

ELECTRONIC WORKSHOPS IN COMPUTING

Series edited by Professor C.J. van Rijsbergen

Ian Ruthven (Ed)

Miro'95

Proceedings of the Final Workshop on Multimedia Information Retrieval
(Miro '95)
Glasgow, Scotland
18-20 September 1995

Paper:

A Note on Logic and Information Retrieval

Fabrizio Sebastiani

Published in collaboration with the
British Computer Society



©Copyright in this paper belongs to the author(s)

A Note on Logic and Information Retrieval

Fabrizio Sebastiani *

Istituto di Elaborazione dell'Informazione, Consiglio Nazionale delle Ricerche
Via S. Maria, 46 - 56126 Pisa (Italy) E-mail: fabrizio@iei.pi.cnr.it

Abstract

What is that makes a “good” logical model of IR? What are the guidelines that we should follow when we want to build one, and how much can we depart from these guidelines and still claim to have a *logical* model of IR? We have been motivated to write this note from our dissatisfaction with the fact that there seem to be many competing, incompatible views of what a logical model of IR should consist of; we think some of these views are misleading.

1 Information Retrieval and modelling

In recent years, researchers in Information Retrieval (IR) have devoted an increasing amount of work to the design of *models* of IR, i.e. of theoretical descriptions of the IR task that could serve both as specifications for building running systems, and as theoretical tools for abstractly investigating the relative effectiveness of systems built along their guidelines.

Modelling is fundamentally an activity of *abstraction*. A model is a description of a system that concentrates on the most important, architectural features of the system, and leaves out details that are believed not to be fundamental to the understanding of how the system works. For instance, a model of an IR system might contain a description of how documents and information needs are represented within the system, but leave out details on the data structures used to store these representations. Of course, more or less details can be left out from the description, depending on the purpose of the model. Different levels of abstraction can then be envisaged, and different models of the same system can then be produced, each at a different level of abstraction.

Abstraction results in *generalization* too, following the familiar principle according to which an abstract (or partial) description is equivalent to the class of concrete (or total) descriptions consistent with it: by leaving out non-fundamental details, the description actually becomes a model of a whole class of systems, namely those which differ from each other only by these details. A taxonomy of models can then be built that allows to classify IR systems and highlight their common premises and fundamental differences.

Traditionally, the activity of modelling is ascribed to two fundamental classes of motivations. Motivations of a *descriptive* nature reveal an intent to simply *acknowledge* “as is” the work of system builders, and explain the characteristics of a class of systems without being hampered by unimportant details, thus achieving clarity of exposition in the description of these systems. Motivations of a *prescriptive* (or *normative*) nature reveal instead an intent to *steer*, or influence, the work of system builders, dictating (or proposing) what the characteristics of a class of systems ought to be.

In the IR case (as in other subdisciplines of computer science), there is a third class of motivations that somehow escapes the previous, traditional classification, and that may be regarded as the primary driving force behind the activity of modelling: motivations of a *predictive* nature reveal an intent to predict the behaviour of a real system by running “abstract experiments” on an artificial, simplified setting. The

*This work has been carried out in the context of the project FERMI 8134 - “Formalization and Experimentation in the Retrieval of Multimedia Information”, funded by the European Community under the ESPRIT Basic Research scheme.

properties of the fundamentals of a system, or class of systems, can then be tested: the experimenter can thus be sure that the results so obtained are not influenced by the (supposedly non-fundamental) features that have been left out of the model.

Whatever the nature of the motivations for building models, however, it seems clear that, in order to achieve the breakthrough in efficiency and effectiveness that nowadays applications demand, a better understanding is needed of IR and of the nature of information, and that such an understanding may be achieved only by abstracting away from particular systems and techniques, thus concentrating on the study of the core, foundational principles underlying the IR endeavour. The recent interest that researchers have shown in the *Logical Model* of IR finds its roots in the belief that formal logic (and its companion discipline, formal semantics) is the discipline that provides the right tools for studying these foundational principles.

Surely, IR is not the first discipline within computer science in resorting to logic and methods for formal analysis. Researchers in other subfields of computer science, such as databases or programming languages, have gained much deeper insights in their discipline by analyzing it by means of logical techniques. In doing so, they have often uncovered deep-seated assumptions that had previously gone undetected, or inconsistencies that undermined the very foundations of the whole discipline, or problems previously unknown. Above all, they have gained a new perspective from able to take advantage of razor-sharp analytical tools whose effectiveness has been widely recognized by the scientific community as a whole throughout this century.

It might even be claimed that some disciplines, such as Artificial Intelligence (AI), have been totally revolutionized by the advent of a “logic-oriented” mentality. By the late 70’s, AI researchers had grown increasingly aware of the need for the methodological tools provided by mathematical logic for the construction of AI models of the domains of interest (see e.g. [20, 24]). Since then, the application of these methodologies has shaken the foundations of AI, and the theories that have withstood the test of logic have gained, as a consequence, strength and wider applicative impact, also arising interest in practitioners and theorists of other subfields of computer science (e.g. database theory) or of science as a whole (e.g. economic theory, analytic philosophy). A representative example of this is offered by AI “frame-based knowledge representation languages”. In the late 70’s there existed a plethora of them, each based on its own custom constructs, each basically incomprehensible to anyone apart from its own designers, who nonetheless remained adamant on the fact that their languages could express what logic could not. The final insuccess of these languages propted the very researchers that had first proposed them (e.g. [1]) to recast them in terms of mathematical logic, resulting in what are now called *description logics* (DLs) (see e.g. [2]). These have turned out to be one of the success stories of AI, to the point that AT&T relies on their inference engine to process billions of dollars’ worth of customers orders [25], and that they are now popular among computer scientists as diverse as database theorists [10] and factory-floor software engineers [13].

Although the history and underlying motivations of IR are profoundly different from those of AI, we think that IR too would benefit from the application of logical tools to the design and evaluation of its models, and our belief is indeed one of the fundamental premises that underlie our past work [27, 32]. The methodology in the use of logic that has been developed in AI and analytic philosophy is relevant also to IR applications, since logical tools are not committed to a specific application or domain to be modelled, and IR systems are indeed one of their possible applications or domains. Establishing logic as a common language between IR theorists and theorists of other subfields of computer science would also help IR in gaining better recognition from within these other disciplines, and in fostering cross-fertilization with them.

2 Incrementality and “modelness”

Information Retrieval, having been around almost since the invention of digital computers, has grown into a greatly complex discipline. A wealth of phenomena on which the relevance of documents to user information needs depends, have been highlighted by researchers, and a corresponding wealth of tools have been developed in the attempt to cope with them. Building realistic, worthwhile logical models of IR thus means modelling all these different phenomena, which makes for an extremely severe task.

In order to make worthwhile progress in this direction, we think that a logical model should be built *incrementally*. According to this line of thought, a relevant number of simplifying assumptions on what IR is should be made in the beginning, so as to allow the provisional design of a simple model of (an “idealized” version of) IR along methodological guidelines previously set. These simplifying assumptions should then be relaxed one by one, leading to the design of successive, more complex models that deal with a more and more complex and realistic picture of IR. Of course, care must be taken that the chosen methodology does ensure the relaxation of these assumptions.

This way of proceeding places then more importance on the requirement that the model of IR being designed satisfy well-formedness adequacy criteria well accepted in the field of formal modelling, than on the requirement that the coverage of the model be comprehensive, i.e. that the model deals with all the wealth of phenomena which are of interest to IR. In some sense, we are saying that models of IR (not only logical ones) can be judged by two different yardsticks: “*modelness*” (or “*logical well-formedness*”) and *coverage*.

An analysis of the logical models of IR proposed so far in the literature reveals, in our opinion, that coverage has been the primary goal to attain for most researchers who have engaged in IR modelling through logic, and that most of the models designed so far hardly satisfy the well-formedness criteria that are standard in applied logic. Instead, we think it is much more fruitful to first design a logically well-formed, albeit simple, model of IR and incrementally make it more respondent to the complex reality of IR, than to start with a logically flawed, although comprehensive, model of IR and incrementally eliminate the (possibly inherent) flaws. This position is not only ours, but underlies practically all work in applied logic. For instance, the problem of modelling logically the diagnosis of faulty circuits has first been solved for circuits with single faults, and only later the problem of multiple faults has been tackled [11]. This and other experiences in applied logic show in our opinion that modelness is to be considered a *sine qua non* condition from the start, and coverage a goal to attain incrementally. The purpose of this paper is then to take modelness issues at heart, and discuss methodologies and criteria for the design of logical models of IR. This will be the topic of Sections 3 to 6.

3 The logical model of Information Retrieval

Although interest in logic on the part of IR researchers may be traced back at least to the early 70’s (see e.g. [7]), the first clear statement that IR should be understood in logical terms is, to our knowledge, due to van Rijsbergen¹. In his 1986 papers [35, 36] and in a number of subsequent ones [37, 38], he proposed that the retrieval of document d as a result of information need n should be seen in terms of the logical formula $P(d \rightarrow n)$, where “ \rightarrow ” is the conditional connective formalized by a logic to be chosen and where $P(\alpha)$ is to be read as “the probability of α ”². Accordingly, the central problem of this way of looking at IR becomes that of selecting the “right” implication connective, i.e. selecting the logic whose implication connective best mirrors relevance: the “ideal” logic should be the one in which $P(d \rightarrow n)$ equals the probability that the document \mathbf{d} represented by d is relevant to the information need \mathbf{n} represented by n . To this respect, van Rijsbergen also argued that material implication (i.e. the implication connective formalized by classical logic) is not adequate for this task. He speculated that the answer might instead lie in the brand of implication formalized by some non-classical logics, also suggesting that the **C2** logic, from the tradition of “conditional” logics [33, 34], might be promising in this respect.

¹Darlington [9] has been suggested by some to be the first researcher to point out that IR could be conceived in terms of logic. However, an examination of his work shows that the notion of IR that he takes into account is more akin to “data retrieval” as in database systems, and he gives no suggestion as to how the ideas presented could be extended so as to take into account information retrieval as we understand it. For this reason, we will not consider Darlington’s work as relevant to our purposes; the same will happen, essentially for the same reasons, with the work of Frisch and Allen [15] and Patel-Schneider et al. [30]).

²In this paper we are taking a slight detour from the terminology used in van Rijsbergen’s papers, as we will take *queries* to be *representations of information needs*. Information needs and documents are the “concrete entities” of our domain, while what we actually deal with in logic is their *representations*; “noise” is introduced in the translation from information needs to their representations (queries) and from documents to their representations. As a consequence, the rightmost symbol in the implication is for us not “the representation of a query”, but “the representation of an information need” (or “a query”)

In such a model, the role of probability is a key one, as logical formulae are inherently imperfect representations of documents and information needs; the relevance of a document to an information need can thus be established only up to a limited degree of certainty. In the discussion that follows, however, we will ignore this issue, and make the simplifying assumption that “perfect” representations of documents and information needs can indeed be obtained, and that they are logical formulae of a simple non-probabilistic (e.g. propositional) language. This is an obviously unreasonable assumption once one attempts to build a *realistic* model of IR. It is however an assumption that will momentarily allow us to concentrate on discussing what characteristics a “good” formal model of IR should have by drawing our examples from a simple, non probabilistic logic. This is in keeping with the policy expoused in Section 2, as the very first “modelness” issues we need to clarify are those affecting very simple models of IR; only after these have been agreed upon may we tackle those concerning more complex ones. As a matter of fact, all the claims we make in this paper for the non-probabilistic case apply equally well once we plug in probability.

van Rijsbergen’s proposal is fascinating for three quite distinct reasons. The first, more obvious reason is that this model provides a bridge with database research, as the logical model of databases (see e.g. [31]) precisely establishes that a tuple t is to be returned as a result of query q iff $t \rightarrow q$.

The second reason is apparent once we consider, instead of the proof-theoretic, “symbol-crunching” level of logic, its model-theoretic, semantic level. In terms of this latter, the logical approach to IR amounts to sanctioning that relevance coincides with (set-)inclusion of information content, or semantics: only documents whose information content includes that of the information need are to be retrieved. Building effective IR models means then designing formal theories of the semantic (or information) content of documents, possibly taking into account context-dependence and other situational aspects of information [21].

The third reason springs from the proof-theoretic level of logic, and is related to the predictive intent of the modelling task discussed in Section 1. The adoption of the logical model means that, once the document and the information need are given a representation, “abstract experiments” can be performed by running a theorem prover for the logic in question and establishing whether $d \rightarrow n$; the logical model of IR is then a *computationally effective* model. The idea of “running abstract experiments” is especially interesting for IR, as it might produce good insights into phenomena and techniques of IR that are not yet well understood. In fact, experimental methodologies of the standard kind, while allowing to determine that certain techniques work better than others on a given experimental setting, do not always guarantee that the same results can be generalized to different experimental settings, and do not always help in understanding the reasons that lie at the root of such behaviour. The logical model, instead, might prove a better experimental tool in this respect. We might thus envisage, for instance, coding a particular IR technique in a given logic; once the fact that this is a faithful encoding of the technique has been established (by cross-checking the results of theorem proving, or proof checking, with those obtained experimentally), the semantics of the logical language could give insight into how and why such results are obtained, and give directions as to how the technique should be implemented in real systems. If two techniques, for some yet unknown reasons, do not fit together well in real IR systems, they can both be encoded in the model, and the reasons for their interaction can be investigated by inspecting the semantics of the language.

4 On the status of $d \rightarrow n$

van Rijsbergen’s proposal has proven very influential, to the extent that a number of papers have appeared in the literature tackling, in a variety of ways, the problem of modelling IR within the above-mentioned paradigm. However, a comparative analysis of these papers reveals that it is far from clear what the logical status of the $d \rightarrow n$ formula should be, in order to indicate that the document represented by d is relevant to the information need represented by n . In particular, authors seem to take different stands on which among the following facts should indicate this:

1. $d \rightarrow n$ is *true* in some particular interpretation of the chosen logic \mathcal{L} ;
2. d is a *logical consequence* of n in \mathcal{L} ;

3. $d \rightarrow n$ is *valid* in \mathcal{L} ;
4. n is *derivable* (or *provable*) from d in \mathcal{L} ;
5. $d \rightarrow n$ is a *theorem* of \mathcal{L} .

These notions mean substantially different things in logic, and it should be apparent that any *logical* model of IR (or of any other real-world phenomenon, for that matter) should take a clear stand on this issue if “modelness” is to be a key concern³. This variety of different approaches opens up questions such as: “Is there one should be based?”; “Is there more than one?”; “Are there ‘wrong’ ones?”. And a problem incumbent in the background is, of course, that if different notions are employed, *comparison* of these models may be difficult to achieve. Bruza and Huibers [3, page 112] have quite appropriately expressed their concern on this:

Research has not yet produced a powerful enough framework whereby information retrieval systems can be compared *inductively* instead of *experimentally*. A breakthrough in this area would mean that a theorem could be proven stating, for example, that vector space retrieval is more effective than Boolean retrieval. Such a result would not only spare us the efforts of experimentation, but more importantly, it would allow us to side step the controversies surrounding the experimental process.

Let us then try to analyse the different stands which have been taken. In his 1989 paper [37, page 85], van Rijsbergen seems to indicate that Option 4 (derivability) is the relevant notion to use:

The proposal of this paper is that all retrieval be based on a well-defined inference mechanism. This requires that objects and queries be given a formal semantics and that retrieval is expressed as a proof.

although also Option 5 (theoremhood) is no doubt consistent with this formulation. Bruza and van der Gaag [4, page 12] explicitly side for Option 4 (derivability):

In this approach, an information object is deemed relevant with respect to a searcher’s request if this request can be proven from the characterization of the object by employing a set of rules of *inference*.

Nie [29, page 478] has Option 1 (truth in a particular interpretation) in mind when he says:

For document D to be a “right” answer for query Q , it must “imply” the query, i.e. $D \rightarrow Q$. (...) The symbol \rightarrow does not signify the “material implication” as in classical logics. A counterexample for the material implication is that an empty document cannot imply a nonempty query (...).

as none among the other four options would result, in classical logics, in the behaviour he describes.

In their review paper on logical models of IR, Chiaramella and Chevallet [6, page 235] agree *de facto* with Nie in embracing Option 1 (truth in a particular interpretation), as their rebuttal of classical logic is based on the truth of material implication in a single interpretation:

³ We recall that: 1) a formula α is a *logical consequence* of a set of formulae Γ when α is true in all the interpretations in which all the formulae in Γ are true; 2) a formula α is *valid* when it is true in all interpretations, i.e. when it is a logical consequence of the empty set; 3) a formula α is *derivable* (or *provable*) from a set of formulae Γ when α can be obtained by applying the rules of inference to the axioms of the logic and the formulae in Γ ; 4) a formula α is a *theorem* when it can be obtained by applying the rules of inference to the axioms of the logic, i.e. when it is derivable from the empty set. While “derivability” and “theoremhood” are *syntactic* (or *proof-theoretic*) notions “truth”, “validity” and “logical consequentality” are *semantic* (or *model-theoretic*) notions. When the syntactic apparatus (i.e. the set of axioms and inference rules) of a logic is *weakly sound and complete* with respect to its semantics, theoremhood coincides with validity; when it is *strongly sound and complete*, we also obtain equivalence between derivability and logical consequentality.

If we consider the third line of this truth table, saying that when a predicate A is false and B is true, then A implies B is true has certainly no intuitive meaning: algebraic considerations somehow overrode common sense interpretations. We have left the domain of reality for the domain of mathematical abstraction.

In their model based on Description Logics (DLs) instead, Meghini *et al.* [27, page 298] embrace Option 2 (logical consequentality)⁴:

The terminological model then sees IR as the task of retrieving, as a response to a query C , all and only those documents i such that $(\text{sing } i) \preceq_{\Omega} C$, where Ω is a TL representation of the document base [and \preceq_{Ω} denotes *hybrid subsumption* between terms]. In other words, IR is the task of retrieving all those documents whose membership in the class denoted by C is a direct consequence of the truth of all the assertions and axioms of Ω .

What stand should one take then, given that we want to take “modelness” seriously and given that these options are in general not equivalent? For ease of exposition, we will make the simplifying assumption that our logic has an inferential apparatus which is strongly sound and complete with respect to its semantics. As we observed in Footnote 3, this will allow us to restrict our discussion to truth, validity and logical consequentality, and everything we will say about validity (resp. logical consequentality) will also apply to theoremhood (resp. derivability)⁵.

4.1 Truth

Let us see why truth is not a suitable notion on which to base a logical model of IR. The main reason is that, if we wanted to identify the relevance of the document represented by d to the information need represented by n with the truth of $d \rightarrow n$, we should also specify *in which interpretation* the truth of $d \rightarrow n$ has to be evaluated; a formula cannot be true *tout court*, simply because truth is not a *property* of formulae, but a *binary relation* between formulae and interpretations. We might naively answer that we should take “the interpretation that corresponds to the real world” (i.e. the interpretation in which the sentence “Glasgow is in Scotland” evaluates to true, “Birds are mammals” evaluates to false, and so on). But what is then the truth value to which “The number of water molecules in my glass is even” should evaluate? In the “interpretation that corresponds to the real world” this sentence surely has a truth value, i.e. that number either is even or is not. The answer is that we do not have a clue to what the “interpretation that corresponds to the real world” is, because our knowledge of the real world (or, more to the point, of our domain of discourse) is *partial*, and sometimes *fallacious* too, even in more mundane matters than those of molecular structure. And IR is no exception:

Partiality is an important feature of an IR system because it is common that it is unknown whether or not an item of information is contained in a document. So, assigning a truth-value to every existing formula that can be defined in the logic is meaningless. [21]

something we can pick out and stipulate to correspond to the real world, simply because we do not have a “direct grasp” (i.e. total and infallible knowledge) of the real world. Henceforth, truth is not simply difficult or impossible to compute; it is just not an object of computation, and to speak of “computing truth” is a bit of an epistemological oxymoron.

⁴In reference to this work, however, it should be noted that neither of the five notions discussed above applies to DLs, as DLs deal with terms rather than formulae, while all these notions apply to *formulae* only. However, a closer look at DLs reveals that hybrid subsumption is the analogue of logical consequentality for DLs. In [27] DLs were actually called *Terminological Logics* (TLs).

⁵Quite obviously, when proposing a particular logical model for IR we should check whether this assumption holds or not in the particular logic we have chosen. In fact, among the logics that have been proposed as foundations for a model of IR, some do not enjoy this property (e.g. propositional modal logic, proposed in [29]).

4.2 Validity

This argument, we think, illustrates that the notion of truth is unsuitable for modelling substantial fragments of reality; indeed, this is the reason why, in devising a logic, a logician is usually not interested in specifying the notion of truth *per se*, but rather in specifying truth as functional to the specification of the notions of validity and logical consequence⁶. notion of validity is defined in terms of truth? The key observation is that valid formulae are true (in any interpretation) in virtue of their *form*, and not in virtue of their *content*. Because of this, we do not need to have a grasp on the real world (i.e. to know which interpretation corresponds to it) to assess the validity of a formula (in the IR case: to assess whether we should retrieve the document or not); we only need to perform a purely symbolic check of the formula itself. For instance, the formula of propositional logic

$$\text{John-likes-football} \vee \neg \text{John-likes-football} \quad (1)$$

is true in any interpretation (i.e. valid). In order to assess its validity it is not necessary to know whether in the interpretation that corresponds to the real world (or in any other particular interpretation, for that matter) John actually likes football or not; it is instead sufficient to apply the well-known syntactic rules for validity checking in propositional logic. This formula remains valid even if we substitute any other propositional formula to the two occurrences of `John-likes-football`, which indicates that content has really nothing to do with validity. The formula

$$\text{John-is-a-man} \rightarrow \text{John-likes-football} \quad (2)$$

instead, may well be true in the interpretation that corresponds to the real world, but is false in others, hence is logically not very interesting.

One might argue that (2) is at least an *informative* formula (“it says *something*”), and that valid formulae like (1), being tautologous, carry no information content. The notion of informativeness that is being hinted at here is the one by Carnap and Bar-Hillel [5] (henceforth “CBH-informativeness”), according to which a formula is the more informative the more countermodels it has (i.e. the more interpretations falsify it)⁷; valid formulae are then minimally CBH-informative, as they have no countermodels. But in the IR case it is not the formula $d \rightarrow n$ that carries information to us; *it is the very fact that it is (or that it is not) valid*, as it informs us whether we should retrieve the document or not. For instance, if we cast the Boolean model of IR in terms of the validity of formula $d \rightarrow n$ in propositional logic, the fact that the formula $(p_1 \wedge p_2 \wedge p_3) \rightarrow (p_1 \vee p_2)$ is valid informs us that a document indexed by terms p_1 , p_2 and p_3 should be retrieved as a result of the information need represented by $p_1 \vee p_2$; and the fact that the formula $(p_3 \wedge p_4 \wedge p_5) \rightarrow (p_1 \vee p_2)$ is not valid informs us that a document indexed by terms p_3 , p_4 and p_5 should not be retrieved as a result of the same information need. In other words, although the valid propositional formula $(p_1 \wedge p_2 \wedge p_3) \rightarrow (p_1 \vee p_2)$ is CBH-uninformative, the meta-formula $VALID((p_1 \wedge p_2 \wedge p_3) \rightarrow (p_1 \vee p_2))$ (where $VALID$ is a meta-predicate symbol) is CBH-informative, as is also the meta-formula $\neg VALID((p_3 \wedge p_4 \wedge p_5) \rightarrow (p_1 \vee p_2))$.

This notion of “meta-level informativeness” is, of course, *task-oriented*, and the task here is IR. In fact, we are not interested in the validity or non-validity of formulae which are not of type $d \rightarrow n$; thus, a meta-formula such as $VALID(p_1 \vee \neg p_1)$ has no information content to us, as it informs us of the validity of a formula that is not meant to represent relevance of documents to information needs.

In sum, the interesting fact is that adopting a form-based notion as validity is, rather than a content-based notion as truth is, *allows us to effectively reason about information content*.

4.2.1 Truth, validity and the “false document problem”

A further hint that validity, rather than truth, is a better way to go comes from the observation that, once we base a model of IR on validity, the model does not suffer from what has been called the *false*

⁶Many logicians take a logic to be the set of its valid formulae.

⁷We would like to thank Gianni Amati for pointing this out to us.

document problem (see e.g. [6, page 240]). According to the supporters of Option 1 (truth in a particular interpretation), propositional logic, and classical logic in general, suffers from the problem that a so-called “false document” (i.e. a hypothetical document which is “about nothing”—we will rather call it a *totally uninteresting document* (TUD)) is deemed relevant to any information need, and this is obviously unsuitable.

In fact, suppose our term language consists of the set of propositional letters $P = \{p_1, p_2, p_3, p_4\}$, so that a TUD is represented by the formula $\neg p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4$; quite obviously, in propositional logic the formula $(\neg p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4) \rightarrow \alpha$ is true for any propositional formula α in *at least one interpretation*, i.e. in the interpretation that makes all propositional letters in P true.

But this does not mean that propositional logic is unsuitable for modelling IR: it just means that a model based on truth is. In fact, it is easy to show that, if Option 3 (validity) is adopted, no “false document problem” obtains, and all and only those documents whose representation has a greater “information content” than the representation of the information need are retrieved. For those documents for which this is not the case (as in the case of a TUD), $d \rightarrow n$ will not be valid, i.e. there will be at least one assignment of truth values to the propositional letters for which $d \rightarrow n$ will be false. For these documents, it may well be that, among the assignments that make $d \rightarrow n$ true, there is some assignment that makes d false and n true, and hence makes the implication $d \rightarrow n$ *true*; but this need not worry us, as long as the implication is not *valid*. This situation will be better illustrated by means of an example.

Example 1 *Let us define a model of IR based on the validity of $d \rightarrow n$ in propositional logic, in the following way. Let us suppose each document is represented by a conjunction of literals (a literal being either a propositional letter or its negation) drawn from an alphabet $P = \{p_1, p_2, p_3, p_4\}$. Relying on propositional logic allows us to express the following facts:*

1. *to explicitly say that d is about p_i ; this is obtained by including p_i as a conjunct in d 's representation;*
2. *to explicitly say that d is not about p_i ; this is obtained by including $\neg p_i$ as a conjunct in d 's representation;*
3. *to take no commitment as to whether d is or is not about p_i ; this is obtained by including neither p_i nor $\neg p_i$ as conjuncts of d 's representation⁸.*

Note that, in this representation, a TUD is represented by the formula $\neg p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4$.

Let us also suppose that each information need is represented by a formula of propositional logic built out of alphabet P . The three types of facts that can be stated in representing documents can also be stated in representing information needs: for instance, if we represent an information need by formula $p_1 \vee \neg p_2$, we mean that the system should retrieve all and only those documents that are either about p_1 or are not about p_2 , regardless of whether they are also about p_3 and/or p_4 . Figure 1 shows validity-based relevance assessments for the three sample documents

$$\begin{aligned} d_1 &= p_1 \wedge p_2 \wedge \neg p_3 \wedge \neg p_4 \\ d_2 &= \neg p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4 \\ d_3 &= p_1 \wedge p_2 \wedge \neg p_4 \end{aligned}$$

with respect to information need

$$n = p_1 \vee (p_2 \wedge \neg p_3)$$

Note that document d_2 is a TUD. As can be gathered from inspection of Figure 1, formulae $d_1 \rightarrow n$ and $d_3 \rightarrow n$ are valid (i.e. the corresponding columns have all T's); this means that the documents represented by d_1 and d_3 are deemed relevant to the information need represented by n , as should indeed happen. The formula $d_2 \rightarrow n$, instead, is not valid (i.e. there is an F at row 16 in the corresponding column); this means

⁸In many approaches to Boolean retrieval, including neither p_i nor $\neg p_i$ as conjuncts of d 's representation is taken to mean that d is not about p_i . This is called the *closed world assumption* (CWA); our arguments would follow similar lines even in this case, as the effect of the CWA is to say implicitly what here is specified explicitly.

| | p_1 | p_2 | p_3 | p_4 | d_1 | d_2 | d_3 | n | $d_1 \rightarrow n$ | $d_2 \rightarrow n$ | $d_3 \rightarrow n$ |
|----|-------|-------|-------|-------|-------|-------|-------|-----|---------------------|---------------------|---------------------|
| 1 | T | T | T | T | F | F | F | T | T | T | T |
| 2 | T | T | T | F | F | F | T | T | T | T | T |
| 3 | T | T | F | T | F | F | F | T | T | T | T |
| 4 | T | T | F | F | T | F | T | T | T | T | T |
| 5 | T | F | T | T | F | F | F | T | T | T | T |
| 6 | T | F | T | F | F | F | F | T | T | T | T |
| 7 | T | F | F | T | F | F | F | T | T | T | T |
| 8 | T | F | F | F | F | F | F | T | T | T | T |
| 9 | F | T | T | T | F | F | F | F | T | T | T |
| 10 | F | T | T | F | F | F | F | F | T | T | T |
| 11 | F | T | F | T | F | F | F | T | T | T | T |
| 12 | F | T | F | F | F | F | F | T | T | T | T |
| 13 | F | F | T | T | F | F | F | F | T | T | T |
| 14 | F | F | T | F | F | F | F | F | T | T | T |
| 15 | F | F | F | T | F | F | F | F | T | T | T |
| 16 | F | F | F | F | F | T | F | F | T | F | T |

Figure 1: Validity-based relevance assessments

that, contrary to what might happen in truth-based approaches to relevance, the TUD represented by d_2 is correctly deemed not relevant to the information need represented by n .

It might be interesting to note that there are indeed formulae representing information needs to which a TUD would be deemed relevant: these are precisely all conjunctions of negated propositional letters that can be built by using alphabet P (and all formulae logically equivalent to them). For instance, the TUD represented by d_2 would be deemed relevant to the information need represented by $\neg p_1 \wedge \neg p_2 \wedge \neg p_3$. But this information need is at least as “pathological” as a TUD is a “pathological” document: and it is perfectly reasonable that a “document about nothing” should be deemed relevant to an “information need about nothing”. ■

Note also that this argument, apart from adding further negative evidence on Option 1 (truth in a particular interpretation) as a suitable notion on which to base a model of IR, has the practical consequence of invalidating a number of arguments (see Section 4) that were made against the suitability of classical logic for IR modelling purposes, and that were essentially based on a truth-based view of IR modelling. Therefore, we feel that the use of classical logics in IR has been too hastily dismissed; we will come back to this point in Section 5.

4.3 Logical consequentality

Let us now discuss the case of logical consequentality. The first thing we should point out is that in Example 1 the same results we have obtained with validity would have been obtained by adopting logical consequentality. This is due to the use of propositional logic as the basis of the simple model adopted in the example: propositional logic is in fact strongly sound and complete with respect to its standard denotational semantics, and this ensures (see Footnote 3) that, for any formulae α and β , $\alpha \rightarrow \beta$ is valid if and only if β is a logical consequence of α .

However, given that we are seeking a general (i.e. independent of the chosen logic) characterization of what a logical model of IR should consist of, it is necessary to assess which is the notion of choice in case the two are not equivalent.

To a first approximation both are by and large suitable, as both are “form-based” notions, rather than “content-based” as truth is: while truth may be checked only with reference to a specific interpretation, validity and logical consequentality may be established *formally* (and thus effectively), by symbolic manipulation only. These two notions are the cornerstones of logic exactly because logic is concerned with *partial and possibly fallacious* knowledge of the world, and both notions specify exactly those inferences which are compatible with this knowledge, however partial and fallacious it may be. For instance, in Example 1 we would like document d_3 to be deemed relevant to information need n : by relying on e.g. logical consequence

this indeed happens, because n is compatible with the partial knowledge we have of document d_3 (partial as it does not say anything on d_3 being about p_3 or not) that is expressed by the formula $p_1 \wedge p_2 \wedge \neg p_4$.

Forgetting about truth, let us now try to assess the relative merits of validity and logical consequentality. One aspect on which validity scores better is that logical consequentality would seem a bit unsuitable given that what we are really interested in is extending the “ $d \rightarrow n$ model” to a “ $P(d \rightarrow n) = r$ model” (see Section 3), i.e. to a model in which we can speak of the *probability* of relevance in terms of a formula $P(d \rightarrow n) = r$ to be read “the probability that d implies n is r ”⁹. Given a suitable logic (e.g. the one discussed in [18]), we can express such formulae as $P(d \rightarrow n) = r$, and define appropriate notions of truth and validity for them. However, in no logic that we know of (although it might perhaps in principle be possible) we can formalize the sentence “the probability that n is a logical consequence of d is r ”, and define (meta-level) truth and validity conditions for it; in fact, in any logic that we know of either n is a logical consequence of d or it is not (i.e. the probability of this fact, to be expressed in the metalanguage, would always be either 1 or 0), and we can hardly imagine a reasonable set of events on which such a probability distribution might be defined.

One aspect on which logical consequentality gets instead a better mark is its more intuitive character. It is quite intuitive, in fact, that relevance of a document to an information need is a *consequence* of the semantic content of the document and of the information need, and possibly of other factors such as the meaning (as specified e.g. in a thesaurus) of the terms involved; and it is quite intuitive that relevance should not depend on the fact that the particular logic adopted has an implication connective in

To sum up our argument we may then say that at a first approximation both notions are suitable (whereas truth is not), and which is best is an issue open to debate and which requires a finer grain of analysis, as each of them has advantages and disadvantages for IR¹⁰.

5 IR and the paradoxes of material implication

Since the very introduction of the logical model of IR, researchers seem to have maintained as a cornerstone of their investigations that classical logic is not adequate for IR modelling. However, in Section 4 we have seen that a number of arguments that had been used against employing classical logic for IR modelling are substantially invalid, as they are based on a misuse of logic in modelling real-world phenomena. It is our contention that also other arguments that have tried to counter classical logic are far from conclusive, and that we should perhaps come to the conclusion that the use of classical logic in IR has been dismissed too hastily¹¹. One such argument is the following.

A number of researchers have recalled that material implication suffers from idiosyncratic behaviour, resulting in what are known as “the paradoxes of material implication” (see e.g. [12] and [17, page 37]), and have implied that this renders material implication an unsuitable starting point in the attempt to model relevance of documents to information needs¹². For instance, it is well-known that the following schemata¹³, all of which have a distinctively counter-intuitive flavour once “ \rightarrow ” is interpreted as “if ... then ...”, are valid in classical propositional logic:

⁹See [32] for a discussion on how this probability might be interpreted.

¹⁰Not surprisingly, in the logical approach to artificial intelligence both approaches are adopted from time to time, with a tendency to adopt logical consequentality in the case of classical logics and validity in the case of those “esoteric” logics which do not enjoy strong soundness and completeness [22].

¹¹Note that classical logic has been criticised on two accounts: 1. it cannot deal with uncertainty, which must be dealt with because representations of documents and information needs are inherently imperfect; 2. even if perfect representations of documents and information needs could be obtained and were formulae of classical logic, it would not be adequate. It is this latter criticism we do not completely agree with.

¹²Actually, none among the papers quoted mentions the “paradoxes” of material implication explicitly; it is however clear that those are the phenomena referred to. We write the word “paradox” in quotes because, in the case we are discussing, it is actually a misnomer: the “paradoxes of material implication” are not cases of *inconsistent* behaviour (as e.g. in the paradox of the Liar), but rather of *counter-intuitive* behaviour.

¹³A *schema* is an expression which stands for the class of formulae of the logic that can be obtained by uniformly substituting formulae of the logic for metavariables; for instance, the schema of propositional logic $\alpha \vee \neg\alpha$ stands for the set of formulae of propositional logic containing e.g. the formula $p \vee \neg p$ and the formula $(p \rightarrow q) \vee \neg(p \rightarrow q)$.

$$\alpha \rightarrow (\beta \rightarrow \alpha) \tag{3}$$

$$\neg\alpha \rightarrow (\alpha \rightarrow \beta) \tag{4}$$

$$(\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha) \tag{5}$$

However, if relevance of documents to information needs is to be modelled by the validity of formula $d \rightarrow n$, our interest as IR theorists in an implication connective is exclusively in terms of its behaviour in the context of a formula $d \rightarrow n$ in which *neither d nor n contain occurrences of the “ \rightarrow ” symbol*. Of course, the representation of an information need may well be a formula $p_1 \rightarrow p_2$. But this need not concern us, as we do not understand *this occurrence of “ \rightarrow ”* as modelling relevance; in fact, the counter-intuitive character of schemata (3)–(5) critically depends on the interpretation of *all* the occurrences of “ \rightarrow ” as “if ... then ...”. For instance, if we equivalently write (4) as $\gamma \rightarrow (\gamma \vee \beta)$ (where $\gamma = \neg\alpha$), this says something which, far from being paradoxical, can be absolutely subscribed to, namely that a document about γ is to be retrieved as a consequence of an information need about γ or something else.

Hence it would seem that the only “paradoxes” that may be of interest to us are those in which “ \rightarrow ” occurs at a level of nesting equal to 1, and that we can instead disregard the others; and, to the credit of classical logic, no “paradoxes” of nesting equal to 1 have ever been pointed out (unless either the consequent or the antecedent are themselves either valid or contradictory, a case which we discuss later).

Actually, that we can disregard schemata (3)–(5) is true only in part. In fact in classical propositional logic, as in any logic in which *modus ponens* is a rule of inference¹⁴, whenever $\alpha \rightarrow \beta$ is valid, if α is valid also β is valid. This means that schemata (3) and (4) *are* of interest to us, because they reflect the behaviour of the logic once either the information need and/or the document are *themselves* represented by valid or contradictory formulae; luckily enough, these cases turn out to be harmless, as the following discussion shows.

In IR terms, Schema (3) is pertinent to the case of information needs represented by valid formulae. In fact, from (3) and *modus ponens* we have that if n is a valid formula then $d \rightarrow n$ is also valid: in IR terms this means that *any* document will be deemed relevant to an information need represented by a valid formula. But note that such an information need is represented by formulae such as $p_1 \vee \neg p_1$ (or its logical equivalents), which corresponds to a request to retrieve “all documents that either are about p_1 or are not about p_1 ”: it is then perfectly reasonable that all documents from the document base should be selected for retrieval, as a query such as $p_1 \vee \neg p_1$ asks exactly for this. It is also clear from this example that valid formulae are representations of “pathological” information needs: any normal information need will be represented by a formula that is neither valid nor unsatisfiable, but true in some proper subset of the set of interpretations of propositional logic.

Schema (4) is instead pertinent to the case of documents represented by unsatisfiable formulae. In fact, from (4) and from *modus ponens* we have that, if $\neg d$ is a valid formula (i.e. if d is unsatisfiable), then $d \rightarrow n$ is also valid: in IR terms this means that a document represented by an unsatisfiable formula will be deemed relevant to *any* information need. But note that such a document is represented by formulae such as $p_1 \wedge \neg p_1$ (or their logical equivalents), which asserts that the document “at the same time is about p_1 and is not about p_1 ”. It is also clear from this example that unsatisfiable formulae are representations of “pathological” documents: as for information needs, any “normal” document will be represented by a formula that is neither valid nor unsatisfiable, but satisfied in some subset of the set of interpretations of propositional logic.

In summary, only a few among the paradoxes of material implication are pertinent to IR modelling, depending on the level of nesting of the “ \rightarrow ” connective; we have shown that the few pertinent ones do not rule out the use of classical propositional logic for IR modelling, as 1) they affect the behaviour of the model only when “pathological” documents and information needs are considered, and 2) even in these cases they arguably do not affect it in an unreasonable way.

¹⁴Not all logics have this property: see e.g. [14].

6 The role of denotational semantics in IR modelling

In Section 4 we have argued that either validity or logical consequentality should be the central notions in any logical model of IR. The denotational semantics of the representation language used in the models under scrutiny has played a central role in our argument¹⁵. In fact, it is by analysing the semantics (i.e. the truth conditions) of the formulae involved in these models that we have been able to fully understand the different impact on IR of the five notions discussed in the previous sections. Had we relied on syntactic and proof-theoretic notions only, we would have been able to perform this analysis only with greater difficulty.

Denotational semantics is extremely important in logical analysis, as it provides a frame of reference for analysing the meaning of the logical machinery used to formalise the application domain of interest. This frame of reference is so outstanding in terms of clarity and simplicity that it has gained the status of a “standard”, both within the logic community [17] and the programming languages community [16]; it thus provides a means of analysing, comparing, designing (and sometimes implementing too) logical and programming languages that constitutes a *de facto* “interlingua” among designer, implementor, user and critic. Its importance in the analysis of logical models of IR has been recognized ever since the very first logical model of IR was proposed: van Rijsbergen [35, 36] makes an explicit reference to “possible worlds semantics” (PWS), a well-known type of denotational semantics, and most proposed logical models of IR appeared since then (e.g. [8, 26, 27, 29, 32]) rely heavily on notions from denotational semantics. No agreement among these researchers on the role that denotational semantics should play in IR modelling; we might go as far as saying that there seems to be no agreement as to what denotational semantics really is. Given that its importance also lies in its being a “standard”, it is very important that consensus can be reached among IR theorists on how to use it.

To this respect we would like to note that while denotational semantics is, as we said above, a frame of reference for analysing the language of a given logic (i.e. the meaning of the connectives, operators and of the other primitives of the language), it is by no means part of this language itself. The entities that populate the world of denotational semantics (e.g. possible worlds, individuals, accessibility relations, etc.) are nothing else than (immaterial) “ideas” with reference to which one may explain, or support one’s intuitions about, the meaning of the (absolutely) language under consideration.

For instance, the notion of “possible world” and that of “accessibility between possible worlds” are useful clues to understanding the nature of the necessity operator of modal logic or the implication connective of conditional logic. However, it is of key importance to recognize that these entities are not themselves data structures open to direct manipulation. The idea that they indeed are, and that they can thus have an active part in the IR process, lurks behind a number of papers dealing with the logical modelling of IR (e.g. [8, 29]). In particular, the fact that the semantics of modal and conditional logics relies on “possible worlds” being grouped into graph-like structures (called *Kripke structures*) has led to the definition of models in which the theorists themselves build these networks by direct manipulation (e.g. by assigning weights to the nodes and to the “accessibility” edges of the Kripke structure). Denotational semantics stipulates instead that the characteristics of semantic structures (e.g. the weights mentioned above) are to be determined only by the logical formulae that appear in one’s representation of the relevant knowledge, and are not themselves open to direct manipulation. The only way to make these structures have certain given characteristics is to *induce* these characteristics by introducing in one’s representation formulae that constrain the semantic structures to have exactly those characteristics¹⁶.

It is therefore our contention that those models that do not comply with this, although they may well be

¹⁵We recall that *denotational semantics* (also known as *model-theoretic* or *Tarskian semantics*) is the standard way of formally specifying the meaning of logical languages. Such a specification is accomplished by postulating the existence of a number of “ways the world could be” (*interpretations*), and of systematically specifying in which of these interpretations the expressions of the language are true. Inference is then defined as the derivation of only those formulae that are true in all the interpretations in which the premises are also true. In the case of propositional logic, discussed in the previous sections, interpretations are usually called *truth-value assignments*.

¹⁶For example, in so-called “normal” modal logics (i.e. supersets of the \mathcal{K} modal logic) reflexivity of Kripke structures is achieved not by adding self-loops to all nodes of these structures, but is induced by introducing in one’s representation all formulae of type $\Box\alpha \rightarrow \alpha$.

interesting models of IR, can hardly lay claim of being related in any significant way to the logic they are supposedly using, and to logic.

It may be interesting to understand why models of IR that do not comply with this criterion have been so popular. We plan to tackle this issue in an extended version of this paper. It turns out that the reason for this is also the reason for the popularity of the truth-based models discussed in Section 4.1; the critical issue here is, again, partiality.

7 Conclusion

We have discussed a number of issues related to the well-formedness of logical models of IR. In particular, we have argued that notions such as validity or logical consequentality, rather than truth, should be the basic building blocks of a logical model of IR. We have also argued that the pertinence to IR of the paradoxes of material implication has been overestimated. The rehabilitation of classical logic as a building block for modelling IR is a direct consequence of these arguments. We have also discussed the role of denotational semantics in IR modelling, and argued for a more faithful compliance with its standard methodology.

We hope this study may play a role in better understanding the potential of the logical approach to IR.

Acknowledgements

I would like to thank Gianni Amati, Iain Campbell, Fabio Crestani, Carlo Meghini, Jian-Yun Nie, Alessandro Saffiotti, Ulrich Thiel and Keith van Rijsbergen for discussing with me the topics of this paper. Thanks also to an anonymous referee that prompted me to clarify my points.

References

- [1] Ronald J. Brachman. A structural paradigm for representing knowledge. Technical Report 3605, Bolt Beranek and Newman, Cambridge, MA, 1978.
- [2] Ronald J. Brachman and Hector J. Levesque. The tractability of subsumption in frame-based description languages. In *Proceedings of AAAI-84, 4th Conference of the American Association for Artificial Intelligence*, pages 34–37, Austin, TX, 1984. [a] An extended version appears as [23].
- [3] Peter D. Bruza and Theo W. Huibers. Investigating aboutness axioms using information fields. In *Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval*, pages 112–121, Dublin, IRL, 1994.
- [4] Peter D. Bruza and Linda C. van der Gaag. Efficient context-sensitive plausible inference for information disclosure. In *Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval*, pages 12–21, Pittsburgh, PA, 1993.
- [5] Rudolf Carnap and Yehoshua Bar-Hillel. Semantic information. *British Journal for the Philosophy of Science*, 4:147–157, 1953.
- [6] Yves Chiaramella and Jean Pierre Chevallet. About retrieval models and logic. *The Computer Journal*, 35:233–242, 1992.
- [7] William S. Cooper. A definition of relevance for information retrieval. *Information storage and retrieval*, 7:19–37, 1971.
- [8] Fabio Crestani and Cornelis J. van Rijsbergen. Probability kinematics in information retrieval. In *Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval*, pages 291–299, Seattle, WA, 1995.

- [9] J.L. Darlington. Theorem proving and information retrieval. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 4*, pages 173–181. Edinburgh University Press, Edinburgh, UK, 1969.
- [10] Giuseppe De Giacomo and Maurizio Lenzerini. Making CATS out of kittens: description logics with aggregates. In *Proceedings of DL-95, 4th International Workshop on Description Logics*, pages 143–147, Roma, Italy, 1995.
- [11] Johan de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.
- [12] H.C. de Swart and Rob P. Nederpelt. Implication. a survey of the different logical analyses of “if . . . then . . .”. *Nieuw Archief voor Wiskunde*, 10:77–104, 1992.
- [13] P. Devanbu and M. Jones. The use of description logics in KBSE systems. In *Proceedings of the 17th International Conference on Software Engineering*, Sorrento, Italy, 1994.
- [14] J. Michael Dunn. Relevance logic and entailment. In Dov M. Gabbay and Franz Guenther, editors, *Handbook of Philosophical Logic*, volume 3, pages 117–224. Reidel, Dordrecht, NL, 1986.
- [15] Alan M. Frisch and James F. Allen. Knowledge retrieval as limited inference. In *Proceedings of CADE-6, 6th International Conference on Automated Deduction*, New York, NY, 1982.
- [16] Michael J. Gordon. *The denotational description of programming languages*. Springer, Heidelberg, FRG, 1979.
- [17] Susan Haack. *Philosophy of logics*. Cambridge University Press, Cambridge, UK, 1978.
- [18] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.
- [19] William L. Harper, Robert Stalnaker, and Glenn Pearce, editors. *Ifs. Conditionals, belief, decision, chance and time*. Reidel, Dordrecht, NL, 1981.
- [20] Patrick J. Hayes. In defence of logic. In *Proceedings of IJCAI-77, 5th International Joint Conference on Artificial Intelligence*, pages 559–565, Cambridge, MA, 1977.
- [21] Mounia Lalmas and Cornelis J. van Rijsbergen. 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, 1993.
- [22] Hector J. Levesque. A logic of implicit and explicit belief. In *Proceedings of AAAI-84, 4th Conference of the American Association for Artificial Intelligence*, pages 198–202, Austin, TX, 1984.
- [23] Hector J. Levesque and Ronald J. Brachman. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93, 1987.
- [24] Drew McDermott. Tarskian semantics, or: no notation without denotation! *Cognitive Science*, 2:272–282, 1978.
- [25] Deborah L. McGuinness and Lori Alperin Resnick. Description-logic based configuration for consumers. In *Proceedings of DL-95, 4th International Workshop on Description Logics*, pages 109–111, Roma, Italy, 1995.
- [26] Carlo Meghini. An image retrieval model based on classical logic. In *Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval*, pages 300–308, Seattle, WA, 1995.

- [27] Carlo Meghini, Fabrizio Sebastiani, Umberto Straccia, and Costantino Thanos. A model of information retrieval based on a terminological logic. In *Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval*, pages 298–307, Pittsburgh, PA, 1993. Published by ACM Press, Baltimore, MD.
- [28] John Mylopoulos and Michael L. Brodie, editors. *Readings in artificial intelligence and databases*. Morgan Kaufmann, San Mateo, CA, 1989.
- [29] Jian-Yun Nie. An information retrieval model based on modal logic. *Information processing and management*, 25:477–491, 1989.
- [30] Peter F. Patel-Schneider, Ronald J. Brachman, and Hector J. Levesque. ARGON: knowledge representation meets information retrieval. In *Proceedings of CAIA-84, 1st IEEE Conference on Artificial Intelligence Applications*, Denver, CO, 1984.
- [31] Raymond Reiter. Towards a logical reconstruction of relational database theory. In Michael L. Brodie, John Mylopoulos, and Joachim W. Schmidt, editors, *On conceptual modelling*, pages 191–233. Springer, Heidelberg, FRG, 1984. [a] Also reprinted in [28], pp. 301–326.
- [32] Fabrizio Sebastiani. A probabilistic terminological logic for modelling information retrieval. In *Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval*, pages 122–130, Dublin, IRL, 1994. Published by Springer Verlag, Heidelberg, FRG.
- [33] Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in logical theory*, pages 98–112. Basil Blackwell, Oxford, 1968. [a] Also reprinted in [19], pp. 41–55.
- [34] Robert C. Stalnaker and Richmond H. Thomason. A semantical analysis of conditional logic. *Theoria*, 36:23–42, 1970.
- [35] Cornelis J. van Rijsbergen. A new theoretical framework for information retrieval. In *Proceedings of SIGIR-86, 9th ACM International Conference on Research and Development in Information Retrieval*, pages 194–200, Pisa, Italy, 1986.
- [36] Cornelis J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29:481–485, 1986.
- [37] Cornelis J. van Rijsbergen. Towards an information logic. In *Proceedings of SIGIR-89, 12th ACM International Conference on Research and Development in Information Retrieval*, pages 77–86, Cambridge, MA, 1989.
- [38] Cornelis J. van Rijsbergen. Probabilistic retrieval revisited. *The Computer Journal*, 35:291–298, 1992.