School of IT
Technical Report

The University of Sydney

ACCRETION-BASED ONTOLOGY LEARNING FOR CONTEXT REASONING
TECHNICAL REPORT 608

JUDY KAY AND WILLIAM NIU

SCHOOL OF IT, THE UNIVERSITY OF SYDNEY

JANUARY 2007

Accretion-based Ontology Learning for Context Reasoning

Judy Kay and William Niu
School of Information Technologies
The University of Sydney
NSW 2006, Australia
+61-2-9351 5711
{judy, niu}@it.usyd.edu.au

ABSTRACT

Accretion-based Ontology Learner (ACOL) is a ontology learner within the SAGO framework. It accumulates evidence that contributes to each relationship between concepts picked up by the ontology learner. This means that, instead of having a fixed relationship between concepts, ACOL retains the evidence about the relationships and calculates, or resolves, the relationship according to a given context based on user, time, and/or location. This can be used to adapt information delivery to different contexts when there are multiple sources for (ontology) learning.

This will be deployed in the MyPlace system, which is a context-aware system designed to provide personalised information delivery in the building of School of Information Technologies at the University of Sydney.

Keywords: ontology learning, evidence accretion, personalised information delivery, context-aware

INTRODUCTION

There are numerous research projects on intelligent systems in ubiquitous computing environment, but it was only in the past three years or so that ontological reasoning had started to attract researchers' attention in this area [5, 1, 9, 10]. Systems then suffered from inadequate context ontologies for more general reasoning due to limitations of manually constructed ontologies. Manual ontology engineering is known to be resource- and time-consuming. There has been work to automate or semi-automate the process [6]. However, to date, no one appears to have applied ontology learning techniques in the UbiComp environments; this seems to be a promising research direction considering the vast number of data sources available to different aspects of UbiComp environments. We are currently adding ontological reasoning support to the MyPlace system [4], which is a context-aware system deployed in the building of School of Information Technologies at the University of Sydney. The aim is to semi-automatically learn and populate a scrutable light-weight ontology using multiple document sources from both domain-specific and general-purposed.

ACCRETION-BASED ONTOLOGY LEARNER

Based on the assumption that the same set of concepts may be related by different relationships in different contexts, we have been developing an ontology learner that accumulates evidence sources for each relationship found. We called it Accretion-based Ontology Learner, or ACOL. It is one of the algorithms in the SAGO (Scrutable Automatically Generated Ontology) suite [7], and evolved from the previous ontology learner MECUREO [2]. A characteristic of ACOL is its ability to retain evidence sources found for later contextualised resolution of relationships. To avoid the cold-start, or data sparsity problem many intelligent systems suffer, we would ideally like to use many document sources in order to make useful, contextualised resolution. These documents may be in different formats, but will need to either be domain-specific (e.g. a building manual for an ontology for a building) or general-purpose (e.g. OpenCyc).

The goal of the learned ontology is to provide contextualised information delivery. It may be made possible by accumulating evidence sources from different documents, especially domain-specific ones. Then, depending on the particular context (e.g. person, location, and time), adaptive information may be delivered to the user. From a system scrutability[1] point of view, as each piece of evidence retains the timestamp and the document source where it was found, it is possible to explain to users why the system presents what it presents and how the information is deduced. For example, when the MyPlace system indicates that Bob is in room 110, it can provide an explanation of such deduction: this is because Bob's mobile phone is detected by the bluetooth sensor in room 110 less than a minute ago.

Source Documents

The domain that we construct the ontology in is our School of Information Technologies building. Some document sources to be used are a part of OpenCyc[2], the building manual, the technical data for the building, handbooks for postgraduate and undergraduate students, and a directory for staff and postgraduate students. A part of OpenCyc has been man-

[1]The word *scrutable* is defined in the American Heritage® Dictionary of the English Language as "Capable of being understood through study and observation; comprehensible."

[2]http://www.cyc.com/opencyc, 23 Nov 2006.

Room	Name	Principal activities	Primary user	Essential links	Desirable links
103	technical workshop	office	technical staff	t1 manager's office	workshop store
220	reception	reception	null	lift, lobby	null

Table 1: Example entries from the technical building data.

ually extracted to form a base ontology which the location ontology is based on [8]. This ontology will serve as one of the existing domain-specific ontologies to be populated upon. Ideally ACOL can take in as many ontologies, both general-purposed and domain-specific, as available and helpful. It may be difficult, in terms of performance, to even take in the full version of OpenCyc.

Both the staff/postgraduate student directory and the technical building data have been parsed into a semi-structured format, i.e. comma separated values. The directory contains basic information for academic, administrative staff and postgraduate students, including their offices or workspace locations, contact information, as well as research and administrative roles they have (e.g. undergraduate coordinator). This may help populate the ontology about users of the building, and can link the users to other ontologies like the location ontology. The building data, on the other hand, provides information about each location in the building and how it was intended to be built. Table 1 illustrated the kinds of data about a location that can be found in the technical building data. Some of the useful ones (i.e. in terms of ontology learning) are the name, normal activities, primary users, as well as essential and desirable near-by locations. These can enrich the location ontology with more inter-relationships and more knowledge pertaining to each location. One aspect to note is that both the directory and the building data suffer from some out-of-dated inaccuracies, e.g. some actual near-by relationships between locations were not built as planned.

The building manual and handbooks for students are both treated and parsed as glossary-like documents. A distance-based algorithm with part-of-speech tagging similar to the approach described in [2] is adopted to extract the relationships in those documents. Instead of amalgamating the evidence sources found and calculating a single weighted relationship for any two concepts, ACOL accumulates and retains each of the sources. Next section gives a more detailed overview of this approach. The manual describes various things, including policies, services, locations, and groups of people in the building. For example, it tells where a fax machine can be found, and where the reception desk is located and its opening hours. The students are divided into three groups: undergraduate, postgraduate coursework, and postgraduate research. There is a handbook for each of the groups outlining policies, resources, and other relevant information for students in that group. Both sources may potentially provide inter-relationships between locations, people, and services.

DISCUSSION & FUTURE WORK

Evaluating a learned ontology is a difficult task, as it is often not possible and/or practical to efficiently quantify the quality of the ontology. That results in a diversified range of evaluation methods depending on the nature of the ontology and its applications. Some common approaches are golden standard, application-based, data driven, and human assessment [3]. As the size of documents being used in this project is relatively small, it may be possible to carry out manual assessment by ontology engineers and/or domain experts on the ontology according to a set of pre-defined criteria and requirements. Since ACOL is the next version of MECUREO, it also makes sense to compare the difference in accuracy and performance between those two approaches.

On another aspect, we will need to implement a reasoning engine in order to make inference about the ontology to be deployed in MyPlace. Some popular reasoning engines for OWL are RacerPro[3], Jena[4], and Pellet[5].

ACKNOWLEDGEMENTS

This project is partially funded by Smart Internet Technology Cooperative Research Centre (SITCRC). We would also like to thank David J Carmichael for converting the technical building data into a CSV file.

REFERENCES

1. A. Agostini, C. Bettini, and D. Riboni. Loosely coupling ontological reasoning with an efficient middleware for context-awareness. In *Mobile and Ubiquitous Systems: Networking and Services, 2005. MobiQuitous 2005. The Second Annual International Conference on*, pages 175–182, July 2005.

2. T. Apted and J. Kay. MECUREO: Ontology and modelling tools. *WBES of the International Journal of Continuing Engineering Education and Lifelong Learning*, 14(3):191–211, 2004.

3. J. Brank, M. Grobelnik, and D. Mladenić. A survey of ontology evaluation techniques. In *SIKDD 2005 at multiconference IS 2005*, 2005.

4. D. Carmichael, J. Kay, B. Kummerfeld, and W. Niu. MyWorkPlace: Personalised information about a ubiquitous computing enabled building. In *Proceedings of the Second International Workshop on Personalised Context Modelling and Management for UbiComp Applications at UbiComp 2006*, 2006.

5. H. Chen, T. Finin, and A. Joshi. An ontology for context-aware pervasive computing environments. *Knowledge Engineering Review*, 18(3):197–207, 2004.

6. A. Gòmez-Pérez and D. Manzano-Macho. OntoWeb deliverable 1.5: A survey of ontology learning methods and techniques. website, 2003. OntoWeb Deliverable.

[3] http://www.racer-systems.com/, 24 Nov 2006.
[4] http://jena.sourceforge.net/inference/, 24 Nov 2006.
[5] http://www.mindswap.org/2003/pellet/, 24 Nov 2006.

7. J. Kay, A. Lum, W. Niu, and L. Li. TARO: Teaching with an automatically retrieved ontology. Technical Report 596, The University of Sydney, Sept 2006.

8. J. Kay, W. Niu, and D. J. Carmichael. ONCOR: Ontology- and evidence-based context reasoner. In *IUI '07: Proceedings of the 12th International Conference on Intelligent User Interfaces*, New York, NY, USA, 2007. ACM Press. to appear.

9. A. Ranganathan, R. E. McGrath, R. H. Campbell, and D. Mickunas. Ontologies in a pervasive computing environment. In *Workshop on Ontologies in Distributed Systems, IJCAI 2003*, 2003.

10. X. H. Wang, D. Q. Zhang, T. Gu, and H. K. Pung. Ontology based context modeling and reasoning using owl. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, 2004.

School of Information Technologies, J12
The University of Sydney
NSW 2006 AUSTRALIA
T +61 2 9351 3423
F +61 2 9351 3838
www.it.usyd.edu.au

ISBN 1 86487 886 X