RETHINKING THINKING:
USING ONTOLOGIES FOR E-TECHNOLOGY ENABLED INTEGRATED PROBLEM SOLVING

CHEN, Zhengxin
Department of Computer Science
University of Nebraska at Omaha
Omaha, NE 68182-0500, USA
zchen@mail.unomaha.edu

Abstract

Real-world problem solving requires integrated problem solving skills. However, in reality, various problem solving techniques have been covered in separate courses in computer science (or other Information Technology-related) curriculum, leaving little time for students to apply learned knowledge together. The advent of E-technology provides an excellent opportunity to change this. In particular, in this paper we take a look at the issue of using ontologies to enable integrated problem solving. We start with a general review on ontologies, pointing out the dual role of ontologies as an e-technology in education. We then focus on the role of ontologies for integrated problem solving by reviewing our recent experiences from a number of student projects in database management system (DBMS), artificial intelligence (AI), data mining (DM) and other senior level courses, pointing out the important role of ontology as a common theme to thread various aspects related to intelligent information systems. These projects have shown exploring ontology-related aspects is an effective way for students to develop skills in integrated problem solving.

Keywords: *intelligent problem solving, ontologies, E-technology in education*

ACM classification: K.3.2, I.2.0, H.2.8

1. **Introduction**

It has long been noted that integrated heuristic decision making is the key for success of intelligent problem solving. Several levels of the integration have been identified (Chen, 1999):

- *Integrated tools*: Here integration is taken care by the commercial tools themselves. This level of integration is convenient for the users, but it offers little flexibility and controllability to them.
- *Integrated use of existing methods*: Here the methods have been developed, but the task of integration is left to the user.
- *Integrated thinking*: This is the most advanced level of integration, and is the most important idea deserving endorsement. *Integrated thinking* refers to an

ability of problem solving guided by appropriate heuristics to employ existing techniques or to develop new methods.

In this paper, we apply this line of thinking to integrated intelligent problem solving process, which involves building intelligent agents (Russell and Norvig 2002). Bienkowski (1998) characterized an agent as "rethink thinking: autonomy, environmental interaction, and reaction." Integrated thinking is the most important aspect of this kind of rethinking.

However, in reality, various problem solving techniques have been covered in separate courses in computer science curriculum, leaving little time for students to apply learned knowledge together. The advent of E-technology provides an excellent opportunity to change this. In particular, in this paper we take a look at the issue of using ontologies to enable integrated problem solving. We start with a general review on ontologies, pointing out the dual role of ontologies as an e-technology in education. We then focus on the role of ontologies for integrated problem solving by reviewing our recent experiences from a number of student projects in database management system (DBMS), artificial intelligence (AI), data mining (DM) and other senior level courses, pointing out the important role of ontology as a common theme to thread various aspects related to intelligent information systems. These projects have shown exploring ontology-related aspects is an effective way for students to develop skills in integrated problem solving.

The rest of this paper is organized as follows. In Section 2 we review basics of ontology and the role of ontologies as an enabling e-technology in education. In Section 3 we summarize student projects illustrating four important aspects related to ontologies. We provide a discussion and conclude our paper in Section 4.

2. **Basics of ontologies and their roles in education**

To make this paper self-contained, we start with a brief review on the basics of ontology.

In information science, an ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain. Ontologies thus serve as common vocabularies to unite various scientific communities. Ontologies are used to describe real-world concepts and knowledge modeling in a Hierarchical structure. A well known quote from Gruber indicates that an ontology "is a formal explicit specification of a shared conceptualization." In practice, "shared conceptualization" is widely accepted by many groups of researchers and practitioners.

Ontologies allow for:

- A common shared vocabulary for data normalization;
- Machine interpretable in standardized format;
- Data fusion: integrating information;
- Semantically annotate data with concepts;
- Query, retrieve and Index information semantically;
- Use synonyms for identifying concepts within text;
- And others.

In our view, the role of ontologies as an enabling technique in e-technology in education is two-fold: (1) using ontologies to develop teaching software/tools and improve teaching; and (2) using ontologies to unify teaching contents for computer science (or IT) education. The first role has been widely acknowledged for some time. For example, in a discussion on Ontology-based E-Education (http://asusrl.eas.asu.edu/ontoEducation/), it has been noted that Ontology has received significant attention in the education community because it makes possible the specification of knowledge including concepts, relationships, classification, and reasoning. Mizoguchi et al. (1997) pointed out that applying ontology to education brings five main advantages, namely: making the education system smart and reflective; delivering explicit knowledge; standardizing vocabulary; ease of communication; and making knowledge reusable. In addition, Breuker and Muntjewerff (1999) proposed the use of ontology in education for indexing and specifying various relationships and structure within the knowledge domain, while Allert et al. (2006) discussed the use of ontologies in learning under the context of knowledge-creation metaphor of learning, which conceptualizes learning and knowing as a social process where people collectively improve their understanding by generating shared knowledge artifacts. However, the second role of ontologies as we identified above, namely, using ontologies to unify learning contents for integrated problem solving, deserves more attention. Below we focus on how to use ontologies as a common theme for integrated problem solving.

3. Sample aspects of ontologies for integrated problem solving

Contributions of ontologies for computer science and IT-related problem solving are due to the following two facts: First, various ontologies such as WordNet (http://wordnet.princeton.edu/) and domain ontologies such as Gene Ontology (http://www.geneontology.org/) are widely available through the Internet, so that they can be used in many applications. Secondly, there are additional e-technology tools to construct domain ontologies on-demand, which can be incorporated into or combined with existing ontologies. Below we describe several course projects involving various aspects of ontologies, which can be viewed as case studies as well:

1. Ontology construction from text;
2. Using ontologies for intelligent information system query answering, such as query relaxation;
3. An ongoing project on knowledge-driven, ontology-guided multi-criteria data mining.

3.1. *Domain ontology construction*

Typically, ontologies are difficult and labour-intensive to create. In order to acquire domain knowledge needed for ontologies, we need domain experts, as well as domain information. Ontology construction from texts deserves particular attention as they provide the largest source of information on Web. Texts in specific knowledge areas form the domain corpus and provide a model of the domain. These considerations justifies why the concept of "ontologies on demand" is so attractive, because it will allow us to quickly construct domain-specific ontologies for knowledge management.

Our understanding of the overall process for ontology construction from texts is summarized in Fig. 1. Construction of ontologies from text often includes many complex sub-tasks occurring within a pipe-lined fashion. Initially concept and relationship discovery is first applied to the result obtained from using NLP tools. From there, various discovery algorithms (including lexico-syntactic pattern discovery, Noun-Verb-Noun patterns discovery, word frequency discovery, as well as association rules, etc.) are applied, and are incorporated with seed ontology (or bag of words), to build domain concepts and relationships.

We have instructed students to collect and select avian/swine flu papers in bioinformatics domain, and used Stanford NLP tool to process the texts. We also provide information to students to visually present the obtained ontology by using the open source package Graphviz, which creates images of graphs that are described by the 'dot' programming language. Thus in our process, we had to convert concepts into nodes and their relations/connections into arrows described in the dot language syntax. A small part of the result is shown in Fig. 2. This project is important to students, not only because it demonstrates that domain ontology can be constructed by themselves, but also because the obtained results can be used in other projects related to bioinformatics.

3.2. *Ontology-guided query relaxation*

Query relaxation refers to automatic generalization of the given query to better suit user's information needs, particularly dealing with the failing query problem: given a query that returns an empty answer, how can one relax the query's constraints so that it returns a non-empty set of tuples? By generalizing a query to capture "neighboring" information, query relaxation has been an effective means for fully exploit users' information needs from databases and information retrieval systems.

Similar to query optimization, query relaxation requires a kind of rewriting of the original query. However, although query optimization is aimed at improved system performance, query relaxation is aimed at better user satisfaction for the contents returned; in particular, in case of empty result, the system will try to find an approximate match. Unlike query optimization which makes use of a set of syntactical transformation rules to rewrite original queries, query relaxation pays attention to semantics via background knowledge for query rewriting. Therefore, query relaxation becomes an interesting "showcase" for intelligent query answering.

In a recent student project assignment, students have been asked to investigate one aspect of making use of background knowledge for query relaxation, i.e., the role of using ontology for XML database query relaxation. In particular, we have explored a practical approach which incorporates WordNet with free software Qexo. Rather than reinventing the wheel, we take advantage of publicly available tools to make query relaxation easy to implement. By integrating with WordNet which serves as ontology, the GUI front-end can assist effective XML query relaxation. Here is an example. The user submitted a query and asked for 5 desired answers. However, this query obtained empty results, since the tag proprietor as appeared in the query does not appear in the XML file. To deal with this failed query, the system now invokes WordNet, and based on the synset provided by WordNet, rewrites the original query to rename the node proprietor as owner.

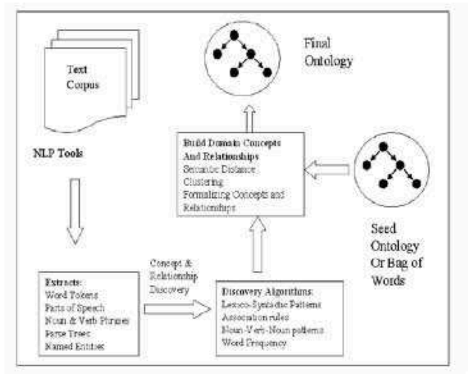

Figure 1. *Ontology construct methods from text*

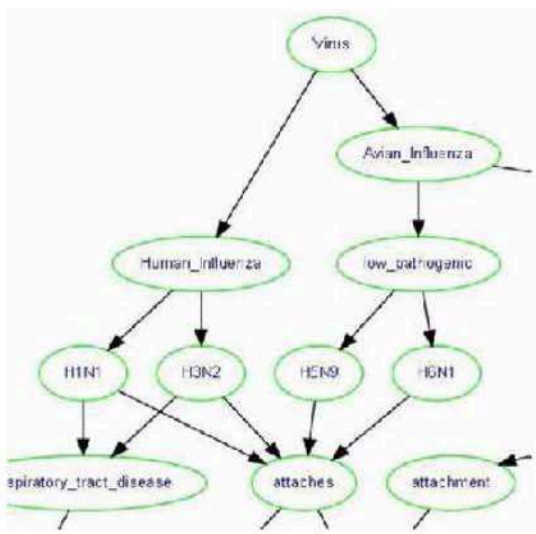

Figure 2. *Part of ontology constructed*

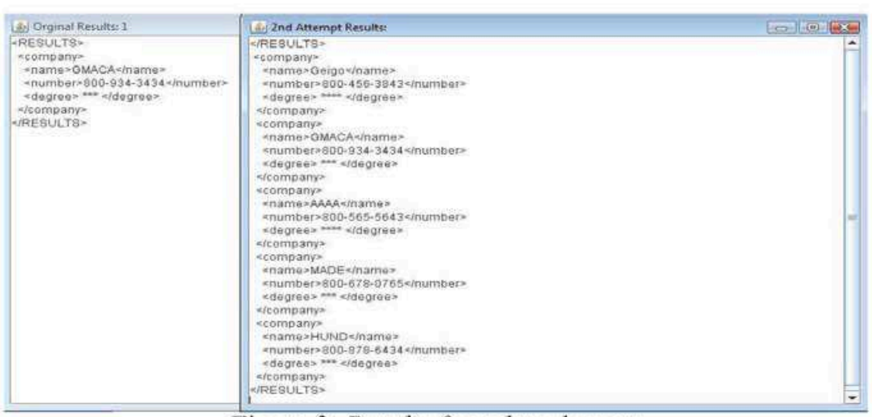

Figure 3. *Results for relaxed query*

The rewritten query is executed automatically, but with only one result returned. This indicates further query relaxation is still needed. By automatically applying extended value constraint method, the system expands the where clause by adding disjunctives with WordNet synset for "car." After automatically execute the relaxed query, the system returned 5 results (as shown in Fig. 3), which satisfies the user's information need.

3.3. *Ontology-guided multi-criteria data mining for business intelligence*

Recently data mining has widely appeared in computer science curriculum. However, in many cases the data mining course covers basic algorithms and applications. Data mining has been developed, thought vigorously, under rather ad hoc and vague concepts. To remedy these problems, the research of foundations of data mining (FDM) (Tsumoto et al., 2002) is intended to establish a solid mathematical/logical foundation for data mining, to integrate its various functionalities/tasks and related activities, and to make data mining a true scientific and engineering discipline. In particular, we promote to take the multi-objective perspective to examine data mining, because it will not only have the potential of offering new algorithms, but also provide a better understanding about the nature of data mining, an important issue concerned in the sub-area of data mining (FDM).

More importantly, following the lead of Yi and Kou (2008), we have started conducting projects based on a framework of integrating domain knowledge and multi-criteria optimization-based data mining methods, where the three major aspects identified are all firmly based on domain ontologies. Below is a brief description of three student sub-projects (corresponding the three major aspects) originated from the collaboration of our research lab with a major local retail business on enhanced business intelligence (BI):

1. Knowledge-driven feature selection: The domain ontology involves customer shopping behavior and the specific industry involved in this business. Overall, around 70 attributes have been identified, making feature selection a critical step. This sub-project incorporates manually constructed domain ontology, as well as generic ontology (such as WordNet) into several existing feature selection algorithms.
2. Data transformation based on domain knowledge: As noted in Yi and Kou (2007), since multi-criteria optimization methods require strictly numeric inputs, appropriate encoding schemes can affect the final classification results dramatically. For this purpose, the framework develops a set of data transformation rules for multi-criteria optimization methods to guide users find appropriate transformation rules. In this sub-project, the domain ontology as mentioned in (1) is incorporated into these data transformation rules involving different data granularities for conducting data mining at various levels.
3. Expert knowledge-based results interpretation: In order to achieve business users' requirements, the framework calls for a user interface to facilitate the communication between data miners and domain experts. In our sub-project, the domain ontology, combined with online analytic processing (OLAP) technology, is used as a useful guide for development of visualization and verbal explanations.

4. **Discussion and conclusion**

In this paper we have focused on using ontologies for E-technology enabled integrated problem solving. Another very important aspect of applying e-technology is learning environment, particularly for virtual learning. In fact, virtual learning environment presents both excellent opportunities as well as exciting new challenges for integrated intelligent problem solving. Due to the complexity of this task, we would conduct it in two major steps: In parallel to the study of integrated intelligent problem solving as reported in this paper, we have also conducted research work related to virtual learning using swarm intelligence (SI) techniques. But the details are not reported here.

In summary, in this paper we discussed the important issue of developing student projects integrated intelligent problem solving, focusing on using general or domain ontologies as a high-level meta-heuristic. We also discussed lessons learned, and relationship with other aspects related to computer science education. We are continuing our effort of guiding student course projects, and we also believe serious efforts must be made for development of integrated problem solving skills.

Acknowledgements

We thank students enrolled in CSCI8350 and CSCI8390 during 2007-2010, as well as several other students in CS department of University of Nebraska at Omaha. Partial support was also provided by grant from National Natural Science Foundation of China (#70901011).

References

1. Allert H., Markannen H., and Richter C., *Rethinking the Use of Ontologies in Learning*, "Proc. 2nd International Workshop on Learner-Oriented Knowledge Management and KM-Oriented Learning (LOKMOL 06)", M. Memmel and D. Burgos (eds.), 115–125, October 2006.
2. Bienkowski M. A., *A Reader's Guide to Agent Literacy*, "SIGART Bulletin", 9(2), 23-28, Fall 1998.
3. Breuker J. & Muntjewerff A., *Ontological Modelling for Designing Educational Systems*, "Technical Report of University of Amsterdam", June, 1999. http://www.ei.sanken.osaka-u.ac.jp/aied99/a-papers/J-Breuker.pdf.
4. Cartelli, A., *Semantics, Ontologies and Information Systems*, "Education: Concerns and Proposals, Issues in Informing Science and Information Technology", 3, 2006.
5. Chen Z., *Computational Intelligence for Decision Support*, CRC Press, 1999.
6. Guber T. R., *Towards Principles for the Design of Ontologies Used for Knowledge Sharing*, "Int. J. Human-Computer Studies", 43(5-6), 907-928, 1993.

7. Mizoguchi R., Ikeda M., and Sinitsa K., *Roles of Shared Ontology*, "AI-ED Research: Intelligence, Conceptualization, Standardization, and Reusability", 1997. http://www.ei.sanken.osaka-u.ac.jp/pub/miz/miz-aied97.pdf
8. Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice Hall, 2002.
9. Tsumoto, S., Lin, T. Y. and Peters, J. F., *A Domain Knowledge-driven Framework for Multi-criteria Optimization-based Data Mining Methods,* "Proc. COMSAC'02", 2002.
10. Yi, P. and Kou, G., *A Domain Knowledge-driven Framework for Multi-criteria Optimization-based Data Mining Methods,* "Proc. 4th Int'l Conf. Networked Computing & Adv. Info. Mgmt", 46-49, 2008.