Journal of Applied Logic 13 (2015) 480–508

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

Similarity-based relaxed instance queries

Andreas Ecke [a,*], Rafael Peñaloza [a,b], Anni-Yasmin Turhan [a]

[a] *Institute for Theoretical Computer Science, Technische Universität Dresden, Germany*
[b] *Center for Advancing Electronics Dresden, Germany*

ARTICLE INFO

Article history:
Received 16 February 2014
Available online 19 January 2015

Keywords:
Description Logics
Instance queries
Concept similarity measures

ABSTRACT

In Description Logics (DL) knowledge bases (KBs), information is typically captured by clear-cut concepts. For many practical applications querying the KB by crisp concepts is too restrictive; a user might be willing to lose some precision in the query, in exchange of a larger selection of answers. Similarity measures can offer a controlled way of gradually relaxing a query concept within a user-specified limit. In this paper we formalize the task of instance query answering for DL KBs using concepts relaxed by concept similarity measures (CSMs). We investigate computation algorithms for this task in the DL $\mathcal{EL}$, their complexity and properties for the CSMs employed regarding whether unfoldable or general TBoxes are used. For the case of general TBoxes we define a family of CSMs that take the full TBox information into account, when assessing the similarity of concepts.

1. Introduction

Description Logics (DLs) are a family of knowledge representation formalisms that have unambiguous logic-based semantics. Each particular DL is characterized by a set of concept constructors, which allow to build complex concepts. Intuitively, *concepts* characterize categories from an application domain. In addition, binary relations on the domain of interest can be captured by *roles*. These in turn can be used to build more complex concepts with the help of a class of concept constructors. The terminological knowledge of an application domain is stored in the *TBox*, that expresses the relationships between concepts. Facts from the application domain and relations between them are represented by assertions about *individuals* in the *ABox*. TBox and ABox together form the DL *knowledge base* (KB).

The formal semantics of DLs allow the definition of a variety of reasoning services. The most prominent ones are *subsumption*, i.e. to compute whether a sub-concept relationship holds between two concepts and

* Corresponding author.
E-mail addresses: ecke@tcs.inf.tu-dresden.de (A. Ecke), penaloza@tcs.inf.tu-dresden.de (R. Peñaloza), turhan@tcs.inf.tu-dresden.de (A.-Y. Turhan).

http://dx.doi.org/10.1016/j.jal.2015.01.002
1570-8683/

instance query answering, where for a given concept all individuals from an ABox that are instances of the query concept are computed. These reasoning services are implemented in highly optimized reasoning systems, see for example [18,19,24,25].

DLs of varying expressivity are the underlying logics for the W3C standardized ontology language OWL 2 and its profiles [26]. This standardization has led to an increased use of DLs and DL reasoning systems in the recent years in many application areas. By now there is a large collection of KBs written in these languages. However, many applications need to query the knowledge base in a more relaxed manner. For instance, in the application area of service matching OWL TBoxes are employed to describe types of services. Here, a user request for a service specifies several requirements for the desired service. These conditions are represented by a complex concept. For such a concept the OWL ABox that contains the individual services is searched for a service matching the specified request by performing instance query answering. In cases where an exact match with the provided requirements is not possible, a 'feasible' alternative should be retrieved from the ABox containing the services to be able to offer an alternative. Essentially, for a given query concept, the system should retrieve all those individuals of the ABox that fulfill the main requirements, while allowing a relaxation of some of the less crutial requirements.

A natural idea on how to relax the notion of instance query answering is to simply employ fuzzy DLs and perform query answering on a fuzzy variant of the initial query concept. However, on the one hand reasoning in fuzzy DLs easily becomes undecidable [7–9] and on the other hand depending on the user and on the request, different ways of relaxing the query concept are needed. For instance, for a request to a car rental company to rent a particular car model in Beijing, it might be acceptable to get an offer for a similar car model to be rented in Beijing, instead of getting the offer to rent the requested car model in London. Whereas for a handicapped user in a wheelchair it might not be acceptable to relax the requested car model from a two-door one to a four-door one. Fuzzy concepts would relax the initial concept in an unspecific and uniform way. In contrast, relaxed instance query answering should allow to

1. choose *which aspects* of the query concept can be relaxed and
2. choose the *degree* to how much these aspects can be relaxed.

The reasoning service addressed in this paper is a relaxed notion of instance querying, such that it allows for a given query concept the selective and gradual extension of the answer set of individuals. We develop a formal definition of this reasoning service in Section 3.

The selective and gradual relaxation of the answer sets returned by instance query answering is achieved by the use of concept similarity measures. A *concept similarity measure* (CSM) yields, for a pair of concepts, a value from the interval $[0, 1]$—indicating how similar the concepts are. To answer a relaxed instance query is to compute for a given concept C, a CSM $\sim$ and a degree t between 0 and 1, a set of concepts such that each of these concepts is similar to C by a degree of at least t, if measured by the CSM $\sim$, and then finding all their instances.

Concept similarity measures are widely used in ontology-based applications. In the bio-medical field, for example the Gene ontology [10], they are employed to discover functional similarities of genes (see e.g. [21, 30]). Furthermore, CSMs are used in ontology alignment algorithms [17]. For DLs there exists a whole range of CSMs, which could be employed for the task of answering relaxed instance queries [6,11,20,34]. In particular the CSMs generated by the framework described in [20] allow users to specify which part of the vocabulary used in their knowledge base is to be regarded more important when it comes to the assessment of similarity of concepts. Thus, the measures generated by this framework naturally allow users to select important features of the query concept and which aspect of the query concept to relax.

We investigate algorithms for computing answers to instance queries relaxed by CSMs for the light-weight DL $\mathcal{EL}$. Our choice for the DL $\mathcal{EL}$ is motivated by the fact that reasoning in $\mathcal{EL}$ has good computational properties—most standard reasoning problems can be solved in polynomial time [4]. Large, well-known