Talking to the Semantic Web - Query Interfaces to Ontologies for the Casual User

Esther Kaufmann

University of Zurich, Dynamic and Distributed Information Systems, Switzerland
kaufmann@ifi.unizh.ch

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 [4]; the use of the logic formalism underlying the Semantic Web is beyond their understanding. So how can we bridge the gap between the logic-based Semantic Web and real-world users, who are ill at ease and, oftentimes, unable to use formal logic concepts?

An often proposed solution to address this problem is the use of natural language interfaces (NLIs). Most NLIs, however, only understand some subset of natural language (NL), but often suggest full understanding, which leads to confusing interaction with users [1]. This mismatch between the users' expectations and the capabilities of a NLI is called the *habitability problem* [5]. Furthermore, the development of NL tools requires computationally intensive algorithms relying on large amounts of background knowledge making the tools highly domain-dependent and inapplicable to new domains or applications [1].

This project proposes to break the dichotomy between full NLIs and formal query approaches regarding them as ends of a *Formality Continuum*. It argues that query interfaces should impose some structure on the user's input to guide the entry but not overly restrict the user with an excessively formalistic language. In this way, we hypothesize that the best solutions for the casual and occasional user lie between the freedom of a full NLI and the structuredness of a formal query language. Furthermore, the use of controlled NLs facilitates to overcome both the habitability problem and the adaptivity barrier of full NLIs. The overarching goal is to turn the vision of the Semantic Web into realization, which can only happen if we bridge the gap between the end-users and the logic-based scaffolding of the Semantic Web.

2   Current State of the Project and Future Research

To support our proposition we have developed two different controlled language interfaces to the Semantic Web that lie in the middle of the Formality Continuum: Ginseng [2] and SWAT [3]. Both allow users to formulate queries in a language akin to English. To preliminarily evaluate our interfaces, we confronted

users with the Ginseng prototype, with the SWAT prototype, and with a SQL interface as well as a keyword-based text search engine as competitors. Using a standardized usability test, we found that both Ginseng and SWAT significantly outperformed the SQL interface. Both interfaces also yielded much better retrieval performances than the competitors.

While the results of our preliminary usability evaluation are promising, many challenges remain. At the current state, the project provides a good basis for a deeper exploration and evaluation of NLIs to Semantic Web knowledge bases. More specifically, we would like to investigate our hypotheses regarding the Formality Continuum to the fullest extent. We, therefore, intend to (1) develop and implement a total of four different NLIs (two of which are extensions of Ginseng and SWAT) for the casual users to query ontologies, (2) thoroughly evaluate the usability and performance of these NLIs by conducting a comprehensive user study allowing us to generate conclusive evidence regarding our hypotheses.

As a first step, we will design and implement two more NLIs. The new interfaces are intended to "veer" towards both ends of the continuum. Consequently, one new interface will allow full NL input, whereas the other one tends to follow the ideas of the formal approaches. The interfaces will be as domain-independent and easily customizable for different ontologies as possible by extracting the necessary underlying frameworks from the knowledge bases. We pursue to avoid complex and tedious full NLP scaffolds on the one hand as well as a formal query interface that shifts all "intellectual" work to the user on the other hand.

As the second step, we plan to perform a thorough evaluation of the interfaces according to well-established methodologies. The evaluation will comprise three elements: (1) a test set evaluation showing the retrieval performance of our interfaces as well a other NLIs, (2) a heuristic usability evaluation to systematically inspect our interfaces to detect their usability problems, and (3) a comprehensive usability study (benchmarking the tools against each other as well as to other existing NLIs). These evaluation elements should provide us with sufficient evidence to answer the question where on the Formality Continuum the best query interface solutions for the casual and occasional user lie.

References

1. I. Androutsopoulos, G. D. Ritchie, and P. Thanisch. Natural language interfaces to databases - an introduction. *Natural Language Engineering*, 1(1):29–81, 1995.
2. A. Bernstein and E. Kaufmann. Making the semantic web accessible to the casual user: Empirical evidence on the usefulness of semiformal query languages. *IEEE Transactions on Knowlwdge and Data Engineering*, under review.
3. A. Bernstein, E. Kaufmann, A. Göhring, and C. Kiefer. Querying ontologies: A controlled english interface for end-users. In *4th Intl. Semantic Web Conf. (ISWC 2005)*, pages 112–126, 2005.
4. A. Spink, W. Dietmar, B. J. Jansen, and T. Saracevic. Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001.
5. C. W. Thompson, P. Pazandak, and H. R. Tennant. Talk to your semantic web. *IEEE Internet Computing*, 9(6):75–78, 2005.