Designing effective retrieval engines for multimedia document repositories*

Carlo Meghini, Fabrizio Sebastiani and Umberto Straccia
Istituto di Elaborazione dell'Informazione
Consiglio Nazionale delle Ricerche
Via S. Maria, 46 - 56126 Pisa (Italy)
E-mail: $\langle lastname \rangle$@iei.pi.cnr.it

1 Introduction

A principled approach to the design of a retrieval engine for multimedia document (MD) repositories starts from the identification of a suitable *retrieval model*, i.e., of a formal specification of the three basic entities of retrieval: documents, users' information needs, and the matching function, which assigns a set of documents to each information need. Modelling the retrieval of MDs requires taking into account (at least) the following three orthogonal dimensions: *form*, i.e. the structural component of a document; *content*, i.e. the meaning of a document; *uncertainty*, i.e. the imprecision in the system's estimation of the relevance of a document to a user information need. When matching documents at the form level, uncertainty affects the system's evaluation of the structural similarity between documents and queries; at the content level, it affects the system's evaluation of the overlap between their information content. A multimedia retrieval model should thus involve a combination of concepts and techniques from *the world of digital signal processing*, which contributes notions for representing the form of documents, and algorithms for assessing the similarity between them and *the world of symbolic processing*, which contributes conceptual models of reality for representing document contents (and related knowledge), and algorithms for reasoning about them in a way that captures relevance. A strong interaction between these two souls is a crucial factor for the adequacy of a multimedia retrieval model. Sadly, this is largely unaccomplished nowadays: current image retrieval models, for instance, either decidedly adhere to one of these two paradigms or can be decomposed into two independent sub-models each belonging to either paradigm. In order to accomplish a full integration, the signal (form) and symbolic (content) dimensions need to be put in relation with each other by the model, so that features pertaining to form can be addressed *from within the same expressions used to address document content*. In the rest of this paper we will briefly illustrate fragments of a model which we are incrementally developing, and which tries to capture these aspects in a unified, well-founded framework.

2 Representing document content and relevance

Following [3], the kernel of our model is based on a *Description Logic* (DL), namely the logic $\mathcal{ALCO}$. Accordingly, the information retrieval process may be viewed as deciding whether, given a KB (i.e. a set of assertions and definitions) Σ containing document representations and thesaural entries, a concept C representing an information need and an individual a uniquely identifying a document, $C(a)$ is a *logical consequence* of Σ (written $\Sigma \models C(a)$).

One of the main concerns of our work has been addressing *relevance*, as the task of information retrieval is defined precisely as that of finding all the documents d that are *relevant* to a given information need

*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.

n. The notion of *tautological entailment*, from the literature on relevance logics, is based on the idea of enforcing a tighter connection in meaning between the premises and the conclusion of any inference licensed by the underlying logic. The obvious connections between this idea and what one wants to capture in document retrieval have led us to incorporate tautological entailment in our DL. Technically, we have done this by switching from a two- to a *four-valued semantics*. Logics of this kind have already been used in KR, and have shown a better computational behaviour than their two-valued analogues (see e.g. [5]). In semantical terms, the interpretation function of $\mathcal{I}$ now maps every concept (resp. role) into a function from $\Delta_{\mathcal{I}}$ (resp. $\Delta_{\mathcal{I}} \times \Delta_{\mathcal{I}}$) to the powerset of $\{t, f\}$, and every object into $\Delta_{\mathcal{I}}$. The interpretation of a concept C can best be understood as consisting of a pair of two separate sets: the *positive extension* $C^{\mathcal{I}}_{+} = \{o \in \Delta_{\mathcal{I}} \mid t \in C^{\mathcal{I}}(o)\}$ and the *negative extension* $C^{\mathcal{I}}_{-} = \{o \in \Delta_{\mathcal{I}} \mid f \in C^{\mathcal{I}}(o)\}$. The case of roles is analogous. The extensions of concepts and roles have to meet certain restrictions, designed so that the formal semantics complies with the informal meaning of constructors. For example, the positive extension of $C_1 \sqcap C_2$ must be the intersection of the positive extensions of C_1 and C_2, and its negative extension must be the *union* of their negative extensions. For reasons of space we detail the four-valued semantics of a few representative operators only:

$$
\begin{array}{rcl}
t \in (\neg C)^{\mathcal{I}}(o) & \text{iff} & f \in C^{\mathcal{I}}(o) \\
f \in (\neg C)^{\mathcal{I}}(o) & \text{iff} & t \in C^{\mathcal{I}}(o) \\
t \in (C_1 \sqcap C_2)^{\mathcal{I}}(o) & \text{iff} & t \in {C_1}^{\mathcal{I}}(o) \text{ and } t \in {C_2}^{\mathcal{I}}(o) \\
f \in (C_1 \sqcap C_2)^{\mathcal{I}}(o) & \text{iff} & f \in {C_1}^{\mathcal{I}}(o) \text{ or } f \in {C_2}^{\mathcal{I}}(o) \\
t \in (\forall R.C)^{\mathcal{I}}(o) & \text{iff} & \forall\, o' \in \Delta_{\mathcal{I}},\ \text{if } t \in R^{\mathcal{I}}(o, o') \text{ then } t \in C^{\mathcal{I}}(o') \\
f \in (\forall R.C)^{\mathcal{I}}(o) & \text{iff} & \exists\, o' \in \Delta_{\mathcal{I}}, t \in R^{\mathcal{I}}(o, o') \text{ and } f \in C^{\mathcal{I}}(o')
\end{array}
$$

An interpretation $\mathcal{I}$ *satisfies an assertion* $C(a)$ (resp. $R(a_1, a_2)$) iff $t \in C^{\mathcal{I}}(a^{\mathcal{I}})$ (resp. $t \in R^{\mathcal{I}}({a_1}^{\mathcal{I}}, {a_2}^{\mathcal{I}})$), *satisfies a definition* $C_1 \sqsubseteq C_2$ (resp. $R_1 \sqsubseteq R_2$) iff ${C_1}^{\mathcal{I}}_{+} \subseteq {C_2}^{\mathcal{I}}_{+}$ (resp. ${R_1}^{\mathcal{I}}_{+} \subseteq {R_2}^{\mathcal{I}}_{+}$), and *satisfies a KB* Σ iff it satisfies all assertions and definitions in Σ. A KB Σ *entails* an assertion α (written $\Sigma \models_4 \alpha$) iff all interpretations satisfying Σ also satisfy α. It can be proven that $\models_4$ is a subset of $\models$, which guarantees the soundness of the entailment relation wrt the two-valued semantics. Also, unlike other four-valued logics such as [5], it licenses a restricted form of *modus ponens,* which we call *modus ponens on roles,* whose consequence is that $\{(\forall R.C)(a_1), R(a_1, a_2)\} \models_4 C(a_2)$ and $\{(\forall R.C_1)(a), (\exists R.C_2)(a)\} \models_4 (\exists R.C_1 \sqcap C_2)(a)$. Also, $\models_4$ avoids the *paradoxes of material implication*; i.e. in our logic (1) not everything is entailed by an inconsistent KB, and (2) a tautology is not entailed by every KB. The possibility of making sense of inconsistent KBs is an advantage in the light of the fact that, given the huge amount of documents belonging to a document base, we can expect mutual inconsistencies in their representations to pop up. The proofs of these properties and a full discussion of our four-valued DL can be found in [4].

3 Document structure and closures

The inadequacy of DLs in representing document structures is due to the fact that they incorporate the so-called *open-world assumption*. Suppose that all the information available about an individual `John` is that it is a `Person` and that it has a `Brother Bill`, who is a `Musician`. This information is not sufficient to infer, for example, that all of `John`'s brothers are musicians, or that `John` is not an elephant. These inferences, as well as a number of others that our intuition would deem correct, could be obtained only at the price of completing the characterization of `John` with all the "negative facts" about it. In order to overcome this problem, we introduce in our language a new "closure" operator $\mathcal{C}$; the expression $\mathcal{C}(a)$ amounts to specifying that the knowledge on individual a is *complete*. We further extend the inference relation of the logic to correctly interpret such a specification by applying a *closed-world assumption* on the specified individual. The resulting entailment relation, $\models_4^c$ is such that from $\Sigma =$ $\{$`Person(John)`, `Brother(John,Bill)`, `Musician(Bill)`$\}$ and $\mathcal{C}$(`John`), it follows that $\langle\Sigma, \{\mathcal{C}($`John`$)\}\rangle \models_4^c$ ($\forall$`brother.Musician)(John)` and $\langle\Sigma, \{\mathcal{C}($`John`$)\}\rangle \models_4^c$ $\neg$`Elephant(John)`. In a somehow dual way, one would like to express the fact that, for instance, all the persons are those *known* to the KB, without bothering to write down an assertion of type $\neg$`Person`(a) for all the scores of individuals that are not persons. This can be done by using a so-called *primitive closure*, i.e. an expression of the form $\mathcal{C}(T)$

where T is either a primitive concept or a primitive role. Finally, we allow closure expression of the form $\mathcal{C}(a, P)$, where a is an individual and P is a primitive role; this is a natural combination of $\mathcal{C}(a)$ and $\mathcal{C}(P)$, meaning that "all P's of a are known". For example, $\mathcal{C}(\texttt{John}, \texttt{Brother})$ means that all brothers of `John` are known. Semantically, closures are modelled by fixing the domain to a non-empty, countable set Δ of symbols called *parameters* (denoted hereafter by p), which are assigned to individuals by a fixed injective function γ. A *c-interpretation* $\mathcal{I}$ is a four-valued interpretation such that $\Delta = \Delta_{\mathcal{I}}$ and such that for each individual a, $a^{\mathcal{I}} = \gamma(a)$. The notion of satisfaction of "open" assertions is extended to c-interpretations in the obvious way. By $\mathcal{M}(\Sigma)$ we will indicate the set of all c-interpretations satisfying Σ.

Satisfaction of closures is defined on the basis of a notion of "minimal knowledge" modelled by so-called *epistemic interpretations*; these latter have already been used for the semantics of epistemic DLs [1]. An epistemic interpretation is a pair $\langle \mathcal{I}, \mathcal{W} \rangle$ where $\mathcal{I}$ is a c-interpretation and $\mathcal{W}$ is a set of c-interpretations. An epistemic interpretation *satisfies* a closure $\mathcal{C}(a)$ iff

- for every primitive concept A, $t \in A^{\mathcal{I}}(\gamma(a))$ iff $t \in A^{\mathcal{J}}(\gamma(a))$ for all $\mathcal{J} \in \mathcal{W}$ and $f \in A^{\mathcal{I}}(\gamma(a))$ iff $t \notin A^{\mathcal{J}}(\gamma(a))$ for some $\mathcal{J} \in \mathcal{W}$, and
- for every primitive role P and parameter $p \in \Delta$, $t \in P^{\mathcal{I}}(\gamma(a), p)$ iff $t \in P^{\mathcal{J}}(\gamma(a), p)$ for all $\mathcal{J} \in \mathcal{W}$ and $f \in P^{\mathcal{I}}(\gamma(a), p)$ iff $t \notin P^{\mathcal{J}}(\gamma(a), p)$ for some $\mathcal{J} \in \mathcal{W}$.

This amounts to saying that, for any c-interpretation satisfying a KB $\langle \Sigma, \Omega \rangle$ and for any "closed" individual a, $a^{\mathcal{I}}$ is allowed in the positive extension of a primitive concept A just in case $A(a)$ is entailed by Σ, in symbols $\Sigma \models_4 A(a)$. As a consequence, the lack of positive information allows, as shown in [4], to infer the corresponding negative information. The case of roles is analogous. The semantics of primitive assertions is perfectly dual; it constrains the extensions of closed primitive concepts and roles with respect to parameters. Finally, the interpretation of closures on role fillers is the obvious combination of that of the other two kinds of closures. An epistemic interpretation *satisfies* a set of closures if and only if it satisfies each closure in the set. A c-interpretation $\mathcal{I}$ *satisfies* the KB $\langle \Sigma, \Omega \rangle$ iff $\mathcal{I}$ satisfies Σ and $\langle \mathcal{I}, \mathcal{M}(\Sigma) \rangle$ satisfies Ω. Essentially, in order to satisfy a KB, a c-interpretation has to satisfy the assertions in Σ and the requirements imposed by closures and detailed above. A document base $\langle \Sigma, \Omega \rangle$ *c-entails* a concept C, written $\langle \Sigma, \Omega \rangle \models_4^c C$, if and only if all c-interpretations satisfying $\langle \Sigma, \Omega \rangle$ also satisfy C.

In order to perform document retrieval, we have developed a sound and complete Gentzen calculus [4], based on the analogous calculus for first order logic. We are developing a prototypical implementation of the calculus in order to experimentally verify the adequacy of the model.

4 Modelling uncertainty

The logic we have described so far is still insufficient for describing *real* retrieval situations, as retrieval is usually not a yes-no question: the representations of documents and information needs which the system (and the logic) have access to are inherently imperfect, and the relevance of a document to an information need can thus be established only up to a limited degree of certainty. Because of this, we need a framework in which, rather than deciding *tout court* whether d is relevant to n, we are able to *rank* documents d_i according to how strongly the system believes in their relevance to n. This means that we need a logic that captures the notion of the *degree of belief DB* of the system in a certain formula. If $C(a)$ represents the fact that document a is relevant to information need C, then we need to be able to compute $DB(C(a))$, which amounts to checking whether, for some $0 \leq r \leq 1$, $DB(C(a)) = r$ follows from the KB that formalises the domain of application. This KB will contain assertions and definitions that are themselves uncertain, as it is exactly the uncertainty that the system has in the premises (namely, in the document contents, in the meaning of queries, and in the fact that thesaural entries can be relied on *tout court*) that determines the uncertainty in the conclusion. For instance, we will need to specify that the system believes with strength r that document `a` deals with topic C, e.g. by means of an expression of type $DB((\exists \texttt{DealsWith}.C))(a) | \texttt{document}(a)) = r_1$ (*conditional probabilistic assertions*). Additionally, we will also need to specify that, for instance, if the system believes with degree r that a document a is about topic C_1, it will believe that it is also about topic C_2 with strength $r_1 \cdot r_2$, e.g. by means of an expression of type $DB(C_2 \mid C_1) = r_2$ (*conditional probabilistic definitions*).

Following the approach first developed in [6] we extend the syntax of our DL with an operator DB, formalising the notion of the system's degree of belief. We give semantics to the DB operator by relying on the notion of *PDL structure*, i.e. a triple $M = \{\Delta, \Psi, \nu\}$, where Δ is as in Section 3, Ψ is a set of epistemic interpretations and ν is a discrete probability distribution on Δ. An assertion $DB(\alpha) = r$ is true in a c-interpretation $\mathcal{I}$ (written $(M, \mathcal{I}) \models DB(\alpha) = r$) belonging to Ψ iff $\sum_{\mathcal{J} \in \Psi \mid (M,\mathcal{J}) \models \alpha} \nu(\mathcal{J}) = r$. This is in keeping with a *possible worlds* reading of degrees of belief, in which the degree of belief in a formula α is a measure of how possible, according to the system, are the states of affairs that make α true. Full details on degrees of belief as applied to description logics and retrieval may be found in [6].

5 Dealing with multimediality

As already pointed out, multimediality plays a crucial role at the level of document form. Here, notions developed in the signal processing area have to be imported into the model *and*, most importantly, *appropriately combined* with the symbolic machinery used to represent document contents and related information. Existing models fall short of this latter requirement. In order to show what we mean by "appropriate combination" of the signal and the symbolic processing level, we briefly outline, in the following, the basic ideas behind an image retrieval model we have recently developed [2]. The image model is based on standard predicate calculus, where predicate symbols are tailored to image form and content representation. Essentially, an image form is queried via instances of the binary predicate symbol I, having as arguments a term denoting a region and a term denoting a colour. In particular, the sentence $I(r, c)$ is true in an interpretation whenever the region named r is of the colour named c. The content of an image is modelled as a set of assertions of the DL previously introduced. Form and content are connected via the binary predicate symbol Map, whose arguments denote a region of the image and an individual of the image's content representation. For instance, the query asking for the images where a person named *giulia* is wearing something pink can be expressed as the sentence $(\exists x)I(x, pink) \wedge Map(x, giulia)$, where the variable x ranges on regions. In this model we allow spatial properties of images to be addressed in queries by using a variant of the interval-based logic defined by Allen for time. A typical iconic query might be "find all pictures with a tree to the left of a house", expressible in our query language as:

$$\begin{aligned}(\exists xy) \quad & (Tree(x) \wedge House(y) \wedge (\forall u)(Map(u, x) \rightarrow ((\forall v)Map(v, y) \rightarrow X\_before(u, v))) \wedge \\ & (\exists wz)(Map(w, x) \wedge Map(z, y) \wedge Y\_overlap(w, z))).\end{aligned}$$

This model is currently being integrated with the model presented in the previous sections. The integration is made possible by the fact that both models comply with the standards of denotational semantics.

References

[1] F. Donini, M. Lenzerini, D. Nardi, A. Schaerf, and W. Nutt. Adding epistemic operators to concept languages. In *Proceedings of KR-92*, pages 342–353, 1992.

[2] C. Meghini. An image retrieval model based on classical logic. In *Proceedings of ACM SIGIR-95*, pages 300–308, Seattle, WA, 1995.

[3] C. Meghini, F. Sebastiani, U. Straccia, and C. Thanos. A model of information retrieval based on a terminological logic. In *Proceedings of ACM SIGIR-93*, pages 298–307, Pittsburgh, PA, 1993.

[4] C. Meghini and U. Straccia. A relevance description logic for information retrieval. In *Proceedings of ACM SIGIR-96*, Zürich, CH, 1996. Forthcoming.

[5] P. F. Patel-Schneider. A four-valued semantics for terminological logics. *Artificial Intelligence*, 38:319–351, 1989.

[6] F. Sebastiani. A probabilistic terminological logic for modelling information retrieval. In *Proceedings of ACM SIGIR-94*, pages 122–130, Dublin, IRL, 1994.