# Description Logic-Based Resource Retrieval



**Simona Colucci**
*Politecnico di Bari, Italy*

**Tommaso Di Noia**
*Politecnico di Bari, Italy*

**Eugenio Di Sciascio**
*Politecnico di Bari, Italy*

**Francesco M. Donini**
*Universitá della Tuscia, Italy*

**Marina Mongiello**
*Politecnico di Bari, Italy*

## INTRODUCTION

*Resource retrieval* addresses the problem of finding best matches to a request among available resources, with both the request and the resources described with respect to a shared interpretation of the knowledge domain the resource belongs to. The problem of resource matching and retrieval arises in several scenarios, among them, personnel recruitment and job assignment, dating agencies, but also generic electronic marketplaces, Web services discovery and composition, resource matching in the Grid. All these scenarios share a common purpose: given a request, find among available descriptions those best fulfilling it, or at "worse," when nothing better exists, those that fulfill at least some of the requirements.

Exact, or full, matches are usually rare and the true *matchmaking* process is aimed at providing one or more "promising" matches to be explored. Non-exact matches should take into account both missing information—details that could be positively assessed in a second phase—and conflicting information—details that could leverage negotiation if the proposed match is worth enough pursuing.

Because of its intangibility, it is now a widely shared opinion that knowledge has to be modeled to make unambiguous the interpretation of any information domain. This disambiguation process is usually obtained through an *ontology*, that is, a specification of a representational vocabulary for a shared domain of discourse—definitions of classes, relations, functions, and other objects (Gruber, 1993).

Once a knowledge domain has been modeled, and several different resources have been described using such a model, issues that need to be faced for efficient knowledge management are: What, if any, kind of retrieval is possible on these resources? How could we benefit both of the model and formalisms used to build the model, in order to perform a "smart" search of described resources matching a request? The above questions focus on important aspects of knowledge-based retrieval:

- formalisms used to model a knowledge domain
- retrieval services that fully use the expressiveness of the formalism to infer new knowledge from the model in order to perform a knowledge-based search

Knowledge domain is modeled with a formalism, whose expressiveness is used in the retrieval process to infer not elicited information from the model. In such a context, choosing this formalism strongly affects the complexity, as well as success probability, of the retrieval process.

In recent years *description logics* (DLs) have been investigated by both the academic and industrial world as a formalism for knowledge representation. Modeling an information domain through the formalism of a DL allows one to employ reasoning services provided by DLs to perform a knowledge-based search. Knowledge domains are formalized in ontologies, which resource descriptions refer to. The use of ontologies allows elicited descriptions to be stored so that information can be inferred from them to retrieve a resource.

The remainder of this article is structured as follows: Background work is revised, including DL basics with associated reasoning services and previous approaches to resource retrieval, including non-logic- and logic-based alternatives. Then, we introduce semantic-based resource retrieval, first highlighting new non-standard inference services and then showing how they can be used for "smart" resource retrieval. Finally, we propose some future trends and draw a conclusion.

# BACKGROUND

## Description Logics Basics

Description, or terminological, logics (Baader, Calvanese, Mc Guinness, Nardi, & Patel-Schneider, 2002; Donini, Lenzerini, Nardi, & Schaerf, 1996) are a family of logic formalisms for knowledge representation. All DLs are endowed of a syntax and a model-theoretic semantics. The basic syntax elements of DLs are: *concept* names, *role* names, *individuals*. Intuitively, concepts stand for sets of objects, and roles link objects belonging to different concepts. Individuals are special named elements of the sets of objects concepts represent.

We give a more formal definition of the outlined basic elements by introducing the concept of semantic *interpretation*.

**Definition 1:** *A semantic interpretation is a pair* $I=(\Delta, \cdot^I)$ *made up of a domain* $\Delta$ *and an interpretation function* $\cdot^I$ *, which maps every concept to a subset of* $\Delta$*, every role to a subset of* $\Delta\times\Delta$*, and every individual to an element of* $\Delta$*.*

Usually, a so-called *Unique Name Assumption* (UNA) is made which ensures different individuals to be mapped to different elements of $\Delta$, *i.e.*, $a^I \neq b^I$ for individuals $a \neq b$.

Every DL allows one to combine basic elements using *constructors* to form concept and role *expressions*. Each DL has its distinguished set of constructors, though all of them provide the *conjunction* of concepts, usually denoted as $\sqcap$. Among the distinguishing concept expressions constructors we enumerate disjunction $\sqcup$ of concepts and complement $\neg$ to close concept expressions under Boolean operations.

Role expressions can be obtained by combining roles with concepts using *existential role quantification* and *universal role quantification*. Other constructs may involve counting, as *number restrictions*.

Many other constructs can be defined, increasing the expressive power of the DL, up to n-ary relations (Calvanese, De Giacomo, & Lenzerini, 1998). Nevertheless, it is a well-known result that usually leads to an explosion in computational complexity of inference services (Brachman & Levesque, 1984). Hence, a trade-off is needed between expressivity and expected performance of reasoning services.

Once expressions have been built, they are given semantics by defining the interpretation function over each construct. Concept conjunction is interpreted as set intersection, and the other Boolean connectives also have the usual set-theoretic interpretation. The interpretation of constructs involving quantification on roles needs to make domain elements explicit.

Concept expressions can be used in *inclusion assertions*, and *definitions*, which impose restrictions on possible interpretations according to the knowledge elicited for a given domain. Definitions are useful to give a meaningful name to particular combinations. Sets of such inclusions are called TBox (terminological box). A TBox, which basically amounts to an ontology, represents a formal, shared, and objective *intensional* knowledge on a domain. Individuals can be asserted to belong to a concept using membership assertions in an ABox. An ABox is the *extensional* knowledge of the domain that can be described based on the TBox. The semantics of inclusions and definitions is based on set containment: An interpretation $I$ satisfies an inclusion $C \sqsubseteq D$ if $C^I \sqsubseteq D^I$ , and it satisfies a definition $C = D$ when $C^I=D^I$. A *model* of a TBox $T$ is an interpretation satisfying all inclusions and definitions of $T$. DL-based systems are equipped with reasoning services: logical problems whose solution can make explicit knowledge that was implicit in the assertions.

DL-based systems usually provide at least two basic reasoning services for $T$:

- **Concept Satisfiability:** Given a TBox $T$ and a concept C, does there exist at least one model of $T$ assigning a non-empty extension to C?
- **Subsumption:** Given a TBox $T$ and two concepts C and D, is C more general than D in any model of $T$?

The previous services can be seen, from a knowledge management perspective, in a more informal way:

- **Concept Satisfiability:** Given an ontology ($T$) modeling the domain we are investigating on and a description (C) of a resource referring to the ontology: Is the information modeled in the description consistent with the one in the ontology?
- **Subsumption:** Given an ontology ($T$) modeling the domain we are investigating on and two resources described by expressions (C, D) referring to the information modeled in the ontology: Is the information about a resource more general than the one related to the other one?