10 LOGICAL IMAGING AND PROBABILISTIC INFORMATION RETRIEVAL

Fabio Crestani

Department of Computing Science
University of Glasgow
Glasgow G12 8QQ
Scotland
fabio@dcs.gla.ac.uk

10.1 INTRODUCTION

In Information Retrieval (IR), probabilistic modelling relates to the use of a retrieval model that ranks documents in decreasing order of their estimated probability of relevance to a user's information need expressed by a query. In an IR system based on a probabilistic model, the user is always guided to examine first the documents which are the most likely to be relevant to his or her need.

The use of a probabilistic model in IR assures us that we can obtain "optimal retrieval performance" once we rank documents according to their probability of relevance with regard to a query (Robertson, 1977). However, this principle, called *The Probability Ranking Principle*, refers only to "optimal retrieval", which is different from "perfect retrieval". Optimal retrieval can be defined precisely for probabilistic IR because optimality has a theoretical basis, which come from a relationship between ranking and the probabilistic interpretation of the measures of performance of IR systems (precision and recall). Perfect retrieval refers to the meeting of an information need. Since IR systems use imprecise representations of documents and information needs, perfect retrieval is not an achievable goal for computer-based systems. Despite that and despite a few criticism (Cooper, 1995), probabilistic models based on the Probability Ranking Principle have been shown to give some of the highest levels of retrieval effectiveness currently obtainable.

One of the major obstacles in designing a probabilistic model for IR is that of finding a method for estimating the probabilities of relevance that is both effective and computationally efficient. Past and present research has made large use of formal probability theory and statistics in order to solve the problems of estimation (Croft and Harper, 1979; Fuhr and Buckley, 1991; Wong and Yao, 1989). In mathematical terms the problem consists of evaluating the probability $P(R \mid q, d)$, that is the probability of relevance given a query q and a document d. The IR system needs to evaluate this probability for every document in the collection, and rank the documents accordingly. Some drawbacks limit its correct application:

1. The concept of relevance (R) must be defined in detail since it is central to the evaluation of $P(R \mid q, d)$. This task has proved to be very difficult and a valid definition of relevance has not been found yet (Seracevic, 1970; Froehlich, 1994).

2. Estimating $P(R \mid q, d)$ has proved very difficult because of the large number of variables involved in the representation of documents in comparison with the small amount of feedback data available about the relevance of documents; a problem sometimes referred to as the "curse of dimensionality" (Efthimiadis, 1996)

The first point is perhaps the most important one. It is not conceivable to estimate the probability of relevance if the concept of relevance is not well defined. The approach followed in this chapter is to consider a *logical definition of relevance*, this corresponds to the so called "system relevance", or relevance as perceived by the IR system (Cooper, 1971).

The chapter is structured as follows. First, in section 10.2, I define the logical definition of relevance. Then, in section 10.3, I describe the space in which I am working: the probabilistic retrieval space. I show the effects in this space of conditionalisation, both in its classical Bayesian form and it a new form called *imaging*. The use of conditionalisation by imaging in IR is described in detail later on in section 10.4. Different forms of imaging are covered. The remaining part of the chapter discusses issues related to the application of imaging in IR. First I show the effects of conditionalisation by imaging on sense resolution (section 10.5), then I present different ways of implementing imaging (section 10.6), and I discuss problems concerning their experimentation and evaluation in the context of IR (section 10.7). The chapter ends with a overview of related work (section 10.8).

10.2 RELEVANCE AS LOGICAL IMPLICATION

Relevance is one of the most important, if not "the fundamental", concept in the theory of IR. The concept arises from the consideration that if a user of an IR system has an information need, then some information stored in some documents in a document collection may be "relevant" to his or her need. In other words, the information to be considered relevant to a user's information need is the information that might help the user satisfy his or her information need. Any information that is not considered relevant to a user's information need, is to be considered "irrelevant" to that same information need. This is a consequence of accepting a dichotomous concept of relevance.

Assuming that information is stored in documents[1], then the goal of the system is to retrieve, in response to a user information need expressed by a query, all and only the relevant documents that are known to the system. To do so an IR system should be able to identify what makes a document relevant to an information need. It is the ability to capture the characteristics of relevance that enables an IR system to make the difficult decision about what to retrieve and what not to retrieve. Thus the concept of relevance, whatever that may be, lies at the very heart of IR.

Despite the fact that this concept is so central, a precise definition of what relevance really means is still to come. Many IR researchers have tackled the problem of defining relevance, but a precise definition with real explanatory power is still long awaited (Seracevic, 1970; Froehlich, 1994).

Currently, there are two different views of the notion of relevance in IR; a rough distinction could be made between:

- *topic-appropriateness*;
- *user-utility*.

The first is related to whether or not a piece of information is on a topic which has some topical bearing on the information need expressed by the user in the query. The latter is related to the ultimate usefulness of the piece of information to the user who submitted the query. In current IR research the term relevance seems to be used loosely in both senses, despite the fact that the difference between them is widely recognised.

Topic-appropriateness can be related to a so called "system-perceived relevance", since it does not involve the judgement of the user and is left completely to the IR system. Because of this, topic-appropriateness can be considered "objective" and be studied only from the point of view of the IR system and of the IR process (Cooper, 1971).

The user-utility view of relevance has a much broader sense than topic-appropriateness and it involves a much deeper knowledge of the user information need and of the purpose of this need. We can relate the user-utility sense of relevance to a so called "user-perceived relevance", where the relevance of a document to a query is left completely to the user judgement. The user-utility notion of relevance is therefore a "subjective" notion, since different users may have completely different views about the relevance or non-relevance of particular documents to a given query (Barry, 1994).

In this chapter I only deal with the topic-appropriateness sense of relevance, since it is free of subjective interpretation and can be treated in the sole context of the IR process.

A logical definition of relevance was considered for the first time in the context of IR by Cooper in a paper written more than 25 years ago (Cooper, 1971). For Cooper *logical relevance* was another name for topic-appropriateness, and he addressed the problem of giving a definition of logical relevance IR by analogy with the same problem in question-answering systems. The analogy goes only as far as having questions with a yes-no (true-false) type of answer, and while Cooper's work started by analysing question-answering systems, later he abandoned the analogy. Relevance is defined by

Cooper as "logical consequence". To make this possible both queries and documents need to be represented by sets of declarative sentences. In the case of a yes-no query, the query is represented by two formal statements of the form p and $\neg p$. The two statements representing the query are called "component statements". A subset of the set of stored sentences is called "premiss set" if and only if the component statement is a logical consequence of that subset. A "minimal premiss set" for a component statement is one that is as small as possible in the sense that if any of its members were deleted, the component statement would no longer be a logical consequence of the set. Logical relevance is defined as a two-place relation between stored sentences and the query represented as component statements (the representation of the information need). A first definition of logical relevance says:

> A stored sentence is logically relevant to (a representation of) an information need if and only if it is a member of some minimal premiss set of stored sentences for some component statement of that need.

This definition of relevance is essentially just a proof-theoretic notion that has been generalised to be applicable to information needs involving more than one component statement.

Although logical relevance was initially defined only for sentences, it can be easily extended to apply to stored documents. A document is relevant to an information need if and only if it contains at least one sentence which is relevant to that need.

In the same paper Cooper attempted to tackle a generalisation of such a definition to natural language queries and documents. However, without a formalised language, no precise definition of the logical consequence relation is at hand, and thus we lose a precise definition of relevance. The problems of ambiguity and vagueness of natural language deny the possibility of extending the previous logical notion of relevance, despite the fact that the general idea of implication in natural language is a reasonably clear one. The definition of relevance, so far as natural language in concerned, is only a definition-in-principle; a conceptual definition, but not yet defined on a mathematical level.

Cooper also tried to tackle the problem of having "degrees of relevance", or as he wrote: "shades of grey instead of black and white"(Cooper, 1971), pp.30. The idea was to extend the system of deductive reasoning used to access logical relevance to a system of plausible reasoning. Cooper argued that plausible or probabilistic inference was not as well defined as deductive inference, even for formalised languages. However, he added that when such tools are formalised enough then this development would become a "sensible and indeed inescapable idea", because it would enable the ranking of documents according to an estimated probability of relevance. What he proposed was to assign a higher probability of relevance to a sentence or a document that has greater probability of belonging to a residual minimal premiss set.

Cooper went on extending the previous definition of relevance to the case of non-inferential systems and to the case of topical queries, but the extensions are not of interest in the context of this chapter. The important thing about Cooper's work is in the fact that he was to first to associate the topic-appropriateness sense of relevance with logical implication and that he recognised the importance of evaluating the probability of such implication to rank documents in relation to their estimated probability of relevance.

Many other researchers followed this idea proposing the use of different logics to capture relevance. One thing that was soon recognised is that the logical implication needed to capture relevance was not the classical material implication.

The reasons why the use of the *classical material implication* $d \supset q$ is not appropriate for IR is in the definition of material implication itself. There are many ways of explaining why material implication is not suitable for IR, the arguments that I report are those addressed in (Chiaramella and Chevallet, 1992).

For reasons of brevity, I repeat only the most important argument, that is the one most often used to dismiss the suitability of classical material implication for IR. The argument relates to the fact that the truth of the material implication $d \supset q$ is to be determined relative to a particular evaluation situation. To determine the truth of $d \supset q$ we have to compare the truth of d with that of q. Using a truth table for $d \supset q$ one can see that when d is false, no matter what the query q is, $d \supset q$ will always be true. Herein lies the problem. In fact, d is true only when d is retrieved, but, given a retrieval situation in which q is submitted, a document d is always false since it has not been retrieved yet. Therefore the real retrieval situation corresponds to the case of d false and such a document is relevant to any query q. This obviously does not provide a suitable definition of relevance.

The idea that a *non-classical form of logical implication* was needed for defining relevance was first proposed in (van Rijsbergen, 1986b). That initial idea was supported with stronger and stronger arguments in (van Rijsbergen, 1986a; van Rijsbergen, 1988; van Rijsbergen, 1989). Several other logic-based approaches to relevance have been proposed since then, like for example (Nie, 1989; Wong and Yao, 1991; Nie, 1992; Chiaramella and Chevallet, 1992; Lalmas and van Rijsbergen, 1993; Meghini et al., 1993; Bruza, 1993), just to mention a few. All of them have used a logical definition of relevance based on a non-classical definition of the implication $d \rightarrow q$. Even a glance at these other approaches is outside the scope of this chapter.

In (van Rijsbergen, 1986a) it was proposed to estimate $P(d \rightarrow q)$ using the following *logical uncertainty principle* that is a modified version of the Ramsey test (Sanford, 1989):

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

However, in that paper Van Rijsbergen did not provide an indication on how "uncertainty" and "minimal" might be quantified. Only a few years later he proposed to estimate the probability of the conditional using a non classical technique from conditional logic called *logical imaging* (van Rijsbergen, 1989). Logical imaging is a technique for evaluating the counterfactual conditional $d \rightarrow q$ first proposed in (Lewis, 1986). However, Van Rijsbergen did not define explicitly how imaging could be used operatively in the context of IR.

That proposal was explored a few years later by Crestani and Van Rijsbergen and a model called retrieval by logical imaging was defined in detail in (Crestani and van Rijsbergen, 1995).

In this chapter I explore further the use of the probability of a conditional, namely $P(d \rightarrow q)$, to estimate the conditional probability $P(R \mid q, d)$, by presenting a class of retrieval models based on different forms of imaging. The presentation of these models follow the definition of the space in which I am working.

10.3 THE PROBABILISTIC RETRIEVAL SPACE

Let us assume that the set of possible relevance judgements contains only the two possible events, relevance and irrelevance (non relevance):

$$\mathcal{R} = \{R, \overline{R}\}$$

According to the Probability Ranking Principle, the task of a probabilistic IR system is to rank documents according to their probability of being relevant:

$$P(R \mid \underline{q}, \underline{d})$$

where $\underline{q}$ and $\underline{d}$ are the real query and the real document.

Unfortunately the IR system can only estimate this probability using the available query and documents representations, q and d. The probability of relevance is therefore only estimated:

$$P(R \mid \underline{q}, \underline{d}) \approx P(R \mid q, d)$$

and the accuracy of this estimate depends very much on the quality of the document and query representation. In the IR terminology $P(R \mid q, d)$ is called the *Retrieval Status Value* (RSV) of a document with regards to a query.

In this chapter I do not deal with the problem of finding an effective representation of documents and queries. This is another very difficult problem. Most IR system assume a poor representation of documents and queries based on the use of index terms that are automatically extracted from the text of document and queries. I follow this same way of representing documents and queries and I ignore the problems related to finding more effective representations. However, I assign to this standard way of representing documents and queries a new semantics that enables me to perform an accurate analysis of the events that take place at retrieval time in a probabilistic retrieval space. By studying these events in the light of the new semantics, it is possible to design a new class of probabilistic retrieval models. This new semantics of the probabilistic retrieval space is based on "Possible World Semantics".

10.3.1 Possible world semantics and imaging

Possible World Semantics (PWS) was introduced in the sixties by (Kripke, 1971) in the context of Modal Logic. In his semantics the truth value of a logical sentence is evaluated in the context of a *world*. PWS has been used in modal systems to give a semantics for necessity, where a sentence is true in every possible world, and possibility, where a sentence is true in at least one possible world[2]. According to PWS we have a set of worlds W and an *accessibility relation* ρ defined on $W \times W$. One world is the actual world. A set of worlds around the actual world are accessible from the latter. This construction is what is needed for the evaluation of a counterfactual implication $y \to x$. The evaluation is based on the notion of *minimal extension* which is in accordance with van Rijsbergen's logical uncertainty principle and the Ramsey test.

According to the notion of minimal extension, the truth value of the conditional $y \to x$ in a world w is equivalent to the truth value of the consequent x in the closest world w_y to w where the antecedent y is true (Lewis, 1986). Ties at this stage, if they occur, are broken at random, to ensure uniqueness of the closest world (but see further on for a generalisation). The passage from one world to another world can be regarded as a form of belief revision, and the passage from a world to its closest is therefore equivalent to the least drastic revision of one's beliefs.

More formally, suppose we have a language L with an infinite set of propositional variables $\{a, b, c, ...\}$, two primitive connectives $\wedge$ (conjunction) and $\neg$ (negation), and parentheses. Suppose we have two sentences (well formed formulae) x and y of L. Moreover we have the additional connectives: $\vee$ (disjunction), $\supset$ (material implication), $\to$ (counterfactual implication), and $\equiv$ (material equivalence) defined in terms of the primitives.

We also assume we have a truth evaluation function τ:

$$\tau(y) = \begin{cases} 1 & \text{if } y \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

τ meets the following two conditions:

$$\tau(\neg y) = 1 - \tau(y)$$

$$\tau(x \ \vee \ y) = max[\tau(x), \tau(y)]$$

Suppose now we have a finite set of possible worlds W. We can extend the truth evaluation function τ to indicate the truth value of a sentence in the context of a world:

$$\tau(w, y) = \begin{cases} 1 & \text{if } y \text{ is true at } w \\ 0 & \text{otherwise} \end{cases}$$

$\tau(w, y)$ is just a special case of $\tau(y)$ once we consider $\tau(y) = \tau(W, y)$, for which y is true if it is true in all possible worlds.

Let w_y be the world most similar to w where y is true according to the accessibility relation. The counterfactual implication $y \to x$ will be true at w if and only if x is true at w_y:

$$\tau(w, y \to x) = \tau(w_y, x)$$

The above technique is called *logical imaging*, or simply imaging, and was proposed in (Stalnaker, 1981).

Imaging was extended in (Lewis, 1981) to the case where there is a probability distribution on W. Let us assume this probability distribution follows the classical rules of probability, and in particular:

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

Then we can go from probabilities of worlds to probabilities of sentences by summing the probabilities of the worlds where a sentence is true:

$$P(y) = \sum_{w} P(w)\ \tau(w,y)$$

From this initial probability distribution P that we call "prior" probability, we can derive a new probability distribution P' so that:

$$P'(w') = \sum_{w} P(w)\ \sigma(w',w,y)$$

where:

$$\sigma(w',w,y) = \begin{cases} 1 & \text{if } w' = w_y \\ 0 & \text{otherwise} \end{cases}$$

This process of deriving the new probability distribution P' from the original P is obtained by transferring the probability of every "not-y world" w to its most similar "y-world".

Lewis showed that $P(y \to x) = P'(x)$ or, using a terminology more appropriate to highlight the imaging process on y, that:

$$P(y \to x) = P_y(x)$$

where $P_y(x)$ is the new probability distribution, called "posterior probability", derived from P by imaging on y. In other words, the probability of the conditional is the probability of the consequent after imaging on the antecedent. The proof is reported in (Lewis, 1981). The interested reader can also look at (Gärdenfors, 1982; Cross, 1994) for more details on the logical imaging process.

Gärdenfors in 1988 proposed a generalisation of the imaging process. The generalisation originated from an attempt to overcome one of the restrictive assumptions Lewis made for Stalnaker's semantics of conditionals (Lewis, 1981). The assumption is related to the "uniqueness" of the world w_y, that is the uniqueness of the most similar y-world to w. In (Gärdenfors, 1988), pp. 110, a generalisation of the imaging process that does not rely on this assumption is proposed[3]. The starting point of the generalisation is the use of a (degenerated) probability function P^w to represent the fact that in any possible world w a sentence y is either true or false. Using our formalism this would be equivalent to assigning a probability to the truth function τ:

$$\tau(w,y) = P^w(y)$$

In the case of imaging:

$$P^w(y) = \begin{cases} 1 & \text{if } y \text{ is true at } w \\ 0 & \text{otherwise} \end{cases}$$

Lewis called such a probability function "opinionated" because "it would represent the beliefs of someone who was absolutely certain that the world w was actual and who therefore held a firm opinion about every question" (Lewis, 1981), pp. 145. Gärdenfors generalised imaging by considering the fact that, instead of having $P^w(y) = 1$ only for a single w_y, we can have $P^w(y) > 0$ for a set of worlds:

$$P^w(y) \begin{cases} = 0 & \text{if } y \text{ is false at } w \\ > 0 & \text{otherwise} \end{cases}$$

with the requirement that $\sum_w P^w(y) = 1$. By analogy with the case of imaging we then have:

$$P^w(y \rightarrow x) = \sum_w P^w_y(x)$$

Hence, taking into consideration the prior probability we go from probabilities of worlds to probabilities of sentences as follows:

$$P(y) = \sum_w P(w)\ P^w(y)$$

From this probability distribution we can derive a new probability distribution P'' so that:

$$P''(w') = \sum_w P(w)\ P^w(w')\ \sigma(w', w, y)$$

where:

$$\sigma(w', w, y) = \begin{cases} 1 & \text{if } w' \in W_y \\ 0 & \text{otherwise} \end{cases}$$

with W_y representing the set of the closest worlds to w where y is true.

It could be proved, with a demonstration similar to the one reported in (Lewis, 1981), pp. 142, that this new probability distribution is the posterior probability distribution derived from the prior probability P by *general logical imaging* on y. In other words, this new probability distribution can be obtained by transferring the probability from every world w to the worlds in W_y, the set of most similar (or closest) worlds to w where y is true. The transfer of probability is performed according to the opinionated probability function P^w. It is easy to prove that Lewis' imaging is just a special case of general imaging when $P^w(y) = 1$ for just one w.

The evaluation of $P_y(x)$ either by imaging or general imaging causes a shift of the original probability P from "not-y-worlds" to "y-worlds" to derive a new probability distribution P_y. Since the transfer of probabilities is directed towards the closest y-worlds, this technique is just what it is needed to implement Van Rijsbergen's logical uncertainty principle described in section 10.1. The probability revision is in fact minimal with regard to the accessibility relation. In section 10.4 I explain how we can use this result in the context of IR.

10.3.2 Imaging versus other forms of conditionalisation

How are imaging and other forms of conditionalisation related? If we restrict ourselves to comparing imaging and Bayesian conditionalisation, then the two techniques can be pictured in the following way.

Conditionalisation by imaging causes a revision of the prior probability on the possible worlds W in such a way that the posterior probability is obtained by shifting the original probabilities from not-y-worlds to y-worlds. Each not-y-world moves its probability to its closest y-world (or set of y-worlds in the case of general imaging). Probability is neither created or destroyed, but just moved around according to the accessibility relation (and to the opinionated probability function, in the case of general imaging).

Bayesian conditionalisation is obtained by cutting off all not-y-worlds and then magnifying the probabilities of the y-worlds so that the posterior probabilities still add up to 1. The magnification is done in the same way for every y-world, thus keeping constant the ratios between the probabilities assigned to these worlds.

It is therefore clear that imaging and Bayesian conditionalisation yield, in general, different results. Gärdenfors proved that there is no prior probability distribution (apart from a trivial one) for which imaging and Bayesian conditionalisation give the same revision (Gärdenfors, 1988), pp. 116. A more detailed comparison between Bayesian conditionalisation and imaging from a logical point of view can be found in (Cross, 1994). An analysis of the consequences of these differences in the context of IR is reported in (Crestani and van Rijsbergen, 1995).

10.3.3 The probabilistic term space

Let us now define the IR space in which we are working.

One of the best known IR models is the Vector Space Model (VSM) (Salton, 1968). In this model a document is represented by means of a vector whose elements represent the presence/absence of certain features in that document like, for example, the presence or absence of some index terms. In the binary case, which I consider for simplicity of exposition, the presence of a 1 in a particular position of the vector indicates that the particular feature relative to that position in the vector is present in the document representation, while a 0 indicates its absence. The document representation space is therefore multidimensional, with as many dimensions as the number of features used to represent documents. A document is represented in this space as a vector. A query is also a vector, and document retrieval is performed as a function of the angle between query and document vectors. The semantics of the VSM is therefore that of a multidimensional geometrical space with a topology and a metrics that enable the evaluation of the RSV of documents with regards to a query. Many IR models use this representation space as the underlying space.

In this chapter I use this very representation model but with a different semantics. The semantics of our representation space is based on the PWS. I use the PWS in the context of IR by considering a term as a possible world. According to this view, proposed in (Amati and Kerpedjiev, 1992; Crestani and van Rijsbergen, 1995), a term is represented as a “vector of documents”. Intuitively this can be understood as “if you want to know the meaning of a term then look at all the documents in which that term occurs". This idea is not new in IR (see for example (Qiu and Frei, 1993)) and it has been widely used for the evaluation of term–term similarity.

	t_1	t_2	t_3	$\cdots$	t_n
d_1	1	1	0	$\cdots$	0
d_2	0	0	1	$\cdots$	1
d_3	1	1	1	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
d_k	1	0	0	$\cdots$	1

Figure 10.1 The classical semantics for the term space.

More formally, let us assume our representation space is made of a set of index terms T, the *Term Space*. The set of index terms T is our set of possible worlds. Assume we also have a document collection D, where each document d is represented using terms (worlds) in T, as depicted in figure 10.1. Then according to our semantics, a document can either be true or not true in the context of a world. To determine this truth value it is sufficient to transpose the representation matrix and consider a document true in the context of a term (world) only if it uses that term in its representation. The matrix depicted in figure 10.2 can therefore be interpreted as representing the truth values of documents in the context of possible worlds.

	d_1	d_2	d_3	$\cdots$	d_k
t_1	1	0	1	$\cdots$	1
t_2	1	0	1	$\cdots$	0
t_3	0	1	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
t_n	0	1	1	$\cdots$	1

Figure 10.2 Application of the PWS to the term space.

The above semantics for the term space can easily be extended to the case of a matrix with real values. In particular, if these values are in the $[0, 1]$ range, then they can be considered as probabilities of truth for a document in the context of a term.

To be able to apply imaging in this context, we also have to assume the presence of a prior probability distribution P on T, assigning to each $t \in T$ a probability $P(t)$ so that $\sum_t P(t) = 1$. This probability reflects the importance of a term in the term space. We call this initial probability distribution "prior" because it reflects the importance of terms prior to the submission to the IR system of any query or the selection of any document as relevant to a user's information need. Once some external information enters the term space, mostly in the form of a query or a relevance judgement (but not necessarily only in these forms) then the importance of a term changes to reflect the new information. Accordingly, the probability assigned to a term changes to reflect the increased or decreased importance of the term. However for it to be considered a probabilistic space, the sum of the probabilities assigned to terms must remain constant (that is equal to 1) and so probabilities are moved around in the term space so that if one term increases its importance then some other terms must decrease their importance in a equal measure. These changes occurring in an IR system at retrieval time are very important in order to understand how IR models work. I believe that a study of the kinematics of probability in IR is very important to understand in detail why some models give a better performance than others. This is what I intend to investigate in the rest of the chapter.

10.4 IMAGING AND INFORMATION RETRIEVAL

In this section I propose a class of IR models based on conditionalisation by imaging. Each model has its own specific probability kinematics that I describe.

To make my analysis clearer, in the context of the term space described in section 10.3.3, I take into consideration a particular document and query and evaluate the RSV according to the different models proposed. Suppose we have a document d represented by terms t_1, t_5, and t_6 and a query q represented by t_1, t_4, and t_6. Each one of these terms has a prior probability associated with it, indicated by $P(t)$. In the following I show how the RSV of document d is evaluated in different ways by different forms of imaging and I explain how the probabilities associated to terms change and shift from term to term in different models. I indicate the new "posterior" probability with $P_d(t)$ to highlight the fact that it is obtained by taking into consideration a particular document d.

10.4.1 Standard imaging

(Crestani and van Rijsbergen, 1995) proposed the use of logical imaging (or standard imaging) in IR for estimating the probability of relevance of a document by means of the probability of the conditional $d \to q$:

$$P(R \mid q, d) \approx P(d \to q)$$

Retrieval by logical imaging (RbLI) is the model that estimates $P(R \mid q, d)$ by $P(d \to q)$, where the latter is evaluated using logical imaging. The application of the PWS on the term space enables us to apply imaging to derive the posterior probability $P_d(t)$ by imaging on d over all the possible terms (possible worlds) t in T. Probabilities are transferred according to the kinematics induced by the imaging process. More formally $P(d \to q)$ can be evaluated in the following way:

$$\begin{aligned} P(d \to q) &= P_d(q) \\ &= \textstyle\sum_t P_d(t)\ \tau(t, q) \\ &= \textstyle\sum_t P(t)\ \tau(t_d, q) \end{aligned}$$

where t_d is the closest term to t for which d is true, or in other words, the most similar term to t that also occurs in the document d. The application of imaging to IR requires an appropriate measure of similarity over the term space, the equivalent of the accessibility relation, to enable the identification of t_d.

According to Van Rijsbergen's logical uncertainty principle, RbLI provides a minimal revision of the prior probability in the sense that it involves no gratuitous movement of probability from one world to dissimilar worlds. In fact, the revision of the prior probability necessary to make d certain is obtained by adopting the least drastic change in the probability space. This is achieved by transferring probabilities from each term not occurring in the document d (a non-d-term) to its closest (or most similar) term occurring in it (a d-term), so that the total amount of the distance covered in the transfer is minimal.

Table 10.1 reports an example of the evaluation of $P(d \rightarrow q)$ by imaging on d. A comparison between the imaging and classical IR is reported in (Crestani and van Rijsbergen, 1995).

Table 10.1 Example of the evaluation of $P(d \rightarrow q)$ by imaging on d.

t	$P(t)$	$\tau(t,d)$	t_d	$P_d(t)$	$\tau(t,q)$	$P_d(t) \cdot \tau(t,q)$
1	0.20	1	1	0.30	1	0.30
2	0.10	0	1	0	0	0
3	0.05	0	5	0	0	0
4	0.20	0	5	0	1	0
5	0.30	1	5	0.55	0	0
6	0.15	1	6	0.15	1	0.15
$\sum_t$	1.00			1.00		0.45

10.4.2 General imaging

Retrieval by general logical imaging (RbGLI), first proposed in (Crestani and van Rijsbergen, 1995), is based on a generalisation of RbLI. This generalisation follows a similar generalisation proposed by (Gärdenfors, 1988) in the context of the Belief Revision theory, as reported in section 10.3.1.

RbGLI is the result of the application of the general imaging technique to IR. General imaging is performed on document d to estimate the RSV of a document d with regards to a query q. Again:

$$P(R \mid q, d) \approx P(d \rightarrow q)$$

but this time the evaluation of $P(d \rightarrow q)$ is performed using the following formula:

$$\begin{aligned} P(d \rightarrow q) &= P_d(q) \\ &= \textstyle\sum_t P_d(t)\, \tau(t,q) \\ &= \textstyle\sum_t P(t)\, (\sum_{t'} P^{t'}(d))\, \tau(t_d, q) \\ &= \textstyle\sum_t (\sum_{t'} P_d^{t'}(t)\, P(t'))\, \tau(t_d, q) \\ &= \textstyle\sum_{t,t'} P_d^{t'}(t)\, P(t')\, \tau(t_d, q) \end{aligned}$$

where $P^{t'}(d)$ is the opinionated probability of d in t', and $P_d^{t'}(t)$ is the opinionated probability of term t in t' which is necessary to evaluate the former, and $\tau(t_d, q)$ is as defined in section 10.4.1.

The application of the above technique to IR requires not only an appropriate measure of similarity over the term space T, as in standard imaging, but also an appropriate opinionated probability function. The probability function $P_d^{t'}(t)$ determines the required movements of probabilities from t' to $t \in T_d$, where $T_d \subset T$ is the set of all d-terms. Such a function depends on the particular document on which the general imaging is performed and on the particular term we are transfering the probability from. The number of functions required is then equal to the product of the number of document multiplied by the number of terms. The number of opinionated probability functions could therefore be very high. It would be interesting to use contextual information together with similarity information to determine how close a term is to another term, with the document setting the context in which the similarity is measured. However, in the meantime we have to make some strong assumptions:

1. The opinionated probability function is independent of the document being considered. This is equivalent to saying that the opinionated probability function is context-independent, that is: $P_d^{t'}(t) = P^{t'}(t)$ for every $d \in D$.

2. The opinionated probability function does not use the similarity value but only the similarity ranking. This means that in the evaluation of how much probability needs to be transferred from t' to t we do not consider the value of the similarity between t' and t, but only the position of t in a ranking of all terms according to their similarity with t'.

These two assumptions enable us to use a single opinionated probability function for every term in the term space. In the example reported in table 10.2, I make use of a very simplistic opinionated probability function, one that, given one term, finds the two closest terms, orders them, and transfers two thirds of the probability to the first one and one third to the second one.

Table 10.2 Example of the evaluation of $P(d \rightarrow q)$ by general imaging on d.

t	$P(t)$	$\tau(t, d)$	t_d	$P_d(t)$	$\tau(t, q)$	$P_d(t) \cdot \tau(t, q)$
1	0.20	1	1	0.33	1	0.33
2	0.10	0	1; 6	0	0	0
3	0.05	0	5; 6	0	0	0
4	0.20	0	5; 1	0	1	0
5	0.30	1	5	0.47	0	0
6	0.15	1	6	0.20	1	0.20
$\sum_t$	1.00			1.00		0.53

10.4.3 Proportional imaging

Retrieval by proportional logical imaging (RbPLI) was first proposed in (Sebastiani, 1996) and formalised in more detail in (Crestani et al., 1996b). It is based on a modification of the general imaging technique that is new in the field of Conditional Logic, but that seems more suitable to the area of IR. RbPLI can be obtained by removing the second assumption reported in section 10.4.2, but keeping the first one. This is equivalent to defining an opinionated probability function that is based on the similarity between terms:

$$P_d^{t'}(t) = \frac{Sim(t, t')}{\sum_{t'' \in d} Sim(t, t'')}$$

where $P_d^{t'}(t)$ is the opinionated probability of term t in t', and $Sim(t, t')$ is a similarity function which evaluates the similarity between two terms.

Great care should be taken in choosing the most appropriate function $Sim(t, t'')$ to make sure that $P_d(t)$ remains a probability function. In particular, we have to make sure that $\sum_t P_d(t) = 1$. I do not address this problem here.

Experiments using this form of imaging have not been performed yet, but it does seem that such a variation of general imaging would be appropriate for IR since it uses directly the similarity between the probability donor and the recipient. However, I foresee that the complexity of such a model will make it very difficult to perform extensive testing on large test collections. In (Sebastiani, 1996) ways are suggested to reduce the burden of the necessary calculation. Experiments are needed to see if these artifices invalidate the advantages of using RbPLI instead of RbGLI.

Table 10.3 reports an example of the evaluation of $P(d \to q)$ by proportional imaging on d. The table reporting the similarity values for each pair of terms has been omitted, but its values are used in the probability transfer.

Table 10.3 Example of the evaluation of $P(d \to q)$ by proportional imaging on d.

t	$P(t)$	$\tau(t,d)$	t_d	$P_d(t)$	$\tau(t,q)$	$P_d(t) \cdot \tau(t,q)$
1	0.20	1	1	0.38	1	0.38
2	0.10	0	1; 6; 5	0	0	0
3	0.05	0	5; 6; 1	0	0	0
4	0.20	0	5; 1; 6	0	1	0
5	0.30	1	5	0.41	0	0
6	0.15	1	6	0.21	1	0.21
$\sum_t$	1.00			1.00		0.59

10.4.4 Mixed imaging

Retrieval by mixed logical imaging (RbMLI) was first proposed in (Crestani, 1996).

Mixed imaging consists of a method for the revision of probability functions that combines Bayesian conditionalisation with imaging. It is based on the fact that, when a measure of similarity between some terms is not available, it is not possible to apply in an accurate way any of the models so far presented. In such eventuality, to avoid losing the probability that is not transferred and therefore having a posterior probability that is not a probability (since it does not sum to one), we have to use some artifices. When one does not have similarity information between terms one can still perform Bayesian conditionalisation. The idea that underlies mixed imaging is that the probabilities of non-d-terms are redistributed to all terms in the document only partly by an imaging method, while the rest is redistributed according to Bayesian conditionalisation. Exactly how much of it is transfered by imaging is determined either by the available similarity information or by a parameter that varies according to the "base" variant of imaging (standard, general, proportional); as an example, in the case of standard imaging, this parameter is the degree of similarity between a non-d-term and its most similar d-term.

Table 10.4 shows an example of the evaluation of $P(d \rightarrow q)$ by mixed general imaging on d, where we suppose that we have no similarity information for term t_3 and t_4.

Table 10.4 Example of the evaluation of $P(d \rightarrow q)$ by mixed general imaging on d.

t	$P(t)$	$\tau(t,d)$	t_d	$P_d(t)$	$\tau(t,q)$	$P_d(t) \cdot \tau(t,q)$
1	0.20	1	1	0.36	1	0.36
2	0.10	0	1; 6	0	0	0
3	0.05	0	-	0	0	0
4	0.20	0	-	0	1	0
5	0.30	1	5	0.40	0	0
6	0.15	1	6	0.24	1	0.24
$\sum_t$	1.00			1.00		0.60

10.4.5 Other forms of imaging

A new variant of standard imaging called *retrieval by Jeffrey's logical imaging* (RbJLI) was suggested in (Sebastiani, 1996). This is an extension of imaging that addresses the revision of probability functions on the basis of "uncertain" evidence. *Jeffrey's conditionalisation* was proposed in (Jeffrey, 1965). It allows conditioning to be based on evidence derived from a "passage of experience", where the evidence can be non-propositional in nature. Jeffrey's conditionalisation has many advantages over Bayesian conditioning, in particular, it enables conditioning on uncertain evidence, and allows order-independent partial assertion of evidence. It can be described by the following formula:

$$P^*(H) = P(H \mid E)\, P^*(E) + P(H \mid \neg E)\, P^*(\neg E)$$

where H is a proposition (the hypothesis) and E some uncertain evidence, since in general $P^*(E)$ the subjective certainty of the evidence usually yields a value $P^*(E) \neq 1$. As can be seen Jeffrey's conditionalisation becomes Bayesian once the evidence is certain and $P^*(E) = P(E) = 1$. The application of the formula to IR, as advocated in (van Rijsbergen, 1992), can be obtained by associating q with H and d with E.

The idea that underlies Jeffrey's imaging, that only part of the probability of a non-d-term is redistributed to the d-terms, is similar to what happens in a possible-worlds view of Jeffrey conditionalisation. The probability that non-d-terms retain, has the effect of allowing the revised probability of the document to be nonzero, as desired. RbJLI can therefore be formalised as follows:

$$P_d^*(q) = \sum_t (P_d(t)\; P^*(d) + P_{\neg d}(t)\; (1 - P^*(d)))\; \tau(t,q)$$

where $P_d(q)$ is the posterior probability of q after imaging on d and $P_d(t)$ and $P_{\neg d}(t)$ are respectively the posterior probability of term t after imaging on d and on the negation of d.

RbJLI has not been experimented yet. Undoubtedly it has a certain appeal, but the complexity of its evaluation will make it difficult to use with large test collections.

(Crestani et al., 1996b) also produced some preliminary negative theoretical results from *retrieval by weighted logical imaging* (RbWLI), a would-be extension of imaging methods that it was hoped would account for those cases in which documents and query are true in possible worlds only to a certain degree, that is when $\tau(t,d)$ and $\tau(t,q)$ assume any value in the range $[0,1]$. In IR, this would particularly suit the case in which documents and information needs are represented by vectors of weighted keywords. Unfortunately, we have proven that in the weighted case, unless $\tau(t,d)$ and $\tau(t,q)$ are probability functions, it is impossible to give a revision of the prior probability into a posterior probability which guarantees that $P_d(d) = 1$.

10.5 IMAGING AND WORD SENSES

In (Crestani et al., 1996a) we reported the initial findings of a study of the sense resolution properties of the RbLI model. A few interesting results can be presented. We compared the probability kinematics of the evaluation of $P(d \to q)$ by imaging on d with that of the evaluation of $P(q \to d)$ by imaging on q. The latter can be evaluated in a way very similar to the former, by shifting probabilities from each term not occurring in the query q (a so called non-q-term) to its most similar one occurring in q (a q-term) (Crestani and van Rijsbergen, 1995). This investigation revealed an unexpected effect that imaging has on certain types of ambiguous terms. We discovered that each of the two forms of imaging behaves differently with regard to the senses of ambiguous terms. The effect can be explained more easily using an example.

Let us imagine a document collection in which the term “bat” appears in a number of documents and that the frequency of occurrence of its term (word) senses is skewed. In most documents, the term is used to refer to a sporting implement, but occasionally it is used to refer to the flying mammal. As the sporting sense of “bat” is predominant in this collection, terms most similar to “bat” (we can assume that similarity is measured by co-occurrence) will be those similar to this one sense. For this example, let us say that the terms most similar to “bat” are “cricket”, “baseball”, “hit”, and “ball”. In terms of imaging, it is these five terms that are most likely to transfer their probabilities to each other.

Now let us look at two documents from this collection. Document d_1 is represented by terms “bat” and “hit”, while document d_2 is represented by terms “bat” and “night”. Document d_1 uses the term “bat” in the sporting sense; document d_2 uses it in the animal sense. Suppose a user enters the two term query, “bat”, “cricket”.

Table 10.5 Evaluation of $P(d_2 \to q)$ by imaging on d_2.

t	$P(t)$	$\tau(t, d_2)$	t_{d_2}	$P_{d_2}(t)$	$\tau(t, q)$	$P_{d_2}(t) \cdot \tau(t, q)$
bat	0.20	1	1	0.95	1	0.95
ball	0.10	0	1	0	0	0
night	0.05	1	3	0.05	0	0
cricket	0.20	0	1	0	1	0
hit	0.30	0	1	0	0	0
baseball	0.15	0	1	0	0	0
$\sum_t$	1.00			1.00		0.95

Let us start by considering the process of imaging on a document. Looking at our example, let us first examine d_2. Since the terms “cricket”, “baseball”, “hit”, and “ball” are more similar to “bat” than to “night”, all their probabilities transfer to this one term. From table 10.5 we can see that this transfer results in document d_2 having an estimated probability of relevance of 0.95.

Table 10.6 Evaluation of $P(d_1 \to q)$ by imaging on d_1.

t	$P(t)$	$\tau(t, d_1)$	t_{d_1}	$P_{d_1}(t)$	$\tau(t, q)$	$P_{d_1}(t) \cdot \tau(t, q)$
bat	0.20	1	1	0.40	1	0.40
ball	0.10	0	5	0	0	0
night	0.05	0	1	0	0	0
cricket	0.20	0	5	0	1	0
hit	0.30	1	5	0.60	0	0
baseball	0.15	0	1	0	0	0
$\sum_t$	1.00			1.00		0.40

However in the case of d_1, this document contains the term “hit”. As this term is also similar to “cricket”, “baseball”, and “ball”, the chances are that the probabilities of some of these terms are likely to be transferred to “hit” instead of “bat”; this is shown in table 10.6. As “bat” is the only query term contained in d_1, this results in

d_1 having a lower estimated probability of relevance than d_2 (see table 10.6), which means that d_2 is ranked higher than d_1!

So what this example seems to show is that imaging on a document gives preference to those documents which contain query terms appearing in unusual contexts. In terms of term senses, the supposition is that this form of imaging will rank higher those documents which hold query terms used in unusual senses.

Table 10.7 Evaluation of $P(q \to d_1)$ by imaging on q.

t	$P(t)$	$\tau(t,q)$	t_q	$P_q(t)$	$\tau(t,d_1)$	$P_q(t) \cdot \tau(t,d_1)$
bat	0.20	1	1	0.70	1	0.70
ball	0.10	0	4	0	0	0
night	0.05	0	1	0	0	0
cricket	0.20	1	4	0.30	0	0
hit	0.30	0	1	0	1	0
baseball	0.15	0	1	0	0	0
$\sum_t$	1.00			1.00		0.70

When imaging on a query, the method of probability transfer is similar to imaging on documents except that the transfer is onto the terms in the query. Unlike imaging on documents this form of imaging is unaffected by the context in which query terms appear. The transfer of probabilities to the query terms is the same regardless of what document is being retrieved. Table 10.7 shows the estimated probability of relevance for d_1. It is left as an exercise to the reader to show that d_2 will be assigned the same score.

A comparison of the two techniques and the relative effects shows that the effect that imaging on documents has on those containing ambiguous query terms is caused because the imaging technique is influenced by all the terms of a document and not just those that appear in the query. It is not clear whether this effect of preferring documents containing query terms in unusual senses or contexts is desirable. Term weighting schemes such as the popular $tf \cdot idf$ do give preference to unusual terms appearing in a document in unusually large quantities (Salton and Yang, 1973). Therefore one might think that this preference for the unusual might indicate that the imaging effect is desirable. However if a user enters a query term it would seem reasonable to expect him or her to intend the most common sense. Evaluations in progress will enable us to make informed judgements.

10.6 IMPLEMENTATION ISSUES

In this section I report on the current work toward the implementation of the above models. This section is divided in three subsections reporting work in progress on implementing imaging using ad hoc techniques, on top of Probabilistic Datalog, and on top of the $\mathcal{L}_1$ Probabilistic Logic.

10.6.1 Ad hoc implementation

First in (Crestani and van Rijsbergen, 1995) and later in (Crestani and van Rijsbergen, 1995) ad hoc implementations of RbLI and RbGLI were proposed and tested on standard test collections. These implementations were developed as a set of C programs on an UNIX platform. The main purpose of these studies was to test the effectiveness and feasibility of RbLI and RbGLI. In this chapter I do not address the implementation details, only the implementation choices made. In particular I report on the choice of functions for prior probability, the accessibility relation, and opinionated probability.

The problem of determining an appropriate prior probability distribution over the set of index terms is one of the oldest problems of IR and many models have been proposed for this purpose (van Rijsbergen, 1979; Robertson and Sparck Jones, 1976). The problem could be translated into finding a measure of the importance of a term in the term space, where this importance is related to the ability of the term to discriminate between relevant and non-relevant documents. The importance of the term in the term space seems a reasonable rationale for a probability function. In our implementation we used the *Inverse Document Frequency* (idf), a measure which assigns high discrimination power to terms with low and medium collection frequency, defined as:

$$idf(t) = -log \; \frac{n}{N}$$

where n is the number of documents in which t occurs, and N is the number of documents in the collection.

Strictly speaking, idf is not a probability measure since $\sum_t idf(t) \neq 1$, however we assumed it to be monotone to $P(t)$. We can use this as an estimate of probability because we require only a ranking of the documents in response to a query; the exact probability values are not used.

The accessibility relation necessary to implement any form of imaging was calculated using the *Expected Mutual Information Measure* (EMIM) as measure of similarity. We chose EMIM because it was used with success by many and because it was a well accepted measure in Lexicography (Church and Hanks, 1989; Brown et al., 1992). The EMIM between two terms is often interpreted as a measure of the statistical information contained in the first term about the other one (or vice versa, it being a symmetric measure). EMIM was calculated as follows:

$$EMIM(t_i, t_j) = \sum_{t_i, t_j} P(t_i, t_j) \; log \frac{P(t_i, t_j)}{P(t_i) \; P(t_j)}$$

where t_i and t_j can assume either the value 1 (for $t \in d$) and 0 (for $t \notin d$). When we apply this measure to the probabilistic term space we can estimate EMIM between two terms using the technique proposed in (van Rijsbergen, 1977), pp. 116. This technique makes use of co-occurrence data which can be derived by a statistical analysis of the term occurrences in the collection. Using this measure we evaluated for every term a ranking of all the other terms according to their decreasing level of similarity with it. We stored this information in a file which was used at run-time to determine for a non-d-term its closest d-term.

The opinionated probability function was approximated using a step function, that is a discrete monotonically decreasing transfer function that transfers from a non-d-term a decreasing fraction of its probability to d-terms ordered in decreasing order of similarity. In particular, to simplify computations, in RbGLI, we transferred probability from each non-d-term only to its first 10 most similar d-terms. The transfer function we used works in such a way that the ith of these 10 terms ordered in decreasing order of similarity always gets a fraction of $P(t)$ that is double the one the $i+1$th gets.

Using the above choice of accessibility relation opinionated probability we performed some experiments comparing RbLI and RbGLI with retrieval using classical IR models. So far these are the only experiments performed on imaging. Results achieved using standard test collection (average size about $6,000$ documents) are reported in (Crestani and van Rijsbergen, 1995; Crestani and van Rijsbergen, 1995). Experiments with a larger test collection (size about $170,000$ documents) are currently in progress (Crestani et al., 1995). The problems encountered in this experimentation are covered in section 10.7.

10.6.2 Implementation of imaging on top of probabilistic datalog

The possibility of implementing imaging on top of Probabilistic Datalog was first proposed in (Rölleke, 1995) and was studied and presented in more detail in (Crestani and Rölleke, 1995).

Probabilistic Datalog (Datalog$_P$) is an extension of stratified Datalog (Hullman, 1988) proposed in (Fuhr, 1995). The basic ideas of Datalog$_P$ are the assignment of probabilistic weights to facts and the computation of the weights of derived facts by means of intensional semantics. For an elaborated description of the complete syntax and semantics of Datalog$_P$ and the evaluation process refer to (Fuhr, 1995; Fuhr and Rölleke, 1995).

The way one can implement RbGLI on top of Datalog$_P$ can be better understood via an example. Let us consider table 10.2 which shows an example of computing the probability $P(d \to q)$ using RbGLI. Now we are going to implement information on term probability, term occurrence and opinionated probability function in Datalog$_P$.

Let us first define the probabilities of the terms (relation *term*) and the occurrence of the terms in a document (relation *docTerm*):

$$
\begin{array}{ll}
0.20 & term(t_1). \\
0.10 & term(t_2). \\
0.05 & term(t_3). \\
0.20 & term(t_4). \\
0.30 & term(t_5). \\
0.15 & term(t_6).
\end{array}
$$

$$
\begin{array}{l}
docTerm(d, t_1). \\
docTerm(d, t_5). \\
docTerm(d, t_6).
\end{array}
$$

$$!term(\_).$$

The facts *term* are declared to be disjoint (*!term(_)*). This corresponds to the disjointness of possible worlds within the imaging model. For the semantics of Datalog$_P$, disjointness means that there exists no world where more than one fact of the relation *term* is true.

The relation *transfer* is used to determine the portion of the probability of a non-d-term t_i to be transferred to a d-term. Thus this relation is used to express the shifting of the probabilities of non-d-terms to d-terms modelling the opinionated probability function. The facts of *transfer* are disjoint with respect to the same disjointness key.

The following are the facts related to the accessibility relation. The clause (*!transfer(dk,dk,_)*) indicates that the first and second attribute form the disjointness key.

$$
\begin{array}{ll}
0.67 & transfer(d, t_2, t_1). \\
0.33 & transfer(d, t_2, t_6). \\
0.67 & transfer(d, t_3, t_5). \\
0.33 & transfer(d, t_3, t_6). \\
0.67 & transfer(d, t_4, t_5). \\
0.33 & transfer(d, t_4, t_1).
\end{array}
$$

$$!transfer(dk, dk, \_).$$

The following Datalog$_P$ program implements standard imaging:

about(D,T) :- docTerm(D,T) & term(T).
about(D,T) :- transfer(D,T',T) & term(T').

The query

?- about(d,t_1).

yields (0.333 ()) as an answer, since

$$
\begin{array}{rcl}
\omega(about(d,t_1)) & = & P(about(d,t_1)) \\
& = & P(docTerm(d,t_1) \wedge term(t_1) \vee \\
& & transfer(d_1,t_2,t_1) \wedge term(t_2) \vee \\
& & transfer(d_1,t_4,t_1) \wedge term(t_4)) \\
& = & 1.0 * 0.2 + 0.67 * 0.1 + 0.33 * 0.2 = 0.333
\end{array}
$$

We can formulate the whole query for all documents about t_1, t_4, and t_6 as

q(D)	*:-*	*about(D,t_1).*
q(D)	*:-*	*about(D,t_4).*
q(D)	*:-*	*about(D,t_6).*
	?-	*q(D).*

The result is (0.533 (d_1)), the same reported in table 10.2 obtained with an ad hoc implementation.

One problem with the implementation of imaging on top of Datalog$_P$ is that it is possible to evaluate the weights and the parameters of the transfer function externally, since Datalog$_P$ does not possess the primitives to enable such calculations. A possible partial solution to this problem was pointed out in (Crestani and Rölleke, 1995). It is necessary to produce these facts externally and pass them to the Datalog$_P$ program that performs all the probability transfer and gathering operations.

On the other hand, from the use of Datalog$_P$ as an implementation platform for imaging we gain the possibility of combining the probability kinematics defined by imaging on terms with other probabilistic knowledge. The expressiveness of the knowledge representation and query language is increased, because typical IR knowledge and queries may be combined with typical database knowledge and queries. Moreover, imaging on Datalog$_P$ allows for implementing any kind of probability function, since the computation of the weight is done externally. A subset of the possible transfer functions could be computed using Datalog$_P$ itself. Sadly, for the opinionated probability function described in section 10.6.1 we have not found a way to compute it internally.

10.6.3 Implementation of imaging on top of probabilistic logic

First (Sebastiani, 1996) and later (Crestani et al., 1996b) proposed an implementation of imaging on top of the $\mathcal{L}_1$ probabilistic logic. The characteristics of the $\mathcal{L}_1$ probabilistic logic are discussed in detail in (Halpern, 1990). The logic allows the expression of real-valued terms of type $w_{\langle x_1,\ldots,x_n\rangle}(\alpha)$, where α is a standard first order formula, with the meaning "the probability that random individuals $x_1,\ldots,x_n$ verify α". It also allows their comparison by means of standard numerical binary operators, resulting in formulae that can be composed by the standard sentential operators of first order logic.

It is also worthwhile to notice that, similarly to what happens in the implementation of imaging on top of Datalog$_P$ (Rölleke, 1995; Crestani and Rölleke, 1995), practically all the entities that participate in the imaging process are given an explicit representation in the language of $\mathcal{L}_1$. However, unlike the implementation of imaging on top of Datalog$_P$, in the implementation of imaging on top of the $\mathcal{L}_1$ probabilistic logic, an explicit representation is given to:

- the formula that computes the prior probabilities of keywords;
- the formula that computes the similarities between keywords;
- the formula that chooses the recipients of a probability transfer and computes the revised probabilities of these recipients.

This hints at the way in which experimentation with different formulae encoding different methods of computation of the above features may be tried. In this sense, the whole information retrieval process is modelled as *a proper theory* of $\mathcal{L}_1$, whose role is that of a platform for experimentation of different models. Such a proper theory is obtained by assembling together various sets of formulae, each representing a class of entities participating in the process.

In the following I describe in detail the characterisation of the IR model based on standard imaging. In order to represent standard imaging, a first subset of formulae is needed to identify keywords and documents. This is necessary, as the domain of interpretation must be restricted to deal with only those types of individuals which are the only entities of interest in the revision processes. Assuming that $\{t_1,\ldots,t_n\}$ is the set of terms by means of which documents are represented, and that $\{d_1, \ldots, d_m\}$ are the documents in our collection, we need the following formulae:

$$\begin{array}{l} Term(t_1) \ \wedge \ldots \wedge \ Term(t_n) \\ Document(d_1) \ \wedge \ldots \wedge \ Document(d_m) \\ \forall x.[x = t_1 \vee \ldots \vee x = t_n \vee x = d_1 \vee \ldots \vee x = d_m] \\ \forall x.\neg(Document(x) \wedge Term(x)) \end{array}$$

Documents and terms are individuals belonging to the domain of discourse of a first order interpretation. This is a key feature of this approach as it is in the implementation of imaging on top of Datalog$_P$. In contrast, in the ad hoc implementation, terms were (propositional) interpretations and documents were propositions.

The next subset of formulae is the one that specifies term occurrence (which documents are indexed by which terms). We represent it by the formula:

$$w_x(Occ(t_i, d_j)) = o_{ij} \qquad o_{ij} \in \{0, 1\}$$

for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$, where o_{ij} is 1 iff t_i occurs in d_j.

Next, the probability of each term t_i is specified by means of the set

$$w_x(x = t_i \mid Term(x)) = p_{t_i} \qquad p_{t_i} \in [0, 1]$$

for all $i = 1, \ldots, n$. The formulae account for the case in which we want to input the probability values p_{t_i} from the outside. Alternatively, the probability values can be computed within $\mathcal{L}_1$ from the available occurrence data, for example their inverse document frequency. Then the formula would become

$$w_x(x = t_i \mid Keyword(x)) = -log(w_y(Occ(t_i, y) \mid Document(y)))$$

We are computing the probabilities of keywords as their inverse document frequency; the formula $w_y(Occ(t_i, y) \mid Document(y))$ is in fact to be read as "the probability that, by picking a random document y, keyword t_i occurs in y". For it to truly represent idf, though, we must assume that documents are picked with equal probability, which we state as

$$\forall xy.(Document(x) \wedge Document(y)) \Rightarrow [w_z(x = z) = w_z(y = z)]$$

which is to be read "if x and y are documents, the probability that by picking an individual at random x is picked is equal to the probability that by picking an individual at random y is picked". Alternatively, we may choose to include the previous three formulae in the representation. In this way, probability values are precomputed "externally" and input to the reasoning process acting as "integrity constraints". In what follows we use the expression $P(t_i)$ as a shorthand for the expression $w_x(x = t_i \mid Term(x))$.

The next subset of formulae is the one that specifies the similarity between terms, that is how similar term t_i is to term t_j for all $1 \leq i, j \leq m, i \neq j$.

$$Sim(t_i, t_j) = s_{i,j}$$

Only similarities between non equal terms are specified; in fact the case $i = j$ is of no interest for imaging methods, and its specification would complicate our formulations. Values $s_{i,j}$ are input from an external source of information. Alternatively, they can be computed from within $\mathcal{L}_1$ from the available occurrence values; for instance, they may be taken to be equivalent to the degree of coextensionality of the Occ predicate and computed by means of the formula:

$$Sim(t_i, t_j) = w_x(Occ(t_i, x) \mid Occ(t_j, x)) \cdot w_x(Occ(t_j, x) \mid Occ(t_i, x))$$

or else be computed according to some other measure of similarity (such as for example, the EMIM measure adopted in (Crestani and van Rijsbergen, 1995)). On the other hand, the above formulae may act as integrity constraints. Further integrity constraints may be added if one's theory of similarity requires one to do so, in order to capture further properties of similarity, like for example symmetry.

The following subset of formulae, specifies for each term, how the most similar term can be computed within $\mathcal{L}_1$ from the already available similarity data:

$$MostSim(t_i, t_{k_i}) \Leftrightarrow \neg\exists t_j.[Sim(t_i, t_j) \geq Sim(t_i, t_{k_i})]$$

Alternatively, one can input the most similar terms from the outside.

Next, we have to show how to calculate the revised probability of term t_i by imaging on document d_j in order to implement the probability transfer function. The revised probabilities are specified for $1 \leq i \leq n$ by

$$\begin{aligned} P_{d_j}(t_i) = w_x(Occ(t_i, d_j)) \cdot [P(t_i) + \\ + \textstyle\sum_{k=1}^{n} [P(t_k) \cdot w_x(\neg Occ(t_k, d_j)) \cdot w_x(MostSim(t_k, t_i)]] \end{aligned}$$

To compute $P_{d_j}(q)$ we have to indicate by which terms the information need q is indexed:

$$w_x(Occ(t_i, q)) = o_i \qquad o_i \in \{0, 1\}$$

The probability $P_{d_j}(q)$ evaluated by imaging on d_j may be then calculated as:

$$P_{d_j}(q) = \sum_{i=1}^{n} w_x(Occ(t_i, q)) \cdot P_{d_j}(t_i)$$

This example of the implementation of (standard) imaging on top of the $\mathcal{L}_1$ probabilistic logic suggests that $\mathcal{L}_1$ is a convenient and powerful platform for fast prototyping since it enables the evaluation of all the information necessary for the imaging process internally. External evaluation is required by the implementation of imaging on top of Datalog$_P$. As pointed out in (Sebastiani, 1996), there are both advantages and disadvantages to having an internal definition/computation of the similarities between terms and their prior and posterior probabilities. The $\mathcal{L}_1$ approach has the advantage of being more self-contained and conceptually attractive, as it requires a minimum amount of data to be provided from outside the reasoning mechanism. Moreover, with a minimal coding effort, different probability kinematics methods may be experimented with and compared. The price to be paid for this is that of efficiency, as reasoning in Datalog_P, a less expressive reasoning tool than $\mathcal{L}_1$, is no doubt more computationally tractable[4]. On the one hand, it is plausible to think that data needing to be computed once and for all (such as similarity data between keywords) may be more efficiently computed outside the logic and subsequently fed to it. On the other hand, the possibility of expressing the probability kinematics methods within the logic seems most definitely desirable, particularly if one accepts the logic as a fast prototyping tool.

10.7 EXPERIMENTATION ISSUES

In this section I briefly report on the current state of evaluation of IR effectiveness of the models based on imaging. Most of the work reported here is still in progress.

10.7.1 Experimentation with small test collections

The first set of experiments on the effectiveness of ad hoc implementation of RbLI compared with a classical IR model is reported in (Crestani and van Rijsbergen, 1995). We compared RbLI with an IR model based on the cosine correlation that employed only the idf weighting measure (to make it fare for RbLI that currently uses only idf). The comparison was performed both on the effectiveness level and on the feasibility level. From the point of view of effectiveness both models produced on the *Cranfield 200* test collection (Cleverdon et al., 1966) equivalent results. However, RbLI was distinctly more computationally expensive, and it was necessary to experiment with a few simplifications of the RbLI model that, though falling short of all the logical requirements of the model, enabled us to retain the same level of effectiveness with a considerably lower computational burden.

A more extensive analysis was performed in (Crestani and van Rijsbergen, 1995). There we compared the effectiveness of four models, two were classical models of IR (models equivalent to the cosine correlation and the probabilistic model) and two were RbLI and RbGLI. The analysis was only performed with theoretical goals since both classical models employed were not optimised using ad hoc weighting and normalisation schemas. The main purpose of the analysis was to compare the probability kinematics of classical IR models with the models based on imaging.

Given the heavy computations necessary to perform RbLI and RbGLI and given the theoretical nature of our goals, we decided to use only small collections. The choice fell on the three test collections that have been extensively studied and used in the field of IR: the *Cranfield 1400*, the *CACM*, and the *NPL* test collections (Sparck Jones and van Rijsbergen, 1976). The results of our test are presented using the standard evaluation technique used in IR.

Without entering into the details of the evaluation, since this is beyond the scope of this chapter, it suffices to say from the results we could observed that:

- any model inducing a probability transfer (like those based on imaging or Bayesian conditionalisation) performs better than any model that does not induce such transfer;

- any model that induces a probability transfer from one term to a set of terms (called "one-to-many" transfer, as in general imaging, for example) performs better than any model in which either there is no transfer or the transfer is from one term to a single other term (called "one-to-one" transfer, as in standard imaging);

- any model that induces a one-to-many transfer that takes into account the similarity between the donor and the receivers (all forms of imaging except standard imaging) performs better than any model with a one-to-many transfer that takes into account the probability ratio between the receivers (Bayesian conditionalisation).

These findings are consistent over the three document collections used in the experimentation.

10.7.2 Experimentation with large test collections

Some further tests are currently being performed using a much larger test collection: the *TREC B* document collection, that is used in the context of the Text Retrieval (TREC) initiative (Harman, 1995). This collection is one order of magnitude larger than the ones used in the tests reported in this chapter, being $170,000$ full text documents. Our initial results on the evaluation of RbLI and RbGLI using this collection, reported in (Crestani et al., 1995), are difficult to interpret since we had to make heavy modifications to our models to be able to cope with the large size of the data. I will not attempt to explain the poor results achieved with this much larger collection in this chapter. I shall wait until we are able to implement fully the RbLI and RbGLI models.

10.8 RELATED WORK

The use of imaging in IR was proposed for the first time in (van Rijsbergen, 1989) without, however, specifying how imaging could be used operatively. To the best of my knowledge, there have been only a few attempts to use imaging in IR. I review these attempts briefly.

(Amati and Kerpedjiev, 1992) proposed two logical models for IR. One of them is based on conditional logic and makes use of imaging for the evaluation of $P(d \rightarrow q)$ and $P(q \rightarrow d)$. They offered two different semantics for the evaluation of the two conditionals. For the evaluation of $P(d \rightarrow q)$ they consider a term as a world, while for the evaluation of $P(q \rightarrow d)$ they consider a document as a world. I see a difficulty in taking a document as a world for the fact that the event d in the conditional statement is also interpreted as world. To deal with this difficulty one would have to make explicit the difference between a document as a fictive object existing in its own right and a partial description of such an object. Rather than doing this I have adopted the approach of considering for both evaluation a term as a world.

(Sembok and van Rijsbergen, 1993) proposed a relevance feedback technique based on the use of imaging. Again, the perspective of a document as a world is used; moreover the similarity between documents is evaluated by means of clustering using nearest neighbour. The similarity measure used for the clustering on the document space is based on the Dice's coefficient, a very simple similarity measure. One should notice that, since most of the power of imaging relies on the correct identification of the closest possible world, it is very important to use the best possible similarity measure for the job.

(Nie et al., 1995) used imaging to include user knowledge, domain knowledge, intentions, and so on in an IR model. In their model both documents and queries are propositions. Possible worlds represent different states of the data set, i.e. possible states of knowledge that can be held by users. A document d is true in a world w if the document is "consistent" (the term is used here in a broad sense) with the state of knowledge associated with that world. Worlds differ because they represent different states of knowledge, and given a metric on the world space, we can identify the closest world to w for which d is true. Imaging can then be used for the evaluation of the certainty of the implication $d \rightarrow q$. We have a similar view in this chapter. Both approaches consider a world as an informative entity, in the context of which a document or a query needs to be checked for consistency. The major advantage of Nie et al. model is that it enables user modelling, and therefore the evaluation of a user-oriented measure of relevance, while the models presented in this chapter take into account only a system-evaluated relevance. However, their model will take time to be developed.

Indirectly related to imaging is the approach proposed by (Wong and Yao, 1991; Wong and Yao, 1995) that suggested estimating $P(d \rightarrow q)$ by $P(q \mid d)$. However, the limitations of this approach are known in the area of logics by the name of "triviality results", and were well illustrated in (Lewis, 1981). Without entering into the details of the demonstration, Lewis' result excluded the use of conditional probabilities as a probabilistic logic dealing with conditionals, that is why he suggested estimating the probability of a conditional by using imaging.

Wong and Yao demonstrated that most of the IR models presently in use can be explained in terms of the *probabilistic inference model*, based on the evaluation of the degree of confirmation (or belief, according to the view taken) of the proposition H given evidence E expressed as $P(H \mid E)$ (Wong and Yao, 1995). Conventional IR models can be obtained by associating either d or q with H or E, and by defining different ways of evaluating the probabilities via probabilistic inference on a concept space. Concepts are disjoint elements of the representation space, or are elements transformed in such a way as to be disjoint. Terms are basic concepts.

An important result in (Wong and Yao, 1995) is that their model subsumes the probabilistic model. Both the probabilistic independence indexing model (Fuhr, 1989) and binary probabilistic independence retrieval model (van Rijsbergen, 1979) can be explained in terms of the probabilistic inference model. Taking that result for granted, it is easy to understand that the presence of a normalisation factor or of likelihood ratios does not change the kinematics of probabilities of these models. So, if they can all be explained in terms of the probabilistic inference model, and if this model is based on the concept of conditional probability, than all the conventional models are based on the same kinematics of probabilities. The amount of probability moved from one concept to another may change, but the principle remains the same: the transfer of probabilities provides the minimal revision of the prior probability that is necessary to make the evidence E certain without distorting the profile of probability ratios on the representation space. Moreover, the view taken by Wong and Yao is purely probabilistic, i.e. only probabilistic inference is used for the evaluation of the uncertainty of the implication $E \rightarrow H$. I extend that view by taking into consideration a semantics of the representation space based on PWS, and this enables the evaluation of $P(E \rightarrow H)$ in a less restrictive way than by pure probabilistic inference. Thus I believe that the use of PWS enables us to design and deal with different and more complex models of probability kinematics, like for example those presented in this chapter.

10.9 CONCLUSIONS

In this chapter we have looked at the current state of the use of imaging in IR to develop models based on new form of probability kinematics. Further research is necessary to prove both theoretically and experimentally the validity of the proposed models.

Acknowledgments

I would like to thank all the people that helped me formalising in an understandable way this theory. This work has been partially supported by a “Marie Curie” research fellowship from the European Community.

Notes

1. Notice that the term document is used here in a very general sense. A document is a piece of text, an image, a sound, etc.
2. I simply refer to the Modal System $S5$ and not to more complex models.
3. In that book he also characterised this generalisation of imaging in terms of a homomorphic condition that does not presuppose any kind of possible world semantics, but we remain faithful to our semantics.
4. Only “theoretical” tractability considerations are taken into account. The $\mathcal{L}_1$ logic has not been implemented, while Datalog$_P$ has been, so an experimental comparison cannot be made yet.

References

Amati, G. and Kerpedjiev, S. (1992). An Information Retrieval logical model: implementation and experiments. Technical Report Rel 5B04892, Fondazione Ugo Bordoni, Roma, Italy.

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

Brown, P., Della Pietra, V., deSouza, P., Lai, J., and Mercer, R. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479.

Bruza, P. (1993). *Stratified Information Disclosure: a synthesis between Hypermedia and Information Retrieval*. Phd thesis, Katholieke Universiteit Nijmegen, The Netherlands.

Chiaramella, Y. and Chevallet, J.P. (1992). About retrieval models and logic. *The Computer Journal*, 35(3):233–242.

Church, K. and Hanks, P. (1989). Word association norms, mutual information and lexicography. In *Proccedings of ACL 27*, pages 76–83, Vancouver, Canada.

Cleverdon, C., Mills, J., and Keen, M. (1966). *ASLIB Cranfield Research Project: factors determining the performance of indexing systems*. ASLIB.

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

Cooper, W.S. (1995). Some inconsistencies and misnomers in probabilistic information retrieval. *ACM Transactions on Information Systems*, 13(1):100–111.

Crestani, F. (1996). Kinematics of probabilistic term weights and term space knowledge in information retrieval. In *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT 96)*, volume 2, pages 836–840, Aachen, Germany.

Crestani, F. and Rölleke, T. (1995). Issues on the implementation of imaging on top of probabilistic datalog. In *Proceedings of the First Workshop in IR, Uncertainty and Logic*. Glasgow, Scotland, UK.

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. Experimenting retrieval by logical imaging on TREC. In *Proceeding of the TREC Conference*, Washington D.C., USA.

Crestani, F., Sanderson, M., and van Rijsbergen, C.J. (1996a). Sense resolution properties of logical imaging. *The New Review of Document and Text Management*, 1:277–298.

Crestani, F., Sebastiani, F., and van Rijsbergen, C.J. (1996b). Imaging and information retrieval: variation on a theme. In *Proceedings of the Second International Workshop on Information Retrieval, Uncertainty and Logic*, pages 27–31, Glasgow, UK.

Crestani, F. and van Rijsbergen, C.J. (1995). Information Retrieval by Logical Imaging. *Journal of Documentation*, 51(1):1–15.

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

Croft, W.B. and Harper, D.J. (1979). Using probabilistic models of document retrieval without relevance information. *Journal of Documentation*, 35:285–295.

Cross, C.B. (1994). Eliminative bayesianism and probability revision. Unpublished paper.

Efthimiadis, E. (1996). Query expansion. *Annual Review of Information Science and Technology*, 31:121–187.

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

Fuhr, N. (1989). Models for retrieval with probabilistic indexing. *Information Processing & Management*, 25(1):55–72.

Fuhr, N. (1995). Probabilistic datalog - a logic for powerful retrieval methods. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 282–290, Seattle, WA, USA.

Fuhr, N. and Buckley, C. (1991). A probabilistic learning approach for document indexing. *ACM Transactions on Information Systems*, 9(3):223–248.

Fuhr, N. and Rölleke, T. (1995). A probabilistic relational algebra for the integration of information retrieval and database systems. Submitted for publication.

Gärdenfors, P. (1982). Imaging and conditionalization. *Journal of Philosophy*, 79:747–760.

Gärdenfors, P. (1988). *Knowledge in flux: modelling the dynamics of epistemic states*. The MIT Press, Cambridge, Massachusetts, USA.

Halpern, J. (1990). An analysis of first-order logics for probability. *Artificial Intelligence*, 46:311–350.

Harman, D. (1995). Overview of the fourth text retrieval conference (TREC-4). In *Proceeding of the TREC Conference*, Gaithersburg, MD, USA.

Hullman, J. (1988). *Principles of Database and Knowledge-Base Systems*, volume I. Computer Science Press, Rockville, MD, USA.

Jeffrey, R. (1965). *The logic of decision*. McGraw-Hill, New York, USA.

Kripke, S.A. (1971). Semantical considerations on modal logic. In Linsky, L., editor, *Reference and modality*, chapter 5, pages 63–73. Oxford University Press, Oxford, UK.

Lalmas, M. and van Rijsbergen, C.J. (1993). A model of an Information Retrieval system based on Situation Theory and Dempster-Shafer theory of evidence. In *Proceedings of the 1st Workshop on Incompleteness and Uncertainty in Information Systems*, pages 62–67, Montreal, Canada.

Lewis, D. (1981). Probability of conditionals and conditionals probabilities. In Harper, W., Stalnaker, R., and Pearce, G., editors, *Ifs*, The University of Western Ontario Series in Philosophy of Science, pages 129–147. D.Reidel Publishing Company, Dordrecht, Holland.

Lewis, D. (1986). *Conterfactuals*. Basil Blackwell, Oxford, UK, 1st edition.

Meghini, C., Sebastiani, F., Straccia, U., and Thanos, C. (1993). A model of Information Retrieval based on a Terminological Logic. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 298–307, Pittsburgh, PA, USA.

Nie, J.Y. (1989). An Information Retrieval model based on Modal Logic. *Information Processing & Management*, 25(5):477–491.

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

Nie, J.Y., Lepage, F., and Brisebois, M. (1995). Information retrieval as counterfactuals. *The Computer Journal*, 38(8):643–657.

Qiu, Y. and Frei, H. (1993). Concept based query expansion. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 160–171, Pittsburgh, PA, USA.

Robertson, S.E. (1977). The probability ranking principle in IR. *Journal of Documentation*, 33(4):294–304.

Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27:129–146.

Rölleke, T. (1995). Imaging on top of probabilistic datalog. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, WA, USA. Poster session.

Salton, G. (1968). *Automatic information organization and retrieval*. Mc Graw Hill, New York.

Salton, G. and Yang, C. (1973). On the specification of term values in automatic indexing. *Journal of Documentation*, 29(4):351–372.

Sanford, D. (1989). *If P, then Q: conditionals and the foundations of reasoning*. Routledge, London, UK.

Sebastiani, F. (1996). Information retrieval, imaging and probabilistic logic. In *Deliverable D3: A Theory of Uncertainty for Information Retrieval*, number 1/96 in FERMI, technical report 7, pages 59–65. ESPRIT Basic Research Action, Project Number 8134 - FERMI.

Sembok, T. and van Rijsbergen, C.J. (1993). Imaging: a relevance feedback retrieval with nearest neighbour clusters. In *Proceedings of the BCS Colloquium in Information Retrieval*, pages 91–107, Glasgow, UK.

Seracevic, T. (1970). The concept of "relevance" in information science: a historical review. In Seracevic, T., editor, *Introduction to Information Science*, chapter 14. R.R. Bower Company, New York, USA.

Sparck Jones, K. and van Rijsbergen, C.J. (1976). Information Retrieval test collections. *Journal of Documentation*, 32(1):59–75.

Stalnaker, R. (1981). Probability and conditionals. In Harper, W., Stalnaker, R., and Pearce, G., editors, *Ifs*, The University of Western Ontario Series in Philosophy of Science, pages 107–128. D.Riedel Publishing Company, Dordrecht, Holland.

van Rijsbergen, C.J. (1977). A theoretical basis for the use of co-occurence data in Information Retrieval. *Journal of Documentation*, 33(2):106–119.

van Rijsbergen, C.J. (1979). *Information Retrieval*. Butterworths, London, second edition.

van Rijsbergen, C.J. (1986a). A new theoretical framework for information retrieval. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 194–200, Pisa, Italy.

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

van Rijsbergen, C.J. (1988). Logics for information retrieval. *Note Recensioni e Notizie - Supplemento*, 1-2:121–124.

van Rijsbergen, C.J. (1989). Toward a new information logic. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 77–86, Cambridge, USA.

van Rijsbergen, C.J. (1992). Probabilistic retrieval revisited. Departimental Research Report 1992/R2, Computing Science Department, University of Glasgow, Glasgow, UK.

Wong, S.K.M. and Yao, Y.Y. (1989). A probability distribution model for Information Retrieval. *Information Processing & Management*, 25(1):39–53.

Wong, S.K.M. and Yao, Y.Y. (1991). A probabilistic inference model for information retrieval. *Information Systems*, 16(3):301–321.

Wong, S.K.M. and Yao, Y.Y. (1995). On modelling information retrieval with probabilistic inference. *ACM Transactions on Information Systems*, 13(1):38–68.