On Information Retrieval and Evidential Reasoning

Lung Albert Chen

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, CA. 94720

ABSTRACT

Many research efforts have been devoted to solving the problems of Boolean systems, which are currently used for Information Retrieval(IR). We propose a new model of IR, which treats the whole process of IR as a process of evidential reasoning. Our model is knowledge based, and theoretically sound. An input query provided by a user, triggers the process of evidential reasoning. The process consists of two parts: automatic query formulation and query evaluation. Automatic query formulation maps a concept given by the user into a set of textual terms. These terms, according to the pieces of evidence given by an expert, have been used by various authors to describe the concept specified in the input query. Query evaluation is an evidence-aggregation scheme, that combines all the pieces of evidence and assigns a Retrieval Status Value(RSV) to each document. A list of documents, ranked according to the RSV, is provided to the user as a response to his or her information request. In our model, inference strength between concept and subconcept is measured by conditional basic probability assignment; and this measure is discounted, chained, and combined based on the Dempster-Shafer(D-S) theory and its extension.

This work was supported by the Sloan Foundation

1. Introduction

Conceptually, an Information Retrieval (IR) system consists of a set of information items(e.g., documents) and a mechanism for answering queries by retrieving appropriate information items [SaM83]. In this research, we focus on document-retrieval systems, and treat information retrieval as a process of retrieving relevant documents according to the information need of the users. Because the notion of *relevance* is subjective, the information request in IR may not be satisfied exactly.

The standard Boolean keyword model is used by most of the commercially available on-line systems, such as DIALOG, LEXIS, and MEDLARS. The Boolean systems have some serious drawbacks. For example, the Boolean systems have inhospitable request formalism; they frequently provide null outputs and overloaded outputs. In addition, the user is not able to place different emphasis on different facets of the search [Coo88].

One way to solve the problems of the Boolean systems is to generalize the query evaluation procedure used by them. In Boolean systems, a document is either relevant or not relevant. There is no possibility for partial relevance. It is important to rank any retrieved document, in respect to a given query, according to its relevance value. Relevance value is also called *retrieval status value*, or *RSV*. There are three major approaches to ranking documents according to relevance: vector space models, probabilistic models, and fuzzy set retrieval models.

In general, these three approaches provide a sound theoretical basis for the development of IR systems. Each approach has successfully demonstrated the ranking capability of its query evaluation process. It should be noted that each approach interprets the meaning of the weights differently. Also, according to previous experimental results, each system provides quite acceptable results.

1.1. Motivation

All these approaches share one common assumption, that is, they assume the user is able to describe his or her information need, in a consistent way, by using the query languages adopted by these systems. For these systems, the query formulation has two phases. The first one is for the user to choose proper terms to describe the information need; and the second one is for the user to place the proper weight on each term . Both steps may be very difficult for general end users, especially when the weights are interpreted differently in different systems. Usually, a user has to go through several experiments in order to get a better result; therefore, many users must consult an expert such as a librarian to perform the query formulation.

The motivation of this research is to design an IR system to replace the expert needed in query formulation. In other words, the user should be able to specify his or her information need through a friendly user interface; and then have the system generate a proper conceptual structure which best describes the information need. This conceptual structure should contain the proper concepts and proper weights, according to the expert's knowledge stored in the system. Since query evaluation is closely related to the query formulation, we need a new model of query evaluation to incorporate the expert's knowledge in query formulation.

1.2. Research Issues

This research focuses on five questions:

(1) Given a query as the evidence of the user's information need, to what degree of belief a single document should be retrieved?

(2) Given a concept, what other concepts and sub-concepts have been used by various authors in describing the concept, and to what degree of belief?

(3) What is the interpretation of the degree of belief, in other words, what is the meaning of the weight?

(4) Given all the pieces of expert's knowledge, how to construct a conceptual structure which best describes the given query?

(5) Given the conceptual structure and the attached degree of belief, what RSV should be assigned to each individual document?

Each of these questions addresses either the system, knowledge acquisition, and/or the user.

2. Early Work

2.1. Vector Space Models

Vector space models are able to rank retrieved documents based on a simple similarity measure, for example, the cosine correlation function used in the SMART system [Sal71]. However, vector space models are not able to process structured queries. Another drawback of vector space systems is that term vectors are assumed to be pairwised orthogonal, otherwise one can not compute the similarity between a query and each individual document. This orthogonal assumption is too restrictive.

2.2. Probabilistic Models

The probabilistic models put retrieval theory on a statistical basis [Boo85]. They recognize that the retrieved documents should, in general, be ranked in descending order of probability of usefulness to the user. This is the Probability Ranking Principle [Coo76], [Rob77], [CoM78], and [RMC82]. Probabilistic models of indexing, also called "Model 1" type approaches to probabilistic retrieval, attempt to apply probability estimation techniques to document indexing. The first model of this kind was proposed by Maron and Kuhns [MaK60]. In their model, the question asked of the indexers is reformulated in helpful ways, and the analysis is extended to multi-term search requests. Probabilistic models of retrieving, also called "Model 2" type approaches, assign probability interpretation of the weights placed on request terms. This kind approach was first explored by Robertson and Sparck-Jones [RoS76], who proposed a formula for the estimation of probability of user satisfaction that could be used as the retrieval rule. The third model, proposed by Robertson, Maron and Cooper [RMC82], unifies the two previous approaches, by using relevance feedback information from the individual user about other documents, and from other users about the individual document.

In probabilistic models, an important assumption is that the probability measure on the event space is uniform. That is, the probability for a request-document pair is independent of the chosen pair. Also, some estimation is necessary for the marginal probabilities; and there is the *curse of dimensionality* [Rij79], which forces the probabilistic models to include only query terms in query evaluation [Bar85].

2.3. Fuzzy Set Models

The fuzzy set retrieval models are based on fuzzy set theory [Zad65]; and they interpret the term weight as a subjective measure of relevance of importance, rather than an objective probabilistic measure. The early works of fuzzy set information retrieval models were done by Tahani and Radecki [Tah76], [Rad76]. They provide the formal mathematical frameworks for fuzzy set models in IR; and they talk about fuzzy query processing with fuzzy predicates embedded in a query, which is treated as a *pseudo document*. Later on, threshold models [Rad79], [BuK81a], [BuK81b] were introduced for any generalized Boolean query; however there are problems of consistency in generalizing Boolean queries to include relevance weights and thresholds [Bue81], [BuK81a]. This difficulty is overcome by Bookstein [Boo80], who proposes that the Boolean operator act differently on the same set depending on the context in which the set is found.

3. Research Background

There are two systems related to this research. The first one is RUle-Based Retrieval of Information by Computer (RUBRIC) [TSD83], [TAC85], which generates a query structure according to the user's knowledge. The second one is Knowledge-Assisted Document Retrieval (KADR) [BBS87], which extends the user's Boolean query according to expert knowledge.

The notion of *evidential reasoning* in IR is introduced in RUBRIC; and it is formalized in KADR, which uses Dempster-Shafer (D-S) theory. Evidential reasoning in both systems treat documents as pieces of evidence in addressing the issue that to what degree a single document is relevant to a given query. RUBRIC uses a strong evidence model since it's goal is to support full text retrieval; while KADR uses evidential reasoning model to support preprocessed document descriptors.

Since we are interested in knowledge-based automatic query formulation and query evaluation, we will focus on these parts in RUBRIC and KADR. Also, we will briefly introduce D-S theory before reaching the KADR system.

3.1. Related Work I: RUBRIC

RUBRIC is a full text retrieval system developed by Tong et. al. It is a rule-based system, which uses rules to represent a user's knowledge, or, the user's preference in retrieving information. A conceptual hierarchy is the result of chaining of rules; and this conceptual structure is used in matching with each document in deciding the RSV of the document.

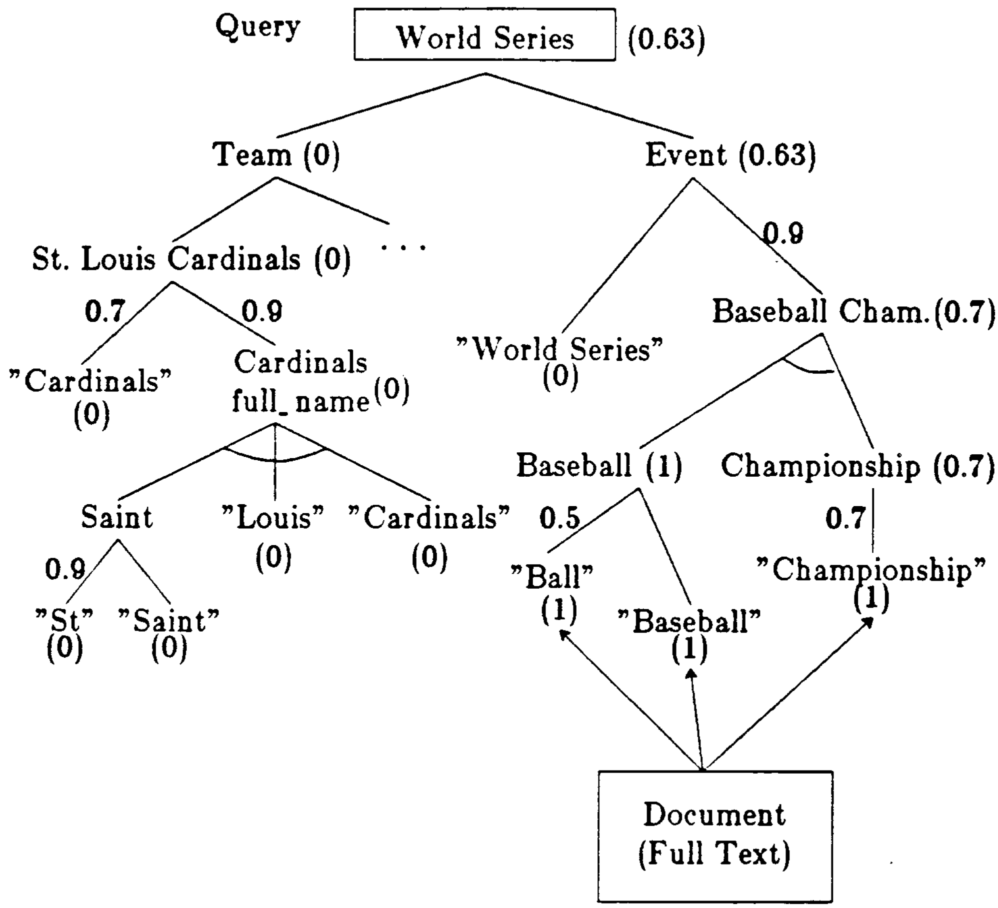

Fig. 1: Example query in RUBRIC

3.1.1. Automatic Query Formulation

Given a single concept which describes the information need of a user, the way RUBRIC generates conceptual structure in response to the given concept is based on the pieces of knowledge provided by the user. As in general expert systems, a piece of knowledge is represented in the form of a rule. These rules are interpreted as a hierarchy of retrieval concepts and sub-concepts. Thus by naming a single concept, the user automatically invokes a goal oriented search of the tree defined by all of the sub-concepts that are used to define that concept. The lowest-level sub-concepts are themselves further defined in terms of pattern expressions in a text reference language, usually, contextual terms. Then, the whole structure is used to assign RSV to each document.

As a simple example of a query, let us use the "1982 World Series " as shown in Fig. 1. Each arc in the tree has an attached relevance value such that the intermediate topics and keyword expressions contribute, according to their relevance values, to the overall relevance that the document has to the root topic. Unlabeled arcs have an implicit relevance value of 1.0; and arcs representing the conjunctions of an AND expression are linked together near their common base. Here, the meaning of relevance value has the same interpretation of the meaning of weights in fuzzy set models, which is a subjective, semantic similarity measure. Two rules used in generating this structure are:

Baseball Championship $\rightarrow$ *Event*, (0.9)

Ball $\rightarrow$ *Baseball*, (0.5)

where the first rule means "Baseball Championship" *implies* "Event" to the degree 0.9, while the second one means that "Ball" is an *evidence* of "Baseball", to the degree 0.5.

3.1.2. Query Evaluation

Evidential reasoning in RUBRIC starts by recognizing a document as the "evidence" on which the system determines the relevance value of that document to the retrieval request. For example, regarding the query structure in Fig. 1., if a document contains the words "ball", "baseball", and "championship", but no other words referred to the example rule-base, then leaf nodes of "ball", "baseball", and "championship" all receive a value of 1.0, showing that there is "strong" evidence that this document is relevant to these concepts; and all the other nodes receive a value of 0.0, which means there is no evidence that this document is relevant to these concepts.

The relevance values at the leaf nodes are then propagated across the rules, toward the root concept. As a result, the nodes of the tree would be assigned the relevance values shown in parentheses in the same figure. In this example, the overall relevance measure between the document and the given query is 0.63. The propagation of the relevance values is governed by the similarity measure on the ark. In this example, a calculus that models conjunction with the minimum operation, disjunction with the maximum operation, and uses arithmetic product as the detachment operator is chosen to propagate the relevance values. The detail of these operations leads toward some issues in selecting uncertainty calculi [ToS85], [BIJ87] which will not be discussed here.

3.1.3. Summary

RUBRIC is a successful example of applying ideas from Artificial Intelligence (AI) in the development of a computer-based aid for information retrieval; the system is fully implemented and commercialized. RUBRIC has several important features: The matching is performed over the whole document. The RSV of a document is a relevance value in the range [0, 1]. Queries are expressed in a language of rules that allows the user to develop hierarchical knowledge structures of retrieval concepts. Also, users are provided with a collection of graphic tools.

One problem is that Rubric assumes that a user will conceptualize a query consistently over time. This is not always the case for the general user. Another problem is that Rubric assumes all users are knowledgeable about the domain in which they are interested and are knowledgeable about the type of documents they want to retrieve. More general cases fail to have this level of sophistication.

3.2. Evidential Reasoning and D-S Theory

The D-S theory of evidence was first introduced by Dempster [Dem67] and extended in subsequent work by Shafer [Sha76]. Because of its theoretical soundness, strength and generality, D-S theory has, in recent years, received increasing attention from AI researchers [GoS84]; and the coherent approach suggested by D-S theory in aggregating pieces of evidence bearing on hypothesis groups is called evidential reasoning [LoG82].

The D-S theory originated from the concept of lower and upper probability induced by a multivalued mapping [Dem67]. A multivalued mapping, Γ, from an evidence space E to a hypothesis space Θ, associates each element in E with a set of elements in Θ, i.e., $\Gamma : E \rightarrow 2^{\Theta}$. The element-subset compatibility relation is denoted by ":$\rightarrow$". Given a multivalued mapping and a probability distribution of space E, a basic probability assignment(bpa) of space Θ, denoted by m: $2^{\Theta} \rightarrow [0,1]$ is induced. The bpa is also called mass assignment, mass distribution, or granular distribution. The basic probability value of a subset F of space Θ is *:

$$m(F) = \sum_{t_i :\rightarrow F} P(t_i) \tag{3.1}$$

where $P(t_i)$ is the probability judgement over $t_i \in E$. The subset F is also called *focal element*; the space Θ is the *frame of discernment*. It is easy to show that a legal bpa must have the following properties:

$$\sum_{F \subseteq \Theta} m(F) = 1, \qquad m(\emptyset) = 0 \tag{3.2}$$

In general, the probability distribution on the space Θ is constrained by the bpa. And an interval $[Bel(X), Pls(X)]$ is used to measure the degree to believe X, which is an arbitrary set in space Θ. Here, the lower bound $Bel(X)$ denotes the *belief* of X which counts the degree of belief *necessarily* committed to the set X, whereas the upper bound $Pls(X)$ is the *plausibility* of X, which expresses the degree of belief *possibly* committed to X. Mathematically, these belief and plausibility functions are defined by

$$Bel(X) = \sum_{F \subseteq X} m(F) \tag{3.3}$$

$$Pls(X) = \sum_{F \cap X \neq \emptyset} m(F) \tag{3.4}$$

* For simplicity, we assume that there is no mapping between the elements of the space E and the empty set.

It should be noted that (1) $Bel(X) = 1 - Pls(\overline{X}) \leq Pls(X)$, (2) $Bel(X) + Bel(\overline{X}) \leq 1$, and (3) the interval $[Bel(X), Pls(X)]$ will reduce to a pointwise Bayesian probability $P(X)$ if all the focal elements are singletons, i.e., no composite set has probability > 0. In this context, $Bel(X) + Bel(\overline{X}) = 1$ [Sha76].

If m_1 and m_2 are two bpas induced from two independent evidential sources, a third bpa, $m(C)$, expressing the pooling of the evidence from the two sources, can be computed by using Dempster's rule of combination:

$$m(C) = (m_1 \oplus m_2)(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i)\, m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)\, m_2(B_j)}, \tag{3.5}$$

where A_i, B_j, and C are focal elements in Θ.

Dempster's rule of combination normalizes the intersection of the bodies of evidence from the two sources by the amount of *nonconflictive evidence* between the sources. This amount is represented by the denominator of the formula. Sometimes, eqn. (3.5) is also expressed as

$$m(C) = (m_1 \oplus m_2)(C) = K \cdot \sum_{A_i \cap B_j = C} m_1(A_i)\, m_2(B_j) \quad , \tag{3.6}$$

where K is the normalization factor, and has the value:

$$K = \frac{1}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)\, m_2(B_j)}$$

D-S theory offers following advantages over other approaches [Yen86], [Liu87], [GoS84].

(1) It allows a coherent expression of information ignorance. That is, commitment of belief in a focal element does not imply commitment of the remaining belief to its negation, part of the belief can be reserved to the "don't-know" choice.

(2) Evidence may bear on groups of hypothesis and be combined in a coherent way. In other words, it is able to model the narrowing of the hypothesis set with the accumulation of evidence.

(3) The D-S theory subsumes the Bayesian theory in that under certain circumstances, it reduces to the Bayesian. The D-S theory is "probability-related", whereas the Bayesian is "probability-based"; therefore the D-S theory does not require full information concerning probabilities.

3.3. Related Work II: KADR

KADR is another knowledge-based retrieval system. Unlike RUBRIC, KADR uses natural language processing technique in query formulation; however, the internal representation of the query is still simple Boolean query. Since KADR proposes a general evidential reasoning model for deciding the RSV of a document, we will focus on the evidential reasoning part in KADR.

3.3.1. Evidential Reasoning in KADR

Evidential reasoning is used in KADR in figuring out the degree of relevance between a given document and a given concept. Concepts in KADR can be treated as index terms. In KADR, a document is represented as a fuzzy set of concepts; in other words, each concept in this set is attached with a degree of membership:

$$d = \{c_{di}, \mu_d(c_{di})\}, \quad i = 1, \ldots, n$$

where d is a document, c_{di} stands for a concept contained in this document and μ_d is the characteristic function of d, which maps from the concept space to the range [0,1]. Also, in KADR, the relation between concepts is represented as a fuzzy measure. An implication relation is given by the expert:

$$I(c_i, c_j) = \mu_I(c_i, c_j),$$

where μ_I measures the degree of relevance between a given pair of concepts. A concept is relevant to itself to the degree 1.0.

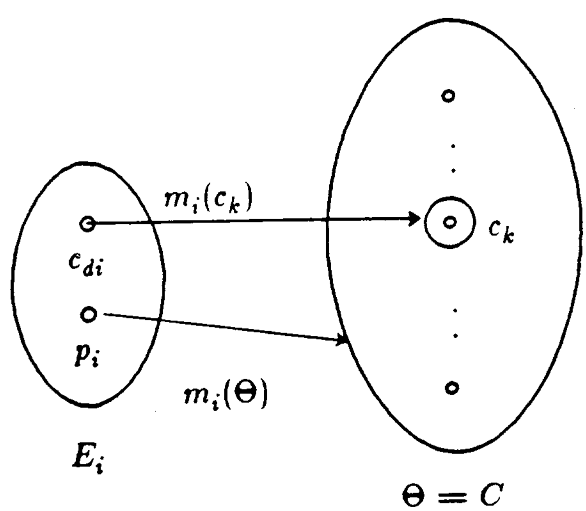

Fig. 2: Mass assignment in KADR

The mass assignment in KADR is decided by the two measures described above. For each concept in a document, or each member in the fuzzy set associated with a given document, an evidence space is constructed for that concept. For a given evidence space, the mapping between it and the concept space is shown in Fig. 2. We have on the left an evidence space E_i which contains concept c_{di}, and on the right the concept space C. A mapping is constructed between c_{di} and each concept implied by it, say c_k. If there is no such c_k, then a mapping is constructed from c_{di} to each concept c_k which implies c_{di}. The mass assignment associated with the mapping between c_{di} and c_k is defined as:

$$m_i(\{c_k\}) = \mu_d(c_{di}) \cdot I(c_{di}, c_k), \tag{3.7}$$

in which m_i stands for the mass assignment of the evidence space associated with the concept c_{di}. Knowing all the mappings from c_{di} to C, the ignorance is measured by:

$$m_i(\Theta) = 1 - \sum m_i(\{c_k\}), \quad c_{di} \rightarrow c_k \tag{3.8}$$

where Θ is the hypothesis space, in this case, C; and $\rightarrow$ means implies or implied by. This ignorance measure is represented by a mapping from a pseudo node p_i in E_i to the whole concept space C.

Given all the evidence spaces and the mappings between them and the concept space. The measure of all the evidence spaces are combined according to Dempster's rule of combining. After combining, each concept of the concept space has a derived mass assignment, which is the measure specifying to what degree the given document is relevant to this concept:

$$r(d, c_k) = m(c_k) = (m_1 \oplus m_2 \oplus \ldots \oplus m_n)(c_k), \tag{3.9}$$

where r is the relevance function between a given document and a given concept, n is the number of concepts in d; and $\oplus$ is the combining operator in D-S theory. Give a query expressed as a Boolean structure of the concepts in C, the relevance values of the concepts are combined using geometric mean and linear interpolation functions, corresponding to the *and* and *or* operators, respectively:

$$v(c_i \text{ and } c_j) = (v(c_i) \cdot (c_j))^{1/2}, \tag{3.10}$$

$$v(c_i \text{ or } c_j) = v(c_i) + (1 - v(c_i)) \cdot v(c_j), \tag{3.11}$$

where c_i and c_j are concepts in the query. Using these functions, given a Boolean query, which is the internal representation of a query in KADR, and a document, the overall relevance value between the query and the given document can be calculated.

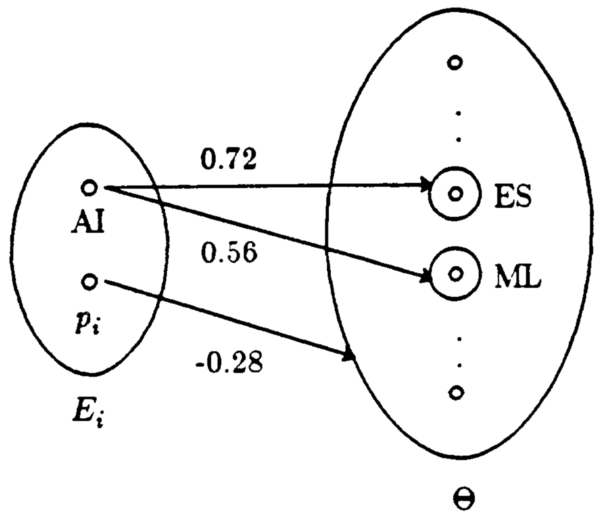

Fig. 3: Counter example I of KADR

3.3.2. Counter Examples

Unfortunately, there are certain situations that the evidential reasoning model in KADR does not work. There are two counter examples. The first one is shown in Fig. 3. Suppose we have a document which is related to the concept AI, to the degree 0.8; and suppose the concept AI implies the concept Expert System (ES), to the degree 0.9, and the concept Machine Learning (ML), to the degree 0.7. According to eqn. 3.7, the mass assignment of the concepts ES and ML is as follows:

$$m_i(ES) = 0.8 \cdot 0.9 = 0.72,$$
$$m_i(ML) = 0.8 \cdot 0.7 = 0.56.$$

where m_i is the measure associated with the evidence space containing the concept AI. In this case, the ignorance measure is equal to:

$$m_i(\Theta) = 1 - 0.72 - 0.56 = -0.28,$$

which is conflict with the D-S theory in which all the measure must be a positive number between [0, 1]. Therefore, the definition of mass assignment in eqn. 3.7 is not justified.

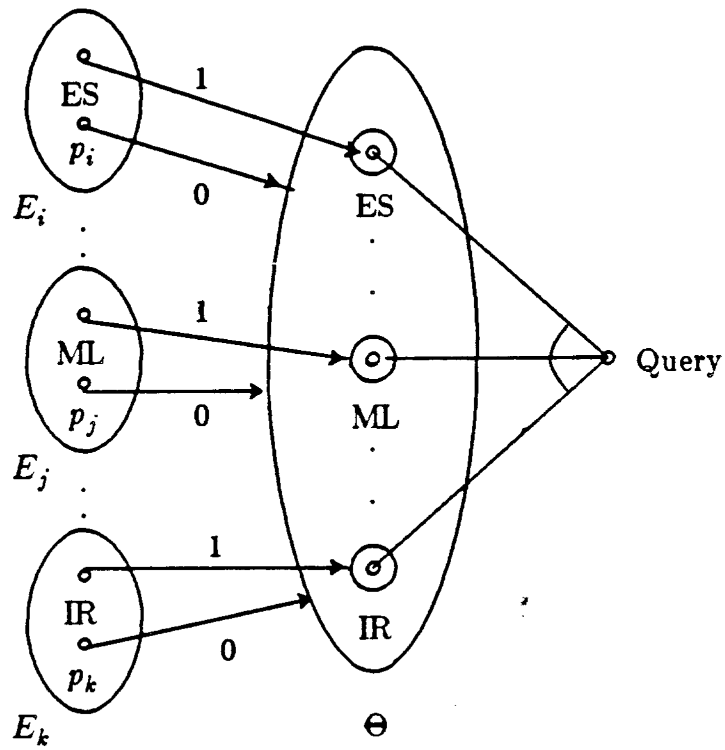

Fig. 4: Counter example II of KADR

The second counter example occurs when we try to see whether KADR subsumes the Boolean system. In Boolean system, there is no imprecise measure; in other words, each document is represented as a crisp set, and the implication relation between concepts is exact matching. This can be represented in KADR that each document contains only concepts which have degree of membership equal to one, and each concept only implies itself, to the degree one. As a result, the ignorance between each evidence space and the concept space is zero.

For example, suppose we have a document which is related to the concepts ES, ML, and IR, and the implication relation associated with them:

$$d = \{(ES,\ 1),\ (ML,\ 1),\ (IR,\ 1)\},$$
$$I(ES,\ ES) = I(ML,\ ML) = I(IR,\ IR) = 1$$

Fig. 4 shows the mapping between the evidence spaces and the concept space. Where E_i, E_j, E_k stand for the evidence space containing the concept ES, ML and IR, respectively. Notice that the ignorance measure is always zero:

$$m_i(\Theta) = m_j(\Theta) = m_k(\Theta) = 0, \tag{3.12}$$

where m_i, m_j, and m_k is the measure of E_i, E_j, and E_k, respectively. Eqn. 3.12 is consistent with the Boolean system.

Now suppose we have a query: ES *and* ML *and* IR. In Boolean system, document d will be retrieved since it contains all the concepts specified in the query. However, in KADR, document d will not be retrieved. This is because the measures of the three evidence spaces conflict with one another, and can not be combined using Dempster's rule. Therefore, the overall relevance function between the query and d is zero.

4. Current Research

This research provides a general framework for knowledge-based information retrieval. Unlike RUBRIC, the automatic query formulation in our model is based on expert's knowledge, rather than user's knowledge. All the pieces of partial information are combined, in a consistent way, in the process of query evaluation. The overall measure of degree of belief, in Shafer's term, between the query structure and a single document, is assigned to that document as it's RSV.

4.1. Automatic Query Formulation

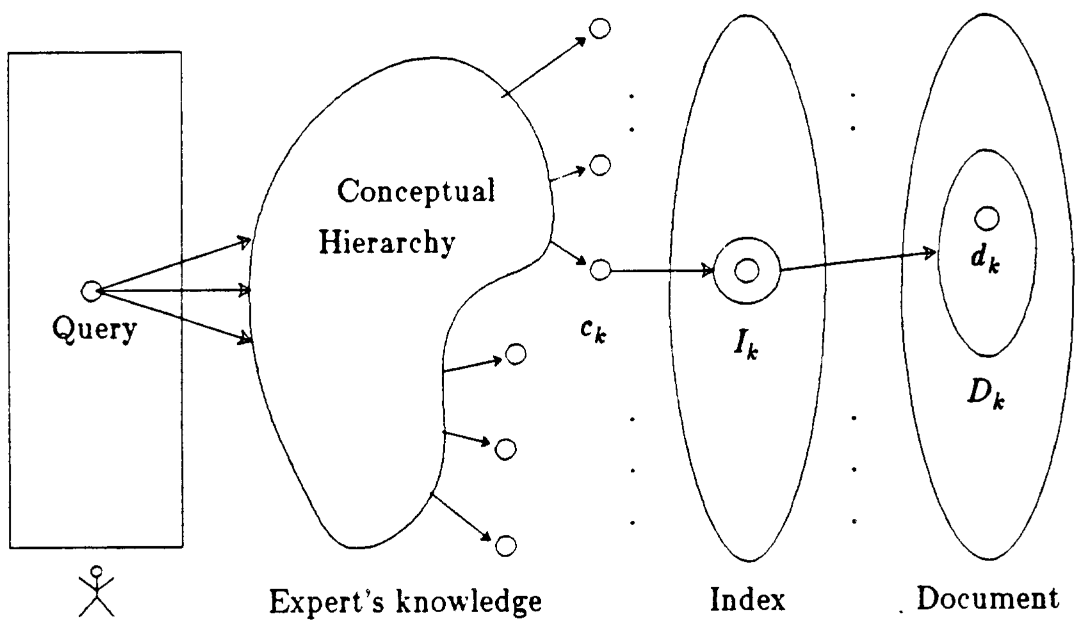

Fig. 5: Overview of our model

An overview of our model is shown in Fig 5. We have an index space which contains all the index terms; and a document space which contains all the documents, which are represented by their descriptors. The mapping between index space and document space is provided by an indexer; in other words, once a descriptor, or an index record, is chosen for each document, an index term will map to a set of documents which have that term in their descriptors. The difference between our model and other IR models is, given a concept as a user's information need, our model maps the concept into a set of index terms which will be used to select relevant documents.

Actually, a conceptual hierarchy is constructed between the concept given by the user, and the set of index terms generated by the system. This hierarchy results from the chaining of the rules given by the expert. This conceptual hierarchy is very similar to the one in RUBRIC; however, there are three major differences between RUBRIC and our model.

The first difference is that in our model, the construction of the conceptual hierarchical is based on expert's knowledge, instead of user's preference. Although a user's preference may better reflect his or her information need, it prevents the general users from accessing the system, as we discussed before. Therefore, we think it is necessary to generate a conceptual structure based on expert's knowledge. As a result, all a general user has to do is to pick the concept which best describes his or her information need, through a friendly user interface. Then, the user can leave all the work to the system. We believe expert's knowledge should provide a more consistent and better retrieval result.

The second difference is that in our model, the conceptual hierarchy is a result of the *forward chaining* of the rules given by the expert, instead of *backward chaining*. In RUBRIC, a concept is placed on the left hand side of a rule; and a concept implied by a sub-concept is placed on the right hand side. Given a concept, we can find out all the sub-concepts which imply this concept by matching it with the right hand side of each rule. The interpretation of these rules is that this concept is defined or described by those sub-concepts. The approach used in our model is to put the concept on the left hand side, and sub-concept on the right hand side. Given a concept, for example, the query concept, all the sub-concepts used to define or describe this concept can be found by matching the concept with the left hand side of each rule. A major benefit of forward chaining is that it not only expands a given query concept into a conceptual hierarchy, but also expands each sub-concept on the leaf nodes into a set of documents.

This leads to the third difference between RUBRIC and our model. In our model, *query* is treated as evidence, instead of *document*. In our opinion, query is the evidence of a user's information need; and conceptual structure is the evidence in deciding which document should be selected. This is different from the approach used by RUBRIC; and this is possible only when the rules and the index mapping have the same direction. The benefit is that a single concept given by the user triggers the process of forward chaining, or evidential reasoning in the sequel; no other control mechanism is necessary.

The query specified by the user can be a compound one. In other words, the connectives *and* and *or* are allowed to describe an information need. The limitation is it should be a simple Boolean query; and the connective *not* is not allowed.

4.2. Evidential Reasoning in Our Model

This section introduces the query evaluation in our model. First, we introduce the theoretical background, which is an extension of D-S theory proposed by Liu [Liu87]; then we explain the interpretation of the meaning of rule uncertainty in our model. A general format of rules is followed to demonstrate the ability of the rules in our model to express the relation between sets of concepts and subconcepts. At last, the evidence aggregation methodology is introduced.

4.2.1. Theoretical Background

The original D-S theory does not support rule uncertainty and chaining of rules. An extension of the D-S theory is proposed by Liu, which solves these problems. In the extended theory, the degree of rule uncertainty is determined through a compatibility relation between the background evidence and rule conclusion. As a result, a "conditional bpa" is constructed to the rule conclusion. The generality of the extended approach actually originates in the generality of the conditional bpa, which may be assigned to an arbitrary condition/conclusion pair [Liu87].

Suppose we have a rule in the form of $A_i \rightarrow C_j$. It is assumed that:

$$A_i = \{a_k \mid a_k \in A_i\} \subseteq \Theta_a$$
$$C_j = \{c_l \mid c_l \in C_j\} \subseteq \Theta_c,$$

where Θ_a is the *antecedent frame* of discernment containing all possible proposition a_k, and Θ_c is the *consequent frame* of discernment containing all possible c_l. In addition, there is a *conditional frame* of discernment $\Theta_{c|a}$ containing all possible *conditional propositions* c_l *given* A_i. The rule $A_i \rightarrow C_j$ is then considered as a subset of $\Theta_{c|a}$:

$$\{c_l \textit{ given } A_i \mid c_l \in C_j\}$$

For each given frame, there is a background frame supporting the evidence of the bpa of the given frame. These background frames are defined as:

$$\Theta'_a = \{a'_m\}$$
$$\Theta'_c = \{c'_n\}$$
$$\Theta'_{c|a} = \{c'_n \textit{ given } A_i\}$$

corresponding to Θ_a, Θ_c, and $\Theta_{c|a}$, respectively.

According to eqn. 3.1, a bpa of the antecedent A_i measures the "reason to believe" A_i by compatible evidence in the background:

$$m(A_i) = \sum_{a'_m :\rightarrow A_i} P(a'_m), \tag{4.1}$$

where $A_i \subseteq \Theta_a$, $a'_m \in \Theta'_a$. In analogy, a *conditional bpa* measures the "reason to believe" the conditional proposition C_j *given* A_i which corresponds to a rule $A_i \rightarrow C_j$:

$$m(C_j \mid A_i) = \sum_{c'_n :\rightarrow C_j} P(c'_n \mid A_i), \tag{4.2}$$

where $C_j \subseteq \Theta_c$, $A_i \subseteq \Theta_a$, and $c'_n \in \Theta'_c$, which is compatible with C_j.

The chaining operation is supported by the extended theory. Suppose the relation between A_i and C_j is defined by some intermediate nodes C_k, that is, we have rules:

$$A_i \rightarrow C_k, \quad \text{and}$$
$$C_k \rightarrow C_j$$

then the following equation describes the chaining operation:

$$m(C_j/A_i) = \sum_{C_k \subseteq \Theta''_c} m(C_j/C_k) \cdot m(C_k/A_i), \tag{4.3}$$

where Θ''_c is a intermediate frame between Θ_c and Θ_a.

4.2.2. Interpretation of Rule Uncertainty

Rule uncertainty in our model is an experience-based belief measure; in other words, the conditional mass assignment in Liu's term. Given a rule between a concept c_i and a sub-concept c_j, the question we ask the expert is: What is the degree of belief that sub-concept c_j will be selected to define concept c_i? or, What is the degree of belief that sub-concept c_j has been used by various authors in describing concept c_i? Here, the terms *degree of belief* and *conditional mass assignment* are used interchangeably; and the *degree of belief* is different from the *belief measure* of a focal element, which is equal to the degree of belief *necessarily* committed to that focal element.

Since a given rule summaries an expert's knowledge and experience, the rule uncertainty is interpreted accordingly. As shown in Fig. 6, given a concept and a sub-concept, c_i and c_j, the first question we ask the expert is for him or her to come up with a list of pieces of knowledge, c_{j1}, c_{j2}, . . ., c_{jn}, which are used to describe the concept c_i, and are semantically compatible with the sub-concept c_j. These pieces of knowledge can be classified into many different categories, for example, definition of concept c_i, knowledge in the text books, conversation with other experts, graphic analogy, the expert's subjective opinion, and so on. As we can see, there is almost no limitation to how an expert can describe the concept he or she has in mind. However, the detail of the process of getting these pieces of knowledge is a problem of knowledge acquisition and will not be discussed here.

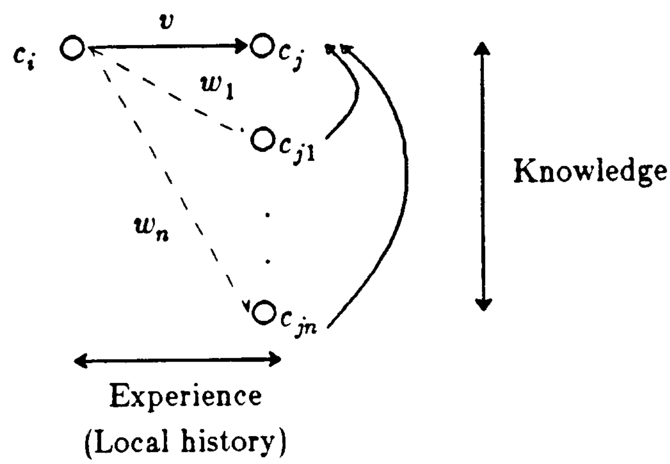

Fig. 6: Rule uncertainty

The second question we ask the expert is to assign a weight to each piece of knowledge, according to his or her experience. Weight w_k is assigned to knowledge c_{jk}, to reflect it's relevant importance. The weight w_j is a number in the range [0, 1]; and $w_1 + w_2 \ldots + w_n = 1$. Also, the weight w_j approximates the conditional probability $P(c_{jk}/c_i)$. This is treated as local history, which is independent from the other pieces of knowledge. Given the weights, the overall degree of belief v of this rule is given by the following equation:

$$v = m(c_j/c_i) = \sum_k w_k \tag{4.4}$$

Given the interpretation of rule uncertainty as conditional mass assignment, the equation of chaining of rules naturally followed according to eqn. 4.3. That is, if we want to know the degree of belief that concept c_i is described by sub-concept c_j, and there are some intermediate sub-concepts c_k between c_i and c_j, then the measure between c_i and c_j is equal to the summation of the pairwise multiplication of the conditional mass $m(c_j/c_k)$ and $m(c_k/c_i)$:

$$m(c_j/c_i) = \sum_k m(c_j/c_k)m(c_k/c_i) \tag{4.5}$$

4.2.3. General Rule

The incompleteness of the knowledge reflected in the rule between the set of concepts and the set of sub-concepts. This incompleteness necessitates the general formulation of the rule. The general format of our rule is:

$$C_i \rightarrow C_j,$$

where C_i is a set of concepts; and C_j is a set of sub-concepts; and the equation for degree of belief measure and chaining becomes:

$$v = m(C_j/C_i) = \sum_k w_k \tag{4.4'}$$

$$m(C_j/C_i) = \sum_k m(C_j/C_k)m(C_k/C_i) \tag{4.5'}$$

With the generalized rules, it is very easy to express the incompleteness of knowledge. That is, the commitment of belief to one statement does not necessarily mean the commitment of the rest of belief to the negation of it. Part of the belief can be reserved to "don't know". For example, in the following three rules associated with concept c_i:

$$c_i \rightarrow c_j, (v_1)$$
$$c_i \rightarrow c_k, (v_2)$$
$$c_i \rightarrow \{c_j, c_k\}, (1-v_1-v_k)$$

the third rule expresses the partial ignorance reserved by the expert in relation to the set of these two sub-concepts. The concept of ignorance was introduced in D-S theory; and the chaining of general rules is made possible by Liu's equation. In the next section, we will introduce the discounting and combining operations of rules in our model.

4.2.4. Evidence Aggregation

In addition to the rules given by the expert in generating the conceptual hierarchy, there are two types of rules. The first one is the indexing rule, which maps an index term into a set of documents. The second one is the default rule, which maps a sub-concept at the leaf node into a textual term, which is the sub-concept itself; the textual term is actually the index term. Unlike the rules described in the previous section, these last two rules are unconditional; or, conditional having the conditional mass assignment equal to one. Each indexing rule represents the knowledge of indexers, who are one type of expert. Each default rule specifies where sub-concepts become textual terms. The default rule is the equivalent of the strong evidential reasoning in RUBRIC.

Now the reasoning path is complete in our model. Starting from a query, which is a concept chosen by a user, through the conceptual hierarchy, then from the leaf nodes to textual terms and from the index terms to the sets of documents, the whole path is linked by rules and the rule uncertainty has a uniform interpretation: conditional mass assignment; therefore, the whole process can be treated as a process of evidential reasoning.

4.2.4.1. Overview

RSV is calculated in the document space, according to the information in the conceptual hierarchy. As shown in Fig. 7, to the left of the dash line is the query, which shows the conceptual hierarchy; to the right of the dash line is the image of this hierarchy which is defined in terms of set of documents; that is, the node of a sub-concept is replaced by a set of documents. In this figure, document sub-space D_i corresponds to the sub-concept c_i. All the uncertainty measures on the links of the conceptual hierarchy are copied to the document space hierarchy.

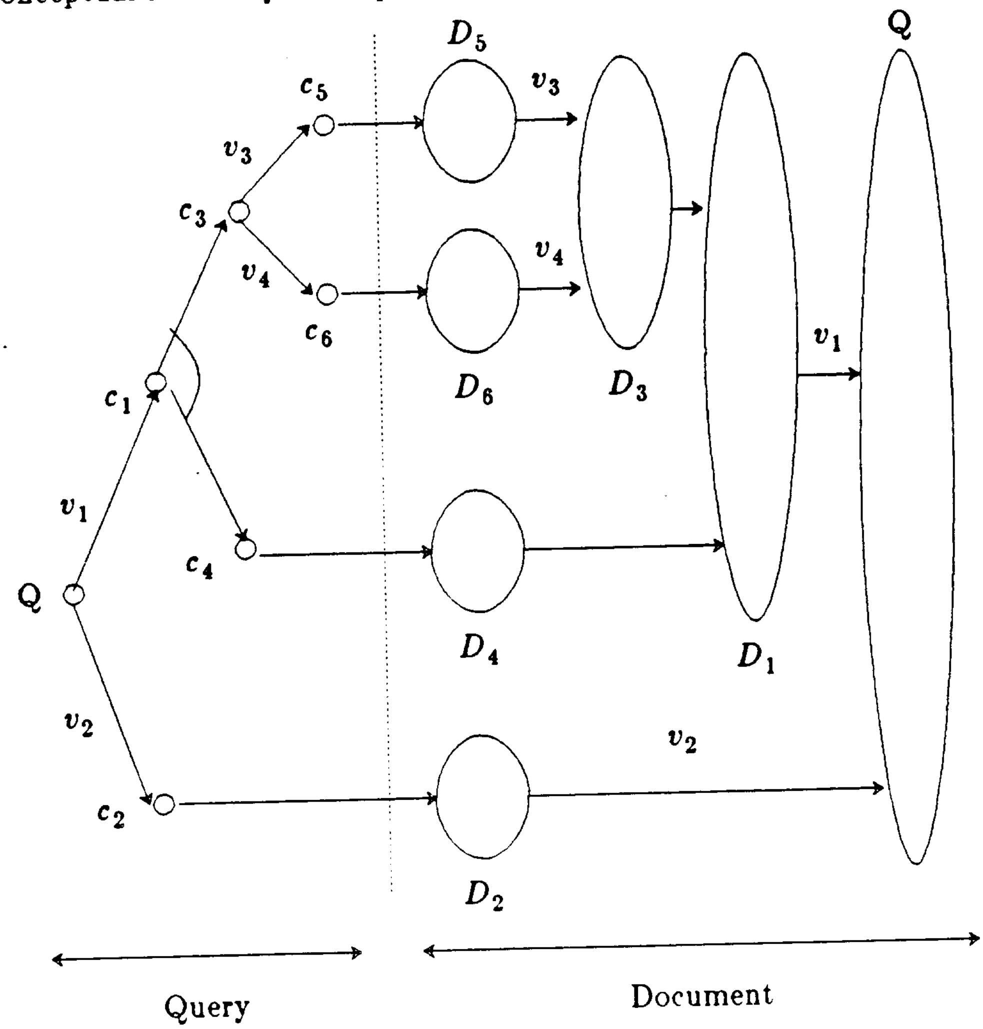

Fig. 7: Evidence Aggregation

The construction of the document space hierarchy starts from the leaf nodes of the conceptual hierarchy. Each leaf node is associated with a set of documents which is indexed by the index term, or sub-concept, at that leaf node. The initial value of the mass assignment for the set of documents at leaf nodes is 1.0; and no discounting is necessary since the leaf nodes are governed by indexing rules. In this example, set D_2, D_4, D_5, and D_6 all have a mass value of 1.0 as the initial value.

After setting the initial value, the mass assignment is propagated to the root of the hierarchy, which is a set of all the documents that will be retrieved. The RSV of each document is decided by the *plausibility* measure of each document, which is the result of applying D-S theory over the document space at root. The way mass distribution is propagated is described in the following section.

4.2.4.2. Basic Operations

The basic operations of propagating mass distribution involve combining the mass value of the sets at the left hand side, or the lower level of the hierarchy, into the mass value of the set at the right hand side, or the higher level of the hierarchy. The combination is controlled by the corresponding node in the query structure, to be more specific, the operation between the nodes which are corresponding to the sets at the left hand side. For simplicity, let's call the operation OP. Usually, there will be more than two sets to be combined. Since the overall combined value is not affected by the order in which they are combined, we treat the combination operation as a binary operation.

In Fig. 8, the part of document hierarchy on the right corresponds to the conceptual hierarchy on the left. The dashed links between the leaf nodes represent this relationship. In this case, concept c_k is described by two sub-concepts c_i and c_j. The rule uncertainty is measured by v_i and v_j respectively. The set of documents associated with the sub-concept c_i is represented by D_i, the same with D_j and D_k. The uncertainty measure between these sets is copied from the conceptual structure on the left. It should be noted that the set D_i and D_j may not be disjoined ; they are shown in separate ovals for the purpose of clarity. Now let us take a look at the way to combine D_i, D_j, and the uncertainty measure of them, into D_k and the uncertainty measure of it.

If OP is an *or* operation, then D_k is equal to the union of D_i and D_j. For any focal element X which has a non-zero mass measure in either D_i or D_j, the derived mass measure of X in the space D_k is equal to the summation of the mass measure of X over space D_i, $m_i(X)$, and the mass measure of X over space D_j, $m_j(X)$, discounted by the uncertainty on the links, v_i and v_j respectively. These are described by the following equations:

$$D_k = D_i \cup D_j \tag{4.6}$$

$$m_k(X) = v_i \cdot m_i\,(X) + v_j \cdot m_j\,(X), \tag{4.7}$$
$$X \subset D_i \cup D_j,\ m_i\,(X) \cdot m_j\,(X) \neq 0$$

If X is contained in D_i only, then eqn. 4.7 is reduced to

$$m_k(X) = v_i \cdot m_i\,(X),$$

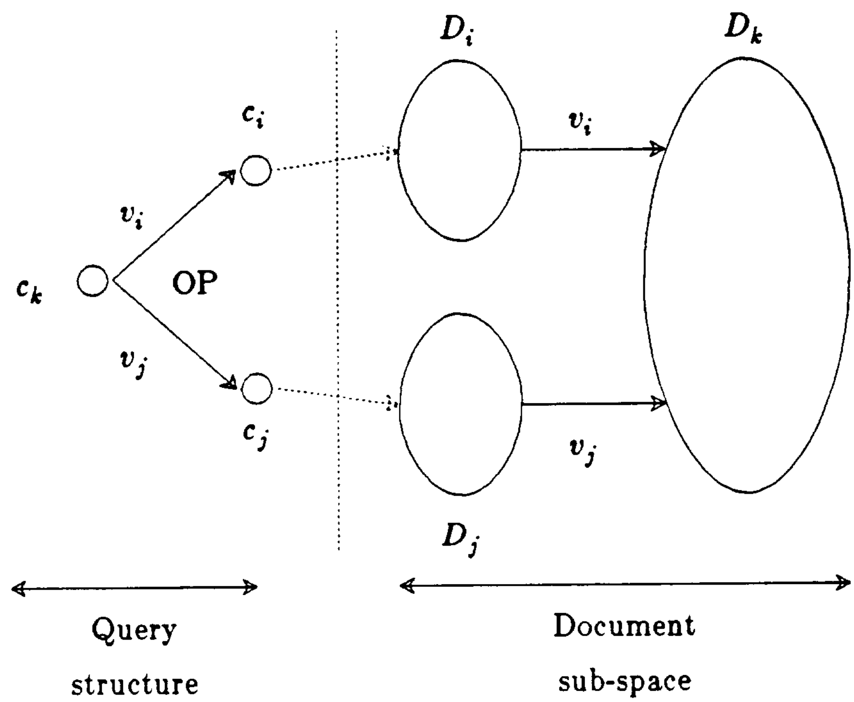

Fig. 8: Basic operation

since the measure of X in D_j is zero. The same arguments hold for the set D_j. If c_i and c_j are leaf nodes, then the set D_i itself is the only subset of D_i that has a none-zero measure; actually, the measure of D_i is assigned to the initial value one. This is also true for D_j. Therefore, the eqn. 4.7 is reduced to

$$m_k(D_i) = v_i,$$
$$m_k(D_j) = v_j.$$

If the OP is an *and* operation, then the intersection of D_i and D_j is assigned to D_k; and the derived measure of D_k is equal to the result of combining the measure of D_i and that of D_j, according to Dempster's rule of combining.

$$D_k = D_i \cap D_j \tag{4.8}$$

$$m_k(X) = (m_i \oplus m_j)(X) \tag{4.9}$$

If the two measures are conflict, in other words, D_k is equal to an empty set, then the measure of X is reset to zero.

In the following two sections, we will discuss the issues of ball-box analogy and combinability of D-S theory, both of which are related to our work.

4.3. Ball-Box Analogy

A ball-box analogy is proposed by Zadeh in explaining D-S theory [Zad85]. We will first describe Zadeh's model, then take a look at the analogy of our model.

4.3.1. Zadeh's Model

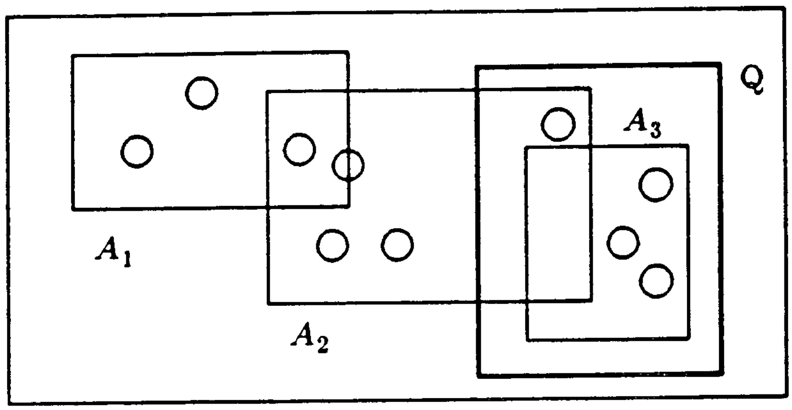

Fig. 9: Ball-box analogy

In Fig. 9, we have several boxes and a set of metal balls. The boxes overlap ; and balls are placed in each box, according to available information. Let's use P_i to represent the fraction of balls put in box A_i. Once a ball is put in a box, the area the ball can travel is limited by the boundary of the box; however, the boundary of the overlap area is penetrable. In other words, if box A_i and box A_j overlap, then the balls put in A_i and A_j are free to travel to the intersection of these two boxes. In our example, $P_1 = 0.2$, $P_2 = 0.5$, $P_3 = 0.3$. These measures are equivalent to the probability measure of the evidence space; and the boxes are equivalent to the focal elements of the hypothesis space. Given the overlap of the boxes, and the measure P_i, we want to know: Given an area, say Q, how many balls are there in the Q?

Because a ball can move around within a box in which it is put, we are not able to come up with a single number in answering the question. Instead, we can only have a range measuring the upper bound and lower bound of the probability of the number of balls in Q. The analogy of the upper bound $Pls(Q)$ and the lower bound $Bel(Q)$ is as follows:

In calculating the *Pls* measure, Q acts as an attractor (e.g. magnet); and all the balls will move toward this area, unless they are constrained by the boundary of a box. The value of the fraction of balls which move to area Q is $Pls(Q)$. On the other hand, in calculating the *Bel* measure, Q acts as a repeller; or, all the other boxes act as attractors. Then, the balls will move away from Q, unless they are constrained by a box within Q. In this case, the fraction of balls still kept in Q is $Bel(Q)$.

4.3.2. Our Model

Although D-S theory is used for our model, the ball-box analogy for our model is a little bit different form Zadeh's model. In our model, besides the information of the overlap of boxes and the number of balls put in each box, we also know the exact location of each box. The whole area in which we can put boxes is divided into small boxes of equal size, as shown in Fig. 10. Each of the small boxes is labeled and represent a document. The location of a given box, regardless of size, is described by its constituent, that is, the small boxes it covers. The question we ask is: Given a magnet box Q, which is the same size as a small box, what is the best location to place Q, such that it can attract most of the balls.

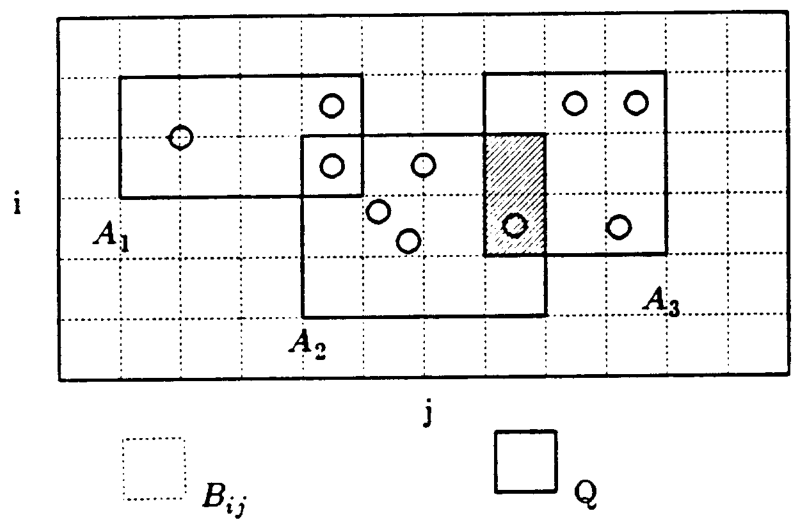

Fig. 10: Ball-box analogy of our model

The answer to this question is provided by calculating which small box has the maximum *Pls* measure, if we put Q there. That is,

$$\max_{i,j} [Pls(B_{ij})] \tag{4.10}$$

In this example, suppose we have the same fraction measure, $P_1 = 0.2$, $P_2 = 0.5$, $P_3 = 0.3$, then the shaded area, which contains two small boxes, has the maximum value of 0.8. The documents represented by this area will be retrieved as the most relevant ones.

4.4. Combinability

4.4.1. Zadeh's Model

One of the problems of the D-S theory is the problem of combinability [Zad85]. If we have two independent evident spaces, which map to the same hypothesis space, the mass distribution of the two evidence spaces can be combined according to Dempster's rule of combination. However, in Zadeh's conjecture, if there is one focal element in one of the mass distributions, which is disjoined from all the focal elements of the other mass distribution, then these two distributions are not combinable. Otherwise, the result of the combining of these distribution will be misleading .

4.4.2. Yen's Model

As pointed out by another researcher Yen, the situation described in Zadeh's conjecture is still combinable, as long as these two distributions are partial information bearing on different attributes [Yen86]. For example, in Fig. 11, we have a relation of students; and there are three different attributes in the relation: sex, age, and address. This relation is also called the parent relation of the distribution information. There are two sources of partial information. The distribution of the number of female students and the distribution of the students who live in California. Suppose we are interested in knowing the distribution of the number of students according to age. Although we have a subset of age, {18, 19}, in the first distribution, which is disjoined from all the subsets of the second distribution, these two distributions are combinable, according to Yen's model. This is because these two distribution are partial information based on different attributes, sex and address, instead of the whole information bearing on the parent relation.

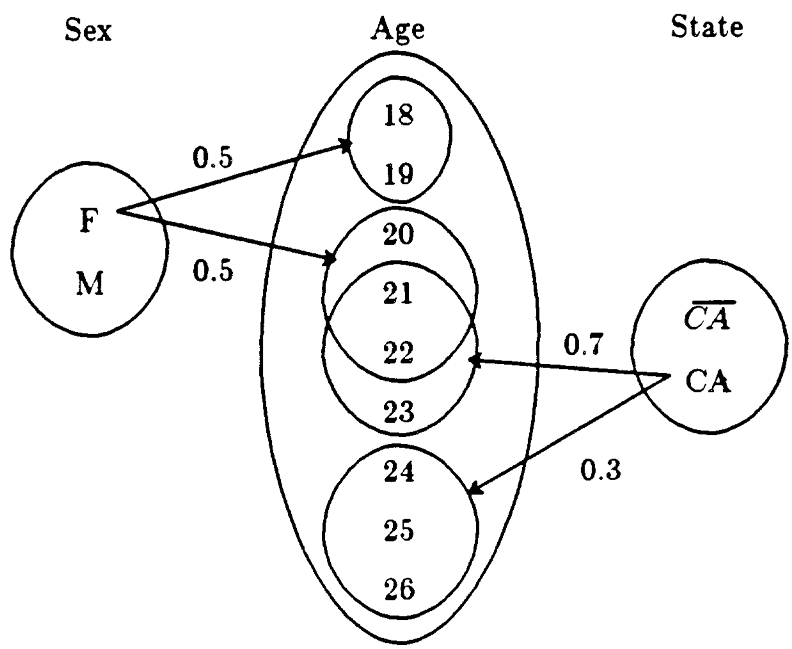

Fig. 11: Combinability

There are two observations in the process of combining these two pieces of information. First, since the set of age {18, 19} is disjoined from all the subsets of the distribution based on students living in California, it is natural to assume that the group of students in this relation who are either 18 or 19 do not live in California. A similar argument holds for male students. Second, after combining, the parent relation is a subset of the original one, which contains female students living in California.

4.4.3. Our Model

In our model, the pieces of evidence we want to combine are based on different aspects, which is similar to the attribute in Yen's model. Therefore, there are pieces of partial information based on different aspects, and two pieces of partial information can always be combined. For example, in Fig. 12, we have a query which tells us the user is interested in knowing articles related to AI and Cognitive Science (Cogsci). According to the result of query formulation, a set of documents, $\{d_1, d_2\}$, is related to one sub-concept AI, to the degree 0.5, and another set of documents, $\{d_7, d_8, d_9\}$, is related to sub-concept Cogsci, to the degree 0.3, and so on. After the combining, we have a new parent relation which contains the documents related to both AI and Cogsci.

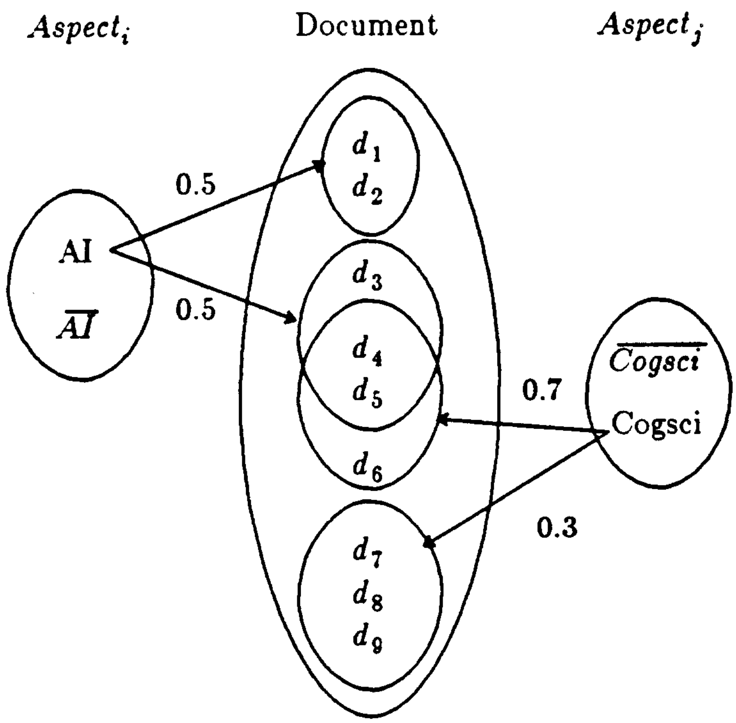

Fig. 12: Combinability of our model

4.5. Correspondence to Boolean System

Let us assume the derived query structure is given by a user of a Boolean system. In the Boolean system, there is no way to express the relative importance of query terms. Our model subsumes the Boolean model in the following way. First, we assume that each link attached to a concept shares the same degree of belief. For example, in Fig. 13, given n sub-concepts associated with a given concept c_i, we assume each sub-concept receives equivalent belief to the degree $1/n$. This process is repeated for every rule; and all the measures are combined as described in our model. All the documents in the final document space which have non-zero RSV will be retrieved. This result is the same as the result produced by a Boolean system.

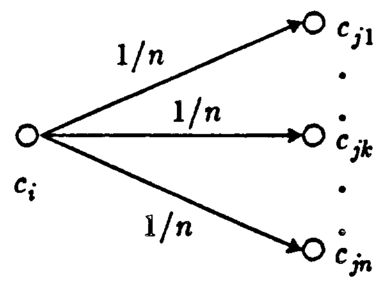

Fig. 13: Reduce to Boolean model

5. Future Work

In this proposal, all the sets are crisp sets. A natural generalization is to incorporate fuzzy sets in our model in the future research. In other words, we would have a rule between a fuzzy set of concepts and a fuzzy set of sub-concepts. More importantly, we would like to adopt fuzzy indexing in our model. That is, the knowledge provided by an indexer is to map a index term into a fuzzy set of documents. We consider this an important step in combining our model with the fuzzy set model of IR.

Another direction of future research is to support more general relation between concepts. That is, there are not only rules between concept and sub-concept, but also rules between related concepts. The difference between these two types of rules needs to be clarified. Also, the measure of the relation between the related concepts and how these two combine these two types of measure need to be studied in the future.

In the process of knowledge acquisition, an important step is to come up with a list of sub-concepts given a concept. Sometimes, it is difficult for an expert to enumerate all the candidate sub-concepts. A heuristic concept structuralization methodology is proposed by Miyamoto et al., which generates thesaurus like structure based on a fuzzy set IR model [MMN83]. The feasibility of this approach will be studied in the future.

We plan to implement a prototype system of our model in the future. The implementation will be done in the environment of a system called Gister * , which is an evidential reasoning system developed at SRI International. Gister provides many graphic tools to support the operations needed for evidential reasoning. However, it does not support the conditional bpa measure. There are two options in supporting this measure in our model. The first choice is to modify the source code of Gister; and the second choice is to use the original Gister, but make explicit the background evidential reasoning associated with each conditional bpa measure. The domain we will make experiment on is the AIList, which is a common place that many AI researchers share their knowledge. The AIList is a message service across the computer network. Each message will be treated as a document; and the topic associated with it will be used in indexing the document.

* Gister is a trademark of SRI International

The standard precision and recall measures will be used in comparing our model with the other models.

6. Conclusion

A general framework of knowledge based IR has been proposed. In this model, IR is treated as a process of evidential reasoning. This process is triggered by the concept selected by a user. Therefore, the query is treated as evidence in selecting the relevant documents for the user. Through out the process, we use a consistent interpretation of the rule uncertainty; and the uncertainty measures are discounted, chained, and combined according to D-S theory and it's extension. No other control knowledge is necessary in aggregating the pieces of evidence in our model. Automatic query formulation in our model is based on expert's knowledge. The general format of rule allows an expert to express his knowledge between sets of concepts and sets of sub-concepts. In other words, partial ignorance of knowledge is properly expressed in our model. In the process of query evaluation, the RSV of a document is calculated by aggregating all the uncertainty measure of the subsets of documents in the document space. A list of documents, ranked according to the RSV, is provided to the user as a response to his or her information request. Also, in this report, we discussed the issues of ball-box analogy and combinability of D-S theory, which are related to our model. We also demonstrated that our model subsumes the Boolean model.

7. References

[Bar85] M. BARTSCHI, An Overview of Information Retrieval Subjects, *IEEE Computers 18(5)* (1985), 67-84.

[BBS87] G. BISWAS, J. C. BEZDEK, V. SUBRAMANIAN and M. MARQUES, Knowledge-Assisted Document Retrieval: II. The Retrieval Process, *Journal of the American Society for Information Science 38(2)* (1987), 97-110.

[BIJ87] P. P. BONISSONE, G. B. P. III and A. L. B. JR, Reasoning with Incomplete and Uncertain Information, *Final Report*, 1987.

[BoS75] A. BOOKSTEIN and D. SWANSON, A Decision Theoretic Foundation for Indexing, *Journal of the American Society for Information Science 26(1)* (1975), 45-50.

[Boo80] A. BOOKSTEIN, Fuzzy Requests, *Journal of the American Society for Information Science 31(3)* (1980), 240-247.

[Boo83] A. BOOKSTEIN, Information Retrieval: A Sequential Learning Process, *Journal of the American Society for Information Science 34(5)* (1983), 331-342.

[Boo85] A. BOOKSTEIN, Probability and Fuzzy Set Application to Information Retrieval, *Annual Review of Information Science and Technology 20* (1985), 117-151, Knowledge Industry Publications, Inc..

[BuK81a] D. A. BUELL and D. H. KRAFT, A Model for a Weighted Retrieval System, *Journal of the American Society for Information Science 32(3)* (1981), 211-216.

[BuK81b] D. A. BUELL and D. H. KRAFT, Threshold Values and Boolean Retrieval Systems, *Information Processing and Management 17(3)* (1981), 127-136.

[Bue81] D. A. BUELL, A General Model of Query Processing in Information Retrieval Systems, *Information Processing and Management 5* (1981), 249-262.

[Coo76] W. S. COOPER, The Suboptimality of Retrieval Ranking Based on Probability of Usefulness, *Technical Report*, 1976.

[CoM78] W. S. COOPER and M. E. MARON, Foundations of Probabilistic and Utility-Theoretic Indexing, *Journal of the Association for Computing Machinery 25(1)* (1978), 67-80.

[Coo83] W. S. COOPER, Exploiting the Maximum Entropy Principle to Increase Retrieval Effectiveness, *Journal of the American Society for Information Science 34* (1983), 31-39.

[Coo88] W. S. COOPER, Getting Beyond Boole, *Information Processing & Management 24(3)* (1988), Pergamon Press.

[Dem67] A. P. DEMPSTER, Upper and Lower Probabilities Induced by Multivalued Mapping, *Annals of Mathematical Statistics 38* (1967), 325-339.

[GoS84] J. GORDON and E. H. SHORTLIFFE, Hypothesis Space, *Tech. Report HPP 84-35* (1984), Stanford University.

[Har75] S. P. HARTER, A Probabilistic Approach to Automatic Keyword Indexing. Part 1. On the Distribution of Speciality Words in a Technical Literature, *Journal of the American Society for Information Science 26(4)* (July 1975), 197-206.

[Kwo87] K. L. KWOK, Some Considerations for Approximate Optimal Queries, *Proceedings of the Tenth Annual International ACMSIGIR Conference on Research & Development in Information Retrieval*, 1987, 19-24.

[Liu87] G. S. LIU, Causal and Plausible Reasoning in Expert Systems, *Ph.D. Thesis*, Computer Science Division. University of California, Berkeley, 1987.

[LoG82] J. D. LOWRANCE and T. D. GARVEY, Evidential Reasoning: a Developing Concept, *Proceeding of the International Conference on Cybernetics and Society*, October 1982, 6-9.

[MaK60] M. E. MARON and J. L. KUHNS, On Relevance, Probabilistic Indexing and Information Retrieval, *Journal of the Association for Computing Machinery 7(3)* (1960), 216-244.

[McC87] P. L. MCCARTY, The Management of Uncertainty in Expert Systems, *Ph.D. Thesis*, Computer Science Division. University of California, Berkeley, 1987.

[MMN83] S. MIYAMOTO, T. MIYAKE and K. NAKAYAMA, Structure Generation on Bibliographic Databases with Citations Based on a Fuzzy Set Model, *IFAC Fuzzy Information*, France, 1983, 225-230.

[Neg73] C. V. NEGOITA, On the Notion of Relevance in Information Retrieval, *Kybernetes 2* (1973), 161-165.

[Rad76] T. RADECKI, Mathematical Model of Information Retrieval System Based on the Concept of Fuzzy Thesaurus, *Information Processing and Management 12* (1976), 313-318.

[Rad77] T. RADECKI, Mathematical Model of Time-Effective Information Retrieval System Based on the Theory of Fuzzy Sets, *Information Processing and Management 13* (1977), 109-116.

[Rad79] T. RADECKI, Fuzzy-Set Theoretical Approach to Document Retrieval, *Information Processing and Management 15(5)* (1979), 247-259.

[Rij79] C. J. V. RIJSBERGEN, *Information Retrieval*, Butterworths, London, 1979.

[RoS76] S. E. ROBERTSON and K. SPARCK-JONES, Relevance Weighting of Search Terms, *Journal of the American Society for Information Science 27(3)* (1976), 129-146.

[Rob77] S. E. ROBERTSON, The Probability Ranking Principle in IR, *Journal of Documentation 33* (1977), 294-304.

[RMC82] S. E. ROBERTSON, M. E. MARON and W. S. COOPER, Probability of Relevance: A Unification of Two Competing Models for Document Retrieval, *Information Technology: Research and Development 1(1)* (1982), 1-21.

[Sal71] G. SALTON, *The SMART Retrieval System -- Experiments in Automatic Document Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[SaM83] G. SALTON and M. J. MCGILL, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.

[Sch87] P. SCHAUBLE, Thesaurus Based Concept Spaces, *Proceedings of the Tenth Annual International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1987, 254-262.

[Sha76] G. SHAFER, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[Tah76] V. TAHANI, A Fuzzy Model of Document Retrieval Systems, *Information Processing and Management 12(3)* (1976), 177-188.

[TSD83] R. M. TONG, D. G. SHAPIRO, J. S. DEAN and B. P. MCCUNE, A Comparison of Uncertainty Calculi in an Expert System for Information Retrieval, *International Joint Conference of Artificial Intelligence*, 1983, 194-197.

[TAC85] R. M. TONG, V. N. ASKMAN, J. F. CUNNINGHAM and C. J. TOLLANDER, RUBRIC: An Environment for Full Text Information Retrieval, *on Research & Development in Information Retrieval*, 1985, 243-251.

[ToS85] R. M. TONG and D. G. SHAPIRO, Experimental Investigations of Uncertainty in a Rule-Based System for Information Retrieval, *International Journal of Man-Machine Studies 22* (1985), 265-282.

[Ton87] R. M. TONG, Conceptual Information Retrieval Using RUBRIC, *Proceedings of the Tenth Annual International ACMSIGIR Conference on Research & Development in Information Retrieval*, 1987, 247-253.

[WoY87] S. K. M. WONG and Y. Y. YAO, A Statistical Similarity Measure, *Proceedings of the Tenth Annual International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1987, 3-12.

[Yen86] J. YEN, Evidential Reasoning in Expert Systems, *Ph.D. Thesis*, Computer Science Division. University of California, Berkeley, 1986.

[YL76] C. T. YU, W. S. LUK and T. Y. CHEUNG, A Statistical Model for Relevance Feedback in Information Retrieval, *Journal of the ACM 23* (1976), 273-286.

[Zad65] L. A. ZADEH, Fuzzy Sets, *Information and Control 8* (1965), 338-353.

[Zad81] L. A. ZADEH, Possibility Theory and Soft Data Analysis, *Mathematical Frontiers of the Social and Policy Sciences*, 1981, 69-129.

[Zad85] L. A. ZADEH, A Simple View of the Dempster-Shafer Theory of Evidence and Its Implication for the Rule of Combination, *AI Magazine 7(2)* (1985), 85-90, AAAI.