

# Robust Query Processing for Personalized Information Access on the Semantic Web

Peter Dolog<sup>1</sup>, Heiner Stuckenschmidt<sup>2</sup>, and Holger Wache<sup>3</sup>

<sup>1</sup> L3S Research Center, Hannover, Germany  
dolog@l3s.de

<sup>2</sup> Universität Mannheim, Germany  
heiner.stuckenschmidt@uni-mannheim.de

<sup>3</sup> Vrije Universiteit Amsterdam, The Netherlands  
holger@cs.vu.nl

**Abstract.** Research in Cooperative Query answering is triggered by the observation that users are often not able to correctly formulate queries to databases that return the intended result. Due to a lack of knowledge of the contents and the structure of a database, users will often only be able to provide very broad queries. Existing methods for automatically refining such queries based on user profiles often overshoot the target resulting in queries that do not return any answer. In this paper, we investigate methods for automatically relaxing such over-constraint queries based on domain knowledge and user preferences. We describe a framework for information access that combines query refinement and relaxation in order to provide robust, personalized access to heterogeneous RDF data as well as an implementation in terms of rewriting rules and explain its application in the context of e-learning systems.

## 1 Introduction

Research in Cooperative Query answering is triggered by the observation that users are often not able to correctly formulate queries to databases that return the intended result. This is even more the case for semantic web systems based on RDF for the following reasons:

- The data accessed often comes from different sources. The internal structure of these sources is not always known.
- The data is semi structured. Sources do not have to describe all aspects of the information resources.
- There is no fixed integrated schema. Each source can have its own schema, sources may make partial use of different available schemas.

With the increasing popularity of RDF as a representation language in domains such as medicine [17] or e-learning [4] this problem becomes more pressing. If RDF query languages are to be used in a large scale we have to make sure that people will be able to formulate meaningful queries. If this is not the case, we have to find ways to still provide the user with the intended results. Cooperative query processing aims at supporting

the user by automatically modifying the query in order to better fit the real intention of the user. Based on the assumed kind of mismatch between the users intention and the formulated query there are different techniques used. We consider two basic mechanisms of cooperative query processing, *query refinement* and *query relaxation* which are briefly presented in the following.

### 1.1 Query Refinement and Relaxation

Due to a lack of knowledge of the contents and the structure of a database, users will often only be able to provide very broad queries, for example in terms of the type of the objects she wants to retrieve and maybe one or two properties. Taking an example from the domain of e-learning, the user might be able to specify that she is looking for a lecture on the Java Programming Language. Learning resources, however, are often annotated with a fair amount of metadata that specifies important information such as the assumed level of expertise and required previous knowledge. In order to select learning resources that are suited for the user, these additional properties have to be specified in the query as well. Dolog et al [4] show that this information can be included into a user query based on a user profile. They describe a method for automatically refining queries with information from the user profile thereby enabling a pre-selection of query answers.

A problem of the automatic refinement of queries lies in the fact that it often overshoots the target instead of too many results an automatically refined query often returns no result at all, because none of the resources exactly matches the users needs. A possible solution to this problem is to successively relax the constraints imposed in the refinement step. Different Techniques for relaxing queries have been proposed in the database area. Gaasterland et al [6] provide a unifying view on different relaxation techniques in terms of replacing subexpressions in the query. In previous work we described an approach for relaxing conjunctive queries over description logic knowledge bases by removing conjuncts from the query in a particular order [16, 18] .

### 1.2 Contributions

In this paper, we build upon existing work on query refinement for personalized information access and extend it in the following ways:

- We describe a framework for information access that combines query refinement and relaxation in order to provide robust, personalized access to heterogeneous RDF data.
- We propose an implementation of the framework in terms of conditional rewriting rules for RDF query patterns.
- We discuss the application of the framework in the context of an existing e-learning system.

## 2 Background

User queries to the open learning environment will consist of one or several keywords related to the topic the user wants to learn about. The result is a list of learning resources including information about the subject and the title of the resource as well as