crestani, F. and van Rijsbergen, C. J. 1995a. Information retrieval by logic imaging. *J. Doc. 51*, 1, 3–17.

Crestani, F. and van Rijsbergen, C. J. 1995b. Probability kinematics in information retrieval. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* (Seattle, Wash.), 291–299.

Crestani, F., Lalmas, M., and van Rijsbergen , C. J. 1998. A study of probability kinematics in information retrieval. *ACM Trans. Inf. Syst. 16*, 3.

Crestani, F., and van Rijsbergen, C. J. 1998. A study of Probability kinematics in information retrieval. *ACM Trans. Inf. Sys.*, 16, 3.

Dubois, D., Farinas del Cerro, L., Herzig, A., and Prade, H. 1997. Qualitative relevance and independence: A roadmap. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence* (*IJCAI-97*), 62–67.

Frakes, W. B. and Baeza-Yates, R., Eds. 1992. *Information Retrieval, Data Structures & Algorithms*. Prentice-Hall, Englewood, Cliffs, N.J.

Huibers, T. W. C. 1996. An axiomatic theory for information retrieval. PhD thesis, Utrecht University, The Netherlands.

Huibers, T. W. C. and Bruza, P. D. 1994. Situations, a general framework for studying information retrieval. In *Information Retrieval*: *New Systems and Current Research, vol. 2*, Taylor Graham, Ed.

Huibers, T. W. C., Lalmas, M., and van Rijsbergen, C. J. 1996. Information retrieval and situation theory. *SIGIR Forum 30*, 1, 11–25.

Hunter, A. 1995. Using default logic in information retrieval. In *Symbolic and Quantitative Approaches to Uncertainty*, Lecture Notes in Computer Science, vol. 946, Springer-Verlag, New York, 235–242.

Hunter, A. 1996. Intelligent text handling using default logic, In *Proceedings of the Eighth IEEE International Conference on Tools with Artificial Intelligence* (*TAI'96*), IEEE Computer Society Press, Los Alamitos, Calif., 34–40.

Hutchins, W. J. 1977. On the problem of 'aboutness' in document analysis. *J. Inf. 1*, 1, 17–35.

Kraus, S., Lehmann, D., and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell. 44*, 167–207.

Lalmas, M. 1998. Logical models in information retrieval: Introduction and overview. *Inf. Proc. Manage. 34*, 1, 19–33.

Lalmas, M. 1997. Information retrieval and Dempster–Shafer's theory of evidence. In *Applications of Uncertainty Formalisms*, A. Hunter and S. Parson, Eds., Chapter 8, Lecture Notes in Computer Science, Springer-Verlag, New York.

Lalmas, M. 1996. Theories of information and uncertainty for the modeling of information retrieval: An application of situation theory and Dempster–Shafer's theory of evidence. PhD thesis, University of Glasgow.

Lalmas, M. and Bruza, P. D. 1998. The use of logic in information retrieval modeling. *Knowl. Eng. Rev*. In press.

Landman, F. W. 1986. *Towards a Theory of Information. The Status of Partial Objects in Semantics*. Foris, Dordrecht.

Losee R. M. 1998. *Text Retrieval and Filtering*: *Analytic Models of Performance*. Kluwer, Hinghamn, Mass.

Losee R. M. 1997. Comparing Boolean and probabilistic information retrieval systems across queries and disciplines. *J. Am. Soc. Inf. Sci. 48*, 2, 143–156.

Maron, M. E. 1977. On indexing, retrieval and the meaning of about. *J. Am. Soc. Inf. Sci. 28*, 1, 38–43.

NIE, J. 1992. Towards a probabilistic modal logic for semantic-based information retrieval. In *Proceedings of the ACM–SIGIR Conference on Research and Development in Information Retrieval* (Copenhagen), 140–151.

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

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

PROPER, H. A. AND BRUZA, P. D. 1999. What is information discovery about? *J. Am. Soc. Inf. Sci. 50*, 9, 737–750.

VAN RIJSBERGEN, C. J. 1993. The state of information retrieval: Logic and information. *Comput. Bull.*, February.

VAN RIJSBERGEN, C. J. 1986a. A non-classical logic for information retrieval. *Comput. J. 29*, 6, 481–485.

VAN RIJSBERGEN, C. J. 1986b. A new theoretical framework for information retrieval. In *Proceedings of the Ninth International SIGIR Conference in Research and Development in Information Retrieval*, 194–200.

VAN RIJSBERGEN, C. J. 1989. Towards an information logic. In *Proceedings of the 12th International SIGIR Conference in Research and Development in Information Retrieval*, pp. 77–86.

VAN RIJSBERGEN, C. J. 1979. *Information Retrieval*, Second edition. Butterworths, London.

VAN RIJSBERGEN, C. J. AND LALMAS, M. 1996. An information calculus for information retrieval. *J. Am. Soc. Inf. Sci. 47*, 5, 385–398.

ROELLEKE, T. AND FUHR, N. 1996. Retrieval of complex objects using a four-valued logic. In *Proceedings of the ACM–SIGIR Conference on Research and Development in Information Retrieval* (Zurich), 206–214.

SALTON, G. 1988. *Automatic Text Processing*. Addison-Wesley, Reading, Mass.

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

SINGHAL, A., BUCKLEY, C., AND MITRA, M. 1996. Pivoted document length normalization. In *Proceedings of the Nineteenth ACM–SIGIR Conference on Research and Development in Information Retrieval* (Zurich), 21–29.

SONG, D. 2000. A commonsense aboutness theory for information retrieval modeling. PhD thesis. The Chinese University of Hong Kong.

SONG, D. AND BRUZA, P. D. 2001. Discovering information flow using a high dimensional conceptual space. In *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval* (*SIGIR'01*) (New Orleans, La., Sept. 9–13), 327–333.

SONG, D., WONG, K. F., BRUZA, P. D., AND CHENG, C. H. 2000a. Fundamental properties of the core matching functions for information retrieval. In *Proceedings of the Thirteenth International Florida Artificial Intelligence Society Conference* (*FLAIRS* 2000) (Orlando, Fl., May 22–24), 118–122.

SONG, D., WONG, K. F., BRUZA, P. D., AND CHENG, C. H. 2000b. Towards a commonsense aboutness theory for information retrieval modeling. In *Proceedings of the Fourth World Multiconference on Systemics, Cybernetics and Informatics* (*SCI* 2000) (Orlando, Fl., July) 23–26.

SONG, D., WONG, K. F., BRUZA, P. D., AND CHENG, C. H. 1999. Towards functional benchmarking of information retrieval models. In *Proceedings of the Twelfth International Florida Artificial Intelligence Society Conference* (*FLAIRS'99*) (Orlando, Fl., May 3–5), 389–393.

TURTLE, H. R. AND CROFT, W. B. 1992. A comparison of text retrieval models. *Comput. J. 35*, 3, 279–290.

XU, J. AND CROFT, B. 1996. Query Expansion Using Local and Global Document Analysis. In *Proceedings of the 19th International SIGIR Conference in Research and Development in Information Retrieval*, pp. 4–11.

Received April 1999; revised May 2001; accepted May 2001