Abstract

According to the logical model of Information Retrieval (IR), the task of IR can be described as the extraction, from a given document base, of those documents $d$ that, given a query $q$, make the formula $d \rightarrow q$ valid, where $d$ and $q$ are formulae of the chosen logic and “$\rightarrow$” denotes the brand of logical implication formalized by the logic in question. In this paper, although essentially subscribing to this view, we propose that the logic to be chosen for this endeavour be a Terminological Logic (TL): accordingly, the IR task becomes that of singling out those documents $d$ such that $d \preceq q$, where $d$ and $q$ are terms of the chosen TL and “$\preceq$” denotes subsumption between terms. We call this the terminological model of IR.

TLs are particularly suitable for modelling IR; in fact, they can be employed: 1) in representing documents under a variety of aspects (e.g. structural, layout, semantic content); 2) in representing queries; 3) in representing lexical, “thesaural” knowledge. The fact that a single logical language can be used for all these representational endeavours ensures that all these sources of knowledge will participate in the retrieval process in a uniform and principled way.

In this paper we introduce Mirtl, a TL for modelling IR according to the above guidelines; its syntax, formal semantics and inferential algorithm are described.