

Querying Ontologies: A Controlled English Interface for End-users

Abraham Bernstein, Esther Kaufmann, Anne Göhring, Christoph Kiefer

University of Zurich, Department of Informatics, Winterthurerstrasse 190,
8057 Zurich, Switzerland
{bernstein, kaufmann, goehring, kiefer}@ifi.unizh.ch,
<http://www.ifi.unizh.ch/ddis>

Abstract. The semantic web presents the vision of a distributed, dynamically growing knowledge base founded on formal logic. Common users, however, seem to have problems even with the simplest Boolean expressions. As queries from web search engines show, the great majority of users simply do not use Boolean expressions. So how can we help users to query a web of logic that they do not seem to understand? We address this problem by presenting a natural language interface to semantic web querying. The interface allows formulating queries in Attempto Controlled English (ACE), a subset of natural English. Each ACE query is translated into a discourse representation structure – a variant of the language of first-order logic – that is then translated into an N3-based semantic web querying language using an ontology-based rewriting framework. As the validation shows, our approach offers great potential for bridging the gap between the logic-based semantic web and its real-world users, since it allows users to query the semantic web without having to learn an unfamiliar formal language. Furthermore, we found that users liked our approach and designed good queries resulting in a very good retrieval performance (100% precision and 90% recall).

1 Introduction

The semantic web presents the vision of a dynamically growing knowledge base that should allow users to draw on and combine distributed information sources specified in languages based on formal logic. Common users, however, were shown to have problems even with the simplest Boolean expressions; the use of the description logic formalism underlying the semantic web is beyond their understanding. Experience in information retrieval, for example, demonstrates that users are better at understanding graphical query interfaces than simple Boolean queries [1]. As queries from web search engines reveal, the great majority of users simply do not use Boolean expressions. Bowen and colleagues even show that people (CS students) who are trained in formulating queries in a logic-based formalism (SQL in their case) are usually inept in composing correct queries in realistically-sized databases rather than the small toy examples used in database classes [2]. *So how can we bridge the gap between the*

(description) logic-based semantic web and real-world users, who are at least ill at ease and, oftentimes, unable to use formal logic concepts?

We address this problem by *presenting a natural language interface to the semantic web*. In its current form the interface provides users with a controlled natural language interface to formulate queries. The controlled natural language used, Attempto Controlled English (ACE) [3, 4], is an unambiguous subset of English, which is translated *automatically* into the N3-style¹ triple-based semantic web query language PQL [5] (which can easily be mapped to query languages such as SquishQL [6]). It provides the users with an almost natural language interface to the semantic web. As experience with controlled languages has shown, they are much easier to learn by end-users than formal languages like logic and are sufficient for querying knowledge bases [7]. We, therefore, believe that the approach presented here has great potential in bridging the gap between the semantic web and its end-users as well as becoming a major enabler for the growth of the semantic web.

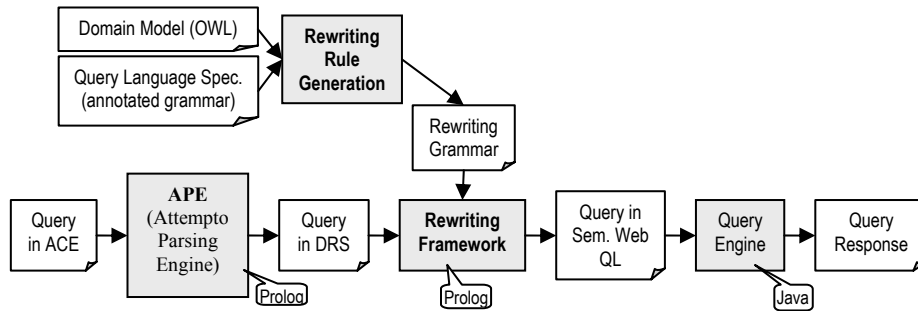


Fig. 1. Overall data flow of the controlled English query interface

The rest of this paper closely follows the data flow of the query interface (Fig. 1). Section 2 introduces Attempto Controlled English (ACE) and the Attempto Parsing Engine (APE). APE translates ACE texts into a discourse representation structure (DRS), a variant of the language of first-order logic introduced by Kamp and Reyle [8]. Section 3 describes the rewriting framework that translates the DRS to the semantic web query language. The translation is based on a rewriting grammar, which was generated using both an OWL-based domain model and a query language specification. The queries are evaluated by a standard query engine not discussed in this paper. Note that we used APE as a black-box component, which uses extended DRSs as internal representations. This allows us to exchange it with another NLP parser should the need arise. Therefore, we did not attempt a direct translation from ACE to N3. In section 4 we provide three evaluations of the approach. We close with a discussion of the current limitations as well as related and future work.

¹ More information about N3 can be found at <http://www.w3.org/DesignIssues/Notation3>

