UNIVERSITY OF OXFORD
COMPUTING LABORATORY

HermiT: Reasoning with Large Ontologies

Grade 7 Research Assistant post

FURTHER PARTICULARS

The objectives of the project are to develop reasoning services that will allow for the successful application of ontology technologies in large scale applications. Ontologies are formal vocabularies of terms, often shared by a community of users. One of the most prominent application areas of ontologies is medicine and the life sciences. For example, the Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) is a clinical ontology which is being used in the UK Health Service's National Programme for Information Technology (NPfIT). Other examples include GALEN, the Foundational Model of Anatomy (FMA), the National Cancer Institute (NCI) Thesaurus, and the OBO Foundry -- a repository containing about 80 biomedical ontologies.

These ontologies are gradually superseding existing medical classifications and will provide the future platforms for gathering and sharing medical knowledge. Capturing medical records using ontologies will reduce the possibility for data misinterpretation, and will enable information exchange between different applications and institutions.

Medical ontologies are strongly related to description logics (DLs), which provide the formal basis for many ontology languages, most notably the W3C standardised Web Ontology Language (OWL). All the above mentioned ontologies are nowadays available in OWL and, therefore, in a description logic. The developers of medical ontologies have recognised the numerous benefits of using DLs, such as the clear and unambiguous semantics for different modelling constructs, the well-understood tradeoffs between expressivity and computational complexity, and the availability of provably correct reasoners and tools.

The development and application of ontologies crucially depend on reasoning. Ontology classification, i.e., organising classes into a specialisation/generalisation hierarchy, is a reasoning task that plays a major role during ontology development: it provides for the detection of potential modelling errors such as inconsistent class descriptions and missing sub-class relationships. For example, about 180 missing sub-class relationships were detected when the version of SNOMED CT used by the NHS was classified using the DL reasoner FaCT++. Query answering is another reasoning task that is mainly used during ontology-based information retrieval; e.g., in clinical applications query answering might be used to retrieve "all patients that suffer from nut allergies".

Despite the impressive state-of-the-art, modern medical ontologies pose significant challenges to both the theory and practice of DL-based languages. Existing reasoners can efficiently deal with some large ontologies, such as NCI, but many important ontologies are still beyond the reach of available tools. For example, none of the existing reasoners can successfully classify either GALEN or FMA.

Furthermore, the amount of data used with ontologies can be orders of magnitude larger than the ontology itself. For example, the annotation of patients' medical records in a single hospital can easily produce data consisting of hundreds of millions of facts, and aggregation at a national level might produce billions of facts. Existing reasoners cannot cope with such data volumes, especially not if ontologies such as GALEN and FMA are used as schemata.

The aim of the HermiT project is to develop scalable reasoning algorithms and a prototypical implementation that can efficiently deal with large and complex ontologies and large data sets. Developing such a reasoner will be critical to the success of many ontology based applications. This will be achieved by:

- Extending our existing hypertableau-based reasoning algorithm to DLs that underpin modern

ontology languages, such as SHOIQ and SROIQ.

- Developing optimisation techniques that will further reduce and-branching and thus improve the performance of TBox reasoning.
- Developing new techniques for query answering by building on the optimisation techniques known from deductive and relational databases.
- Developing a new DL reasoner, based on the existing prototype, to extend it based on the above mentioned theoretical results, and to conduct an empirical performance evaluation.

Selection criteria

Essential:

- A first degree in Computer Science or related discipline.
- Applicants should have, or be about to obtain, a PhD in computer science or a related discipline and/or have some relevant postgraduate experience
- Programming experience, preferably in Java or C++.
- Interests or capabilities in logics and automated reasoning.
- Ability to give technical presentations and prepare reports/papers for publication.
- A willingness to collaborate with others and work effectively as a member of a team.

Desirable:

- Experience in the implementation of algorithms for automated reasoning.
- Experience of working in collaborative or interdisciplinary environments.

Salary and Benefits

Salary will be on the University grade 7 scale (currently £28,839 - £35,469 p.a). The post is available for an immediate start, will run until May 2011, is pensionable and includes an annual leave entitlement of 38 days per year, inclusive of public holidays and university closed periods.

Application Procedure

Applications should be in the form of a letter of application (clearly stating the post title) setting out how the candidate meets the selection criteria, and supported by a full curriculum vitae, together with the names and addresses of two referees. These should be emailed (most formats accepted) to job05@comlab.ox.ac.uk or alternatively, posted to:

The Administrator
Oxford University Computing Laboratory
Wolfson Building
Parks Road,
Oxford OX1 3QD

to arrive by the closing date of **Tuesday 20th January 2009.** Applications received after this time will not be considered.

Candidates must ask their referees to consider the further particulars and email the reference directly to job05@comlab.ox.ac.uk or, alternatively, post it to the above address (fax (+44 1865 283532) so that references arrive by the closing date.

The policy and practice of the University of Oxford require that all staff are offered equal opportunities within employment. Entry into employment with the University and progression within employment will be determined only by personal merit and the application of criteria which are related to the duties of each particular post and the relevant salary structure. In all cases, ability to perform the job will be the primary consideration. Subject to statutory provisions, no applicant

or member of staff will be treated less favourably than another because of his or her age, sex, marital or civil partnership status, sexual orientation, religion or belief, racial group or disability.

Applicants who would need a work visa if appointed to the post are asked to note that under the UK's new points-based migration system they will need to demonstrate that they have sufficient points, and in particular that:

(i) they have sufficient English language skills (evidenced by having passed a test in basic English, *or* coming from a majority English-speaking country, *or* having taken a degree taught in English)

and

(ii) that they have sufficient funds to maintain themselves and any dependants until they receive their first salary payment.

Further information is available at:

http://www.ukba.homeoffice.gov.uk/workingintheuk/tier2/generalarrangements/eligibility/.

All data supplied by applicants will be used only for the purposes of determining their suitability for the post and will be held in accordance with the principles of the Data Protection Act 1998 and the University's Data Protection Policy, but if the person appointed to the post is a migrant sponsored under the UK's new points-based migration system, we are required to retain all applications for the duration of the sponsorship.