Information Retrieval, 2, 15–24 (2000)

Another Look at the Logical Uncertainty Principle

CJ VAN RIJSBERGEN
Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, Scotland

Received November 18, 1998; Revised September 16, 1999

> "Part of our knowledge we obtain direct; and part by argument. The theory of probability is concerned with that part which we obtain by argument, and it treats of the different degrees in which the results so obtained are conclusive or inconclusive."
>
> Keynes, 1929, p. 3.

Abstract. The Logical Uncertainty Principle is re-examined from the point of classical logic. Two interpretations are given, an objective one in terms of an axiomatic theory of information, and a subjective one based on Ramsey's theory of probability.

Keywords: uncertainty, information theory, Bayesian inference, Ramsey test, information retrieval model

Introduction

In classical logic, arguments aim at demonstrating certainty, as Keynes (1929) would have it, thus they claim to be conclusive. There is no doubt that most arguments are inconclusive and that classical logic only has a limited role to play in most contexts. This is especially true of arguments in law, for example, when establishing conclusions beyond reasonable doubt, or when the balance of the probabilities are in favour of the conclusion.[1] The same is true in information retrieval; for example, we only establish the relevance of a document to a degree or probability.

The early models in information retrieval have created the illusion that retrieval is an exact science. A decision to retrieve is based on the success of a match, or in logical terms the successful interpretation of a query in a document. Thus if a query such as $(A \wedge B) \vee C$ is specified using terms A, B, C then all those documents that contain A and B, or C are retrieved. If a document representation is thought of as a possible interpretation of the query, then all those interpretations which make the query true are retrieved, that is, all the models. Alternatively we can describe this as retrieving all the documents which prove the query. Thus stated, the role of logic is explicit, and the underlying matching process can be interpreted as the application of simple deduction. This works because we identify the presence of a term with its achieving truth in the document, i.e. we define a simple semantics for index terms.

The limitations of this classical approach are immediately apparent, or after a little thought, once one has tried to use systems based on it. When users use IR systems they seek documents relevant to their information need, or documents about topics they describe in their queries. To assume that such documents can be found by exact matching, or through classical logic, is to miss the point. Many documents that fail to imply the query deductively will turn out to be relevant. Relevance is a matter of degree. The central problem of IR is how to model and measure that degree of relevance.

There are two extreme approaches possible. The first is to have a crude model of retrieval but to involve a user in the retrieval process as part of a feedback loop. The second is to have a refined model of retrieval that delivers relevant documents with little interaction. There is no doubt that the first approach is currently more attractive especially now that the technology is able to support it adequately. The model of retrieval must be such that the user can grasp its meaning and the way it functions. The model should be seen as a tool with which a user explores the store of information. Such a tool can be based on deductive logic but as shown above that is not enough.[2]

To step outside classical logic we must consider the role of *evidence* in IR. Another way of looking at retrieval is as a form of inference or argumentation where evidence allows us to derive conclusions with a degree of certainty. If one thinks of a query as a hypothesis of what each document in the system might be about, then each document provides evidence in support of that hypothesis to a different degree. There is a long tradition how this form of plausible inference might be described within a logical framework e.g. degree of partial entailment, degree of inferential soundness (Waismann 1930). In a way, they all assume that provability, entailment, soundness etc. is a matter of degree. In the past there has been considerable controversy about this. Here, the approach to this form of plausible inference will be to express it using standard theoretical tools such as Probability Theory and Information Theory but in relation to classical limits. By this is meant that in the *limit* my analysis behaves in a classical way but that it deviates from it in a well controlled manner. This will become clearer as we proceed.

As an example, let us take a classical law of inference Modus Ponens

$$A, A \supset B \Rightarrow B \qquad \text{MP}$$

This is an accepted law of logic, it states that from A and $A \supset B$ one can infer B. It says nothing about the prior truth of A or $A \supset B$. To use it we need to establish the truth (or certainty) of A and $A \supset B$ and hence derive the truth (or certainty) of B. To present MP in this way is somewhat misleading in that often, in practice, we only know A and wish to infer B. In such a situation we seek to establish $A \supset B$ so that we can indeed infer B. So far this view of the matter is entirely classical.

In most familiar situations of reasoning neither A nor $A \supset B$ is known with certainty, instead we only know them with a degree of certainty. To simplify the discussion we will concentrate initially on uncertainty associated with the conditional information. How to express this? Traditionally we express it by a conditional probability, namely $p = P(B \mid A)$, which reads given A the probability of B is p. The assumption here is that A is known with certainty and that B may be hypothetical. We can now claim that $P(B \mid A)$ is a measure of the certainty of the inference $A \Rightarrow B$. To see this consider the following

$$P(B \mid A) = P(A \supset B \mid A)$$

$$\begin{aligned} \text{Because} \qquad P(A \supset B \mid A) &= P(\neg A \vee B \mid A) = P((\neg A \vee B) \wedge A)/P(A) \\ &= P(B \wedge A)/P(A) \text{ by distribution} \\ &= P(B \mid A) \end{aligned}$$

In the case of MP we assume A and $A \supset B$ with certainty. But as we mentioned before we usually do not know these with certainty, so a less extreme case is when we know A with certainty and have a probability $P(A \supset B \mid A)$ measuring the certainty of $A \supset B$ given A.[3] As it happens this is the same as $P(B \mid A)$ and so it seems reasonable to assume that $P(B \mid A)$ is a measure of the certainty of the partial entailment $A \Rightarrow B$. Note that what we have done here, to infer B from A, is to import some conditional information the measure of which we associate with the degree of partial entailment.

This way of looking at things once again demonstrates that we are arguing from the limiting case given by classical logic. To drive this point home consider the following definition given by Popper (1968):

$$A \Rightarrow B \text{ iff } \forall X\ P(B \mid A, X) = 1$$

This constitutes a kind of terminating condition for inferring B from some information, namely, we continue to augment the antecedent until any further addition stabilises the probability. In the limiting case, when $A \Rightarrow B$ it is clear that the information content of B is included in that of A. But when $P(B \mid A) = p$, a measure of how much of the information is *not* included in A is given by $1 - p$. Later, it will be argued that a measure of the augmenting premises $X_1, X_2, \dots X_n$ such that $P(B \mid A, X_1, X_2, \dots X_n) = 1$ is a measure of the certainty of $A \Rightarrow B$.

Let us now return to the situation where the antecedent A itself is uncertain. The first thing to realise here is that when using Bayes' Theorem to compute $P(B \mid A)$ there is a probability associated with A but that this is the prior probability and that at the point of inference $P(A \mid A) = 1$, i.e. A is known with certainty (given). The situation we are in is one in which A (typically the evidence) is not known with certainty itself, the simplest case of this would be if A itself was conditional on another event E which was known so that

$$P(B \mid E) = P(B \mid A, E)P(A \mid E) + P(B \mid \neg A, E)P(\neg A \mid E)$$

But it is not always possible to find such a certain event E. Hence we must consider the case where the *passage of experience* (Jeffrey 1983) gives rise to a new probability function P^* which is a revision of P in the light of such non-propositional experience (instead of E) and for which $P^*(A) \neq 1$ in general. Now to measure the uncertainty of the inference $A \Rightarrow B$ when A itself is uncertain I would propose the Jeffrey conditional

$$P^*(B) = P(B \mid A)P^*(A) + P(B \mid \neg A)P^*(\neg A).$$

Clearly when $P^*(A) = 1$ we are back in the situation where A is certain. The interpretation of $P(B \mid A)$ and $P(B \mid \neg A)$ are now somewhat problematical because they are, as it were, probabilities of subjunctives. We interpret $P(B \mid A)$ as $P(A \rightarrow B)$ where '$\rightarrow$' is now not the material conditional. Such a conditional would be logically stronger than the material conditional, i.e. A $\rightarrow$ B only if A $\supset$ B, and since it is logical stronger more informative.

The logic of uncertainty

In the above a pattern emerges: to evaluate the uncertainty of an inference leading from premises to conclusions we measure the extent to which we need to augment the premises to infer the conclusion with certainty. We can describe this as measuring the strength of argument leading from premises to conclusion, or as the extent to which premises (evidence) entail the conclusion. Furthermore the information we add to the premises is typically in the form of conditional information. Hence it is critical that we evaluate the uncertainty of a conditional statement appropriately; the case for this was first made in Van Rijsbergen (1986).

First let us consider the case without probabilities. To analyse this case we will need to introduce possible world semantics. An intuitive understanding of a possible world is that it is a complete specification of how things are, or might be, down to the finest semantically relevant details. For our purposes, we identify documents with possible worlds.

Let s be a partial description of a document—this might be a set of sentences, or just a single index term—q being a request. In deciding whether to retrieve a document we would need to evaluate $s \rightarrow q$, that is whether $s \rightarrow q$ is true or not. If s is true in a document d then $s \rightarrow q$ is true providing q is true. If s is not true in a document then we go to the nearest document d' to d in which it is true and consider whether q is true. If q is true in d' then $s \rightarrow q$ is true in d, otherwise false. This is a simple example of evaluating what is now commonly known as a Stalnaker conditional originally formulated by Stalnaker (1970) and has been the subject of extensive research ever since, see for example the excellent collection of papers edited by Eells and Skyrms (1994).

To give a simple example, s might be an index term, q the same or a different index term. If $s = q$, then $s \rightarrow q$ is true follows trivially for those documents in which q occurs. The more interesting case is when $s \neq q$. In that case, to establish $s \rightarrow q$ in d find the nearest d' in which s occurs and check for the occurrence of q. If the semantics is more complex and allows relationships between index terms, then we can handle more complicated inferences, perhaps via a thesaurus. For example, let s = FORTRAN, q = PROGRAMMING LANGUAGE and for argument sake let FORTRAN be false in d but true in d' the nearest such document. If q, that is PROGRAMMING LANGUAGE now evaluates to true in d', which it will because Fortran is a programming language, then $s \rightarrow q$ (in d). More examples are given in Lalmas (1998) and Ounis (1998). The above process illustrates what is now widely known as the *Ramsey test*, recent work on it may be found in Lindström and Rabinowicz (1992), whereas the original statement by Ramsey appears in Mellor (1976). It might be summarised as follows:

> To evaluate a conditional, first hypothetically make the minimal revision of your stock of beliefs required to assume the antecedent. Then evaluate the acceptability of the consequent on the basis of this revised body of beliefs.

In document retrieval we are often faced with the situation where $s \rightarrow q$ is assumed false because s does not logically imply q. That is, assuming the truth of the sentences (index terms) in a document we cannot arrive at q. Boolean retrieval is an excellent example: given a truth valuation for the terms describing a document, we retrieve those documents which imply q (make q true for that valuation), and do not retrieve if q evaluates false. On the other hand, Ramsey's belief revision approach to evaluating a conditional suggests that a given document should be revised in a minimal way that makes s true. If, after that revision, q is true, then $s \to q$ is true and d should be retrieved. There are a number of ways of making this revision. One could restrict the revision to selecting a nearest document in which s is true, in which case no interaction from the user would be required. Or, one could involve the user in expanding the information contained in the document under consideration. Or, finally, one could do document expansion automatically using information already stored in the system, perhaps with the aid of a thesaurus as illustrated above. We will return to this notion of *minimal revision* and discuss it in information-theoretic terms later.

The possible world semantics and the Ramsey test motivate a generalisation for application to Information Retrieval. This new principle was called the Logical Uncertainty Principle and was first introduced in more detail by Van Rijsbergen (1986); that paper is reproduced in Crestani et al. (1998) together with a number of closely related papers. The reader is referred to the latter book for a thorough discussion. The Principle states:

> Given any two sentences x and y; a measure of the uncertainty of $y \to x$ relative 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 $y \to x$.

In this principle '$\to$' is left unspecified, it might be interpreted as 'answers' or 'infer'. A classical interpretation would be that '$\supset$' is the material conditional, and then establishing the truth of $y \supset x$ through the Deduction Theorem would allow us to infer x from y. There is then also a straight forward interpretation of minimal information to be added. It turns out that $y \supset x$ is the logical weakest parcel of extra information that together with y will yield x via the use of Modus Ponens. So here we have an example of adding information to a minimal extent (Van Rijsbergen 1989).

Extra information revisited

Let us now examine the concept of extra information a little more closely. To do this we first cast the analysis in terms of information carried by evidence d about a hypothesis q. The intention is that we think of a document description as providing evidence through its relative information content for or against a query. An axiomatic definition of such a measure $\mathrm{I}\langle d \mid q\rangle$ was given by Hilpinen (1970). Not all the details of the measure are reproduced here, for that the reader should consult Hilpinen's original paper. Moreover the measure is merely illustrative, proofs are not given, and it is very likely that other measures could be used; its main purpose is to give an example of how "additional information" can be connected to probability. For readability's sake, the angle brackets are used in $\mathrm{I}\langle .|.\rangle$ the information measure and the round brackets in $\mathrm{P}(.|.)$ the probablity measure.

Q1: If $\vdash d \equiv d'$ and $\vdash q \equiv q'$, then $\mathrm{I}\langle d \mid q\rangle = \mathrm{I}\langle d' \mid q\rangle$ and $\mathrm{I}\langle d \mid q\rangle = \mathrm{I}\langle d \mid q'\rangle$.
Q2: $0 \leq \mathrm{I}\langle d \mid q\rangle \leq 1$.
Q3: If $\vdash d \supset q$ then $\mathrm{I}\langle d \mid q\rangle = 1$.

Q4: If $\vdash d \vee d' \vee q$ then $\mathrm{I}\langle d \wedge d' \mid q\rangle = \mathrm{I}\langle d \mid q\rangle + \mathrm{I}\langle d' \mid q\rangle$.
Q5: $\mathrm{I}\langle d \vee d' \mid q\rangle = \mathrm{I}\langle d \mid q\rangle \mathrm{I}\langle d' \mid q \vee d\rangle$.

$\mathrm{I}\langle d \mid q\rangle$ is defined as a quantitative measure; it is a function of arguments which are sentences of a language. $\mathrm{I}\langle d \mid q\rangle$ is assumed to have a value only for q that is not logically true; this is plausible since our evidence will have nothing to say about tautologies. The plausibility of these axioms derive from the theorems they entail and the extent to which these match our intuitions about information. There is no claim intended that this measure is in some sense unique; it is an example of a measure of information that will enable us to give computational bite to the Logical Uncertainty Principle.

I will now briefly discuss these axioms; a more detailed discussion can be found in Crestani et al. (1998). Q1 says that logically equivalent sentences carry the same information about a hypothesis, and one piece of evidence carries the same information about different hypotheses. It is reasonable to assume that, if $\vdash d^4$ then $\mathrm{I}\langle d \mid q\rangle = $ min., i.e. tautologies contain no information; also, if $\vdash d \supset q$ then $\mathrm{I}\langle d \mid q\rangle = $ max., i.e. d contains all the information in q.

From this we derive that

$$\text{If} \vdash d \supset q \text{ but not } \vdash d' \supset q' \text{ then } \mathrm{I}\langle d \mid q\rangle \geq \mathrm{I}\langle d' \mid q'\rangle.$$

By convention we take max. as 1 and min. as 0, hence our axioms Q2 and Q3. Under some conditions information is simply additive. To make this precise we need the notion of ‘information in common’. Traditionally this is presented by a disjunction; so, $d \vee d'$ carries the information common to d and d' although each may carry more. If $d \vee d' \vee q$ is a logical truth[5] then d and d' contain no information in common about q as Hilpinen (1970) put it, hence axiom Q4. If on the other hand d' and d do convey common information about q then

$$\mathrm{I}\langle d \wedge d' \mid q\rangle = \mathrm{I}\langle d \mid q\rangle + \mathrm{I}\langle d' \mid q\rangle - \mathrm{I}\langle d \vee d' \mid q\rangle$$

Keynes (1929) in his Theory of Probability presented a formulation of ‘the weight of an argument’ which increased with the accumulation of evidence. We can express this here as

$$\mathrm{I}\langle d \wedge d' \mid q\rangle \geq \mathrm{I}\langle d \mid q\rangle$$

Another way of putting this is to say that the information carried increases with the strength of argument,

$$\text{If} \vdash d \supset d' \text{ then } \mathrm{I}\langle d \mid q\rangle \geq \mathrm{I}\langle d' \mid q\rangle,$$

or the information decreases with the weakening of the argument,

$$\mathrm{I}\langle d \vee d' \mid q\rangle \leq \mathrm{I}\langle d \mid q\rangle.$$

It turns out that the above is implied by Q2 and Q4. The following theorems also follow from Q1–Q5:

$$\mathrm{I}\langle \neg d \mid q\rangle = 1 - \mathrm{I}\langle d \mid q\rangle, \text{ and}$$
$$\text{If } \vdash q \supset d \text{ then } \mathrm{I}\langle d \vee d' \mid q\rangle = \mathrm{I}\langle d \mid q\rangle \mathrm{I}\langle d' \mid d\rangle$$

The above axioms define a relative measure of information. It is possible to define an absolute measure by re-expressing the axioms of I, replacing q everywhere with a logically false hypothesis f:

F1: If $\vdash d \equiv d'$ then $\mathrm{I}\langle d \mid f\rangle = \mathrm{I}\langle d' \mid f\rangle$
F2: $0 \leq \mathrm{I}\langle d \mid f\rangle \leq 1$
F3: If $\vdash \neg d$ then $\mathrm{I}\langle d \mid f\rangle = 1$
F4: If $\vdash d \vee d'$ then $\mathrm{I}\langle d \wedge d' \mid f\rangle = \mathrm{I}\langle d \mid f\rangle + \mathrm{I}\langle d' \mid f\rangle$
F5: $\mathrm{I}\langle d \vee d' \mid f\rangle = \mathrm{I}\langle d \mid f\rangle \mathrm{I}\langle d' \mid d\rangle$

These axioms are satisfied by other well-known measures of information. Moreover $1 - \mathrm{I}\langle d \mid f\rangle$ satisfies the axioms of the probability calculus, so we can define

$$\mathrm{P}(d) = 1 - \mathrm{I}\langle d \mid f\rangle$$

The distinction between absolute and relative information is highlighted by

$$\mathrm{I}\langle d \mid q\rangle = \frac{\mathrm{I}\langle d \vee q \mid f\rangle}{\mathrm{I}\langle q \mid f\rangle}$$

or in words the information carried by d relative to q is equal to the ratio of the absolute information in common between d and q and the absolute information carried by q, i.e. the proportion of the information carried by d in common with q. The information $\mathrm{I}\langle d \mid f\rangle$ is often called the *transmitted* information. It is also straight forward to show that

$$\begin{aligned}\mathrm{I}\langle d \mid q\rangle &= \frac{1 - \mathrm{P}(q \vee d)}{1 - \mathrm{P}(q)} \\ &= 1 - \mathrm{P}(d)\frac{1 - \mathrm{P}(q \mid d)}{1 - \mathrm{P}(q)} \\ &= \mathrm{P}(\neg d \mid \neg q)\end{aligned}$$

Thus by defining a measure of information we have recovered a probability function. In many ways a concept of information is more fundamental to the development of a model for IR than probability (Van Rijsbergen and Lalmas 1996).

Given the axiomatic definition of $\mathrm{I}\langle .|.\rangle$ can we use it to formulate the Logical Uncertainty Principle?[6] The answer is Yes, I think! It goes as follows. If $\mathrm{I}\langle d \mid q\rangle \neq 1$, that is not the case that $\vdash d \supset q$, then according to LUP we need to augment d by d' in such a way that $\mathrm{I}\langle d \wedge d' \mid q\rangle = 1$. By the additivity axiom when

$$\vdash d \vee d' \vee q \text{ we have that}$$
$$1 = \mathrm{I}\langle d \wedge d' \mid q\rangle = \mathrm{I}\langle d \mid q\rangle + \mathrm{I}\langle d' \mid q\rangle$$

Hence $\mathrm{I}\langle d' \mid q\rangle$ is a measure of the uncertainty associated with $d \vdash q$. We can of course give this a probabilistic expression since

$$\begin{aligned} \mathrm{I}\langle d' \mid q\rangle &= 1 - \mathrm{I}\langle d \mid q\rangle \\ &= \mathrm{P}(d)\frac{1-\mathrm{P}(q \mid d)}{1-\mathrm{P}(q)} \\ &= \mathrm{P}(d \mid \neg q) \end{aligned}$$

The details of this relation are not important except that it involves the conditional probability relating d and q, which earlier I identified as being illustrative of measuring the extra information needed to make an inconclusive inference go through. One way of looking at this analysis is that we have identified a concept of information that in the context of inference leads to a logical probability.

A subjective interpretation of LUP

There are other ways of proceeding. For example a subjective concept of probability is generated by betting odds. If the LUP is of universal applicability it should also capture that view of probability. Let us see if it does. I illustrate with an example taken directly from Ramsey (Mellor 1976, especially *circa* pp. 76–77). If d is a proposition about which we have a degree of belief say $p = m/n$, q might be a statement about which is the right way at a cross-roads. Let us suppose that arriving at the right destination has a reward r and arriving at the wrong destination a reward w. We, furthermore, assume that the total reward resulting from that decision will be

$$e = npr + n(1-p)w = nw + np(r-w)$$

Now let us assume that at some distance from the cross-roads there is a truthful source. The cost of consulting this source is $f(x)$ where x is the distance from the cross-roads. Knowing the way, the total reward will be

$$nr - nf(x)$$

Clearly deciding to consult the source is attractive provide

$$f(x) \le (r-w)(1-p)$$

in fact one will consult up to a distance d such that

$$f(d) = (r-w)(1-p).$$

This relationship defines the degree of belief, or probability

$$p = 1 - \frac{f(d)}{r-w}$$

Analysing this in terms of the LUP, without consultation we have an expectation, e, which involves an unknown uncertainty, that is, p. To determine this uncertainty we allow for information to be added, a measure of which is $f(x)$. There is a point d beyond which $f(d)$ would exceed the expected gain without consultation. The measure of uncertainty, or degree of belief in d is defined in terms of this extremal point. Or as Ramsey put it, "I propose therefore to use the distance I would be prepared to go to ask, as a measure of the confidence of my opinion".

This analogy should not be stretched too far but it seems clear that in defining the probability p in this way we again allow for the evidence to be added in so that the inference (which way to go) is certain. It is the cost of this information balanced against the expected gain that determines the probability of the statement d.

Conclusion

In this paper I have motivated the Logical Uncertainty Principle and interpreted it in two distinct ways. Firstly, through defining information axiomatically and showing how this notion of information leads to a measure of uncertainty for the inference deriving a query from a document. Secondly by showing a subjective interpretation of the Principle based on Ramsey's theory of probability.

Acknowledgments

This paper has benefited from enjoyable discussions with Gianni Amati and the close reading of Frank Ramsey's papers.

Notes

1. There is an extensive discussion of this approach to evidential reasoning in Schum's recent treatise on the subject.
2. There is a third way in which the retrieval process is rather complex, and possibly mysterious to the user, but which achieves its results through the ease with which feedback can be accomplished.
3. This case where the major premise A is true but the minor premise A $\supset$ B is uncertain or unkown is similar to the classical syllogism enthymemes. As Anderson and Belnap (1961) would put it, "Asked whether the enthymatic inference is valid (...), we may answer either 'No, your premises are simply insufficient for the conclusion' or, 'Yes, provided you mean to be using the obviously required premise [minor premise] (which we grant that you are, we being in a tolerant mood.
4. It is not obvious how best to interpret this *limit*, one can think of it as setting a lower limit to the information measure, or, another way is to think of it as $\vdash d \supset \mathrm{T}$, where T is the truth sentence.
5. This can also be formulated as $\vdash \neg(d \vee d') \supset q$ (with thanks to the referee).
6. I will abbreviate Logical Uncertainty Principle with LUP from here on.

References

Anderson AR and Belnap ND (1961) Enthymemes, The Journal of Philosophy, 713–723.

Crestani F, Lalmas M and Van Rijsbergen CJ, Eds. (1998) Information Retrieval: Uncertainty and Logics, Kluwer Academic Publishers, Dordrecht.

Eells E and Skyrms B (1994) Probability and conditionals: Belief revision and rational decision, CUP, Cambridge.
Hilpinen R (1970) On the information provided by observations. In: Hintikka J and Suppes P, Eds., Information and Inference, Reidel, Dordrecht, pp. 97–122.
Jeffrey RC (1983) The logic of decision, 2nd ed., University of Chicago Press, Chicago.
Keynes JM (1929) A treatise on Probability, Macmillan, London.
Lalmas M (1998) Logical models in information retrieval: introduction and overview. Information Processing and Management, 34:19–33.
Lindström S and Rabinowicz W (1992) The Ramsey test revisited. Theoria, LVII:131–182.
Mellor DH, Ed. (1976) Foundations: Essays in Philosophy, Logic, Mathematics and Economics: FP Ramsey, Routledge & Keegan Paul, London.
Ounis I (1998) Un modäle d'indexation relationnel pour les graphes conceptuels fondé sur une interprétation logique, Ph.D. thesis, Université Joseph Fourier, Grenoble.
Popper KR (1968) On rules of detachment and so-called inductive logic. In: Lakatos I, Ed., The problem of inductive logic, North Holland, Amsterdam, pp. 130–139.
Schum DA (1994) Evidential foundations of probabilistic reasoning, Wiley, New York.
Stalnaker R (1970) Probabilities and conditionals. Philosophy of Science, 37:64–80.
Van Rijsbergen CJ (1986) A non-classical logic for information retrieval. The Computer Journal, 29:481–485.
Van Rijsbergen CJ (1989) Towards an information logic. Proceedings of the Twelfth Annual International ACM:SIGIR Confernce on Research and Development in Information Retrieval, ACM Press, pp. 77–86.
Van Rijsbergen CJ (1992) Probabilistic retrieval revisited. The Computer Journal, 35:291–298.
Van Rijsbergen CJ and Lalmas M (1996) An information calculus for information retrieval. Journal of the American Society of Information Science, 47:385–398.
Waismann F (1930) Logische Analyse des Wahrscheinlichkeitsbegriffs. Erkenntnis, 1:228–248.