KnowledgeWeb
Realizing the Semantic Web
EU Network of Excellence FP6-507482
http://knowledgeweb.semanticweb.org

Scalability, Approximation & Performance

Introduction

A realistic realisation of the Semantic Web will see huge amounts of data being exposed and described using ontologies. Proposed languages for the Semantic Web such as RDF Schema and OWL are rooted in formal, logical languages. This provides a solid foundation, but can potentially introduce issues of scalability. The expected size of the future Semantic Web may well be orders of magnitude larger than anything that traditional logical inference mechanisms can deal with. These scalability issues are not due to any flaws in the design of OWL – high computational complexity is inherent in expressive knowledge representation and reasoning tasks.

Measuring the performance of tools assists users in making decisions about which tools are appropriate for their tasks. Benchmarks can support the effective measurement of tool performance.

This briefing highlights work undertaken by members of the KnowledgeWeb network in the provision of scalable reasoning technique including approximation, distributed reasoning and optimisation, and the investigation of benchmarking.

Approximation

Approximation techniques take inference problems and replace them with simpler inference problems in such a way that either the soundness or completeness (but not both) of the solutions is preserved. Relaxing soundness means that we are no longer guaranteed that all the answers we get are necessarily correct. Relaxing completeness means that we may be being given a partial solution, e.g. not all possible answers are being returned.

Approximate Instance Reasoning

Description Logics (DLs) provide the formal underpinnings for the Web Ontology Language OWL. Reasoning with Description Logic languages is often split into two: *TBox* reasoning essentially deals with schema or concept level inferences, while *ABox* reasoning deals with instances or individuals. State of the art DL reasoners are showing themselves to be capable of handling TBox reasoning tasks for large ontologies. However, most systems break down when presented with large amounts of ABox data.

Screech is a system that provides approximate ABox reasoning. In this case, Screech provides complete reasoning, but trades off soundness for time. The rationale for the approach is based on the observation that deductive databases have been shown to be efficient in dealing with large numbers of facts. Screech transforms an OWL T-Box into a disjunctive datalog program. This program is then passed to a disjunctive datalog reasoner along with the A-Box data.

Evaluations based on publicly available ontologies and data sets suggest that the technique is feasible.

Robust Query Answering

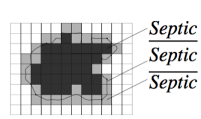
Co-operative Query Processing supports the user by rewriting queries. For example, if a query returns no results, *relaxing* it by replacing or deleting parts of the query can help to provide the "best" results meeting some criteria.

Such a relaxation approach has been demonstrated in the context of a Human Resources use case, where data exchange between employers, applicants and job portals is based on a set of vocabularies which provide shared terms to describe occupations, industrial sectors and job skills. It is often the case that a query for applicants matching a specific request returns no results. A principled rewriting can provide successive loosenings of the query, resulting in a collection of acceptable answers.

Multiple Perspectives

Ontologies are often not really designed independently of the task at hand, but are rather developed to meet the needs of a particular system or task. One solution to the problem involves the representation of different viewpoints on the same ontology, that better reflect the actual needs of the application at hand. Approximate reasoning based on "limited vocabularies" can provide support for such an approach.

Rough Description Logics



In some domains, concepts cannot be precisely defined, but instead are restricted by approximations. An approach known as Rough Description Logics (Rough DL) addresses these issues through an extension of the classical DL approach. A syntax and semantics has been defined, with reasoning tasks supported through classical DL reasoning.

The approach has been demonstrated in a proof of concept example modelling Sepsis in Clinical Trials using data taken from the Dutch National Intensive Care Evaluation.

In this example, it is difficult to provide criteria that characterise precisely when sepsis is present, and different clinical trials used different entry criteria for patient selection. However, upper and lower approximations can be given, allowing a comparison between trials, potentially spotting discrepancies or inconsistencies bewteen the selections.

Reasoning with Inconsistency

A semantic approach has been proposed for reasoning with inconsistent ontologies. Google distances can be used to develop semantic relevance functions to reason with inconsistent ontologies. The methods trade-off computational cost for inferential completeness, and provide attractive scalability. Test results using the PION system with realistic ontologies show that the semantic approach can significantly improve reasoning performance over a syntactic approach.

HermiT

HermiT is a theorem prover for expressive DLs, implemented using a novel hypertableau reasoning algorithm. The key aspect of this algorithm is that it is much less non-deterministic than the existing tableau algorithms, and thus offers the potential of a truly scalable DL reasoning system suitable for application in the Semantic Web. HermiT is still at the prototype stage, but preliminary results on large ontologies such as those used in medicine (for example the NCI ontology developed by the National Cancer Institute in the US) are encouraging.

Benchmarking

Benchmarking supports consistent measurement and analysis of the performance of tools and applications. Within KnowledgeWeb, attention has been paid to the issue of interoperability between heterogeneous applications. Although languages such as RDF, RDF(S) and OWL are standardised, tooling is still somewhat immature, and problems in interoperation have been reported.

A suite of benchmarks have been developed that test the interoperation of tools, in terms of their ability to exchange ontologies represented using OWL. The benchmarks are themselves described semantically, allowing automation of the benchmarking process.

Evaluation of the leading tools and APIs shows that, even with standardised languages, there are a number of areas where interoperation could be improved upon. However, the presence of a set of benchmarks can only serve to assist developers in producing conformant applications.

Benchmarks for tools that process inconsistent ontologies have also been developed, along with evaluations of a number of different approaches to handling inconsistencies – syntactic vs semantic; linear extension vs. multi-step extension; and debugging vs. reasoning.

Publications & Resources

P. Dolog, H. Stuckenschmidt, H. Wache. *Robust query processing for personalized information access on the semantic web* in Proceedings of FQAS, 2006.

P. Hitzler, D. Vrandecic. *Resolution-based approximate reasoning for OWL DL.* in Proceedings of ISWC'05, 2005.

H. Stuckenschmidt. *Towards multi-viewpoint reasoning in OWL ontologies*, in Proceedings of ESWC'06, 2006.

S. Schlobach, M. Klein, L. Peelen *Description Logics with Approximate Definitions - Precise Modeling of Vague Concepts* in Proceedings of IJCAI'07, 2007

S. Schlobach, Z. Huang, R. Cornet, F. van Harmelen, *Debugging Incoherent Terminologies*, to appear in Journal of Automated Reasoning.

KnowledgeWeb Deliverable D2.1.6.3. *Report on benchmarking of processing inconsistent ontologies.*

B. Motik, R. Shearer, I. Horrocks. *Optimized Reasoning in Description Logics using Hypertableaux* In Proceedings of CADE-21, 2007.

KnowledgeWeb Deliverable D1.2.2.1.2: *Benchmarking the interoperability of ontology development tools using OWL as interchange language*

Contact

For more information about the results presented here or the KnowledgeWeb Network of Excellence, please see the project web site

http://knowledgeweb.semanticweb.org

or contact the Network Manager.

Alice Carpentier
DERI, University of Innsbruck
Technikerstraße 13
Innsbruck 6020
AUSTRIA
alice.carpentier@deri.org

KnowledgeWeb was an EU-IST Framework 6 Network of Excellence.

KnowledgeWeb Consortium: Centre for Research and Technology Hellas, GR; France Telecom, FR; Free University of Bozen-Bolzano, DE; Freie Universität Berlin, DE; Institut National de Recherche en Informatique et en Automatique, FR; L3S Research Center, DE National University of Ireland Galway, IR; The Open University, UK; Universidad Politécnica de Madrid, ES; University of Innsbruck; AT University of Karlsruhe, DE; University of Liverpool, UK; University of Manchester, UK; University of Sheffield. UK; University of Trento, IT; Vrije Universiteit, NL; Vrije Universiteit Brussel, BE; École Polythechnique Fédérale de Lausanne, CH