International Journal of Engineering Research & Management Technology ISSN: 2348-4039

March- 2015 Volume 2, Issue-2

Email: editor@ijermt.org www.ijermt.org

Tools and Methods for Extracting and Developing Ontology

Vaishali Tyagi, Shilpi Saxena
M. Tech Student
Department of Computer Science & Engineering
SRM University, NCR campus
Modinagar

Divya Gupta
Assistant Professor
Department of Computer Science & Engineering
SRM University, NCR campus
Modinagar

ABSTRACT:
Ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. It provides a standardized vocabulary for that domain and the relationships between those concepts. Extracting information helps the users to salvage their needed information from the bulk of knowledge and ontology techniques can be used for extracting creative information. Majority of educators, being non-technically savvy, experience difficulties in using conventional ontology editors and thus, may benefit from the suggested ontological elements that would provide them with some starting points in the process of ontology building and extensive amount of domain related material is typically available in the form of electronic texts, slides, handouts, and the like, which can be used as an input for the ontology learning tools. This study specifically aimed at investigating how the utility and usability of the ontology extraction tools is perceived as well as how to perceive the quality of the resulting ontologies and to evaluate the approaches used and readiness of the ontology extraction tools to be used for the development of domain ontologies that meet the identified set of requirements specific for advanced e-learning systems.

Keywords: Ontologies, Text2Onto, OntoGen, TEL, Semantic web

1 INTRODUCTION

Ontologies, a mean for formally expressing the shared semantics of a certain domain, are the cornerstone of any Semantic Web-based solution. In previous researches, the advantages of ontology supported e-learning systems have been explored. In particular, combined use of content structure ontology, content type ontology, and domain ontology and how this could significantly improve the search over learning content repositories was demonstrated. It is also shown that if these three kinds of ontologies are complemented with a user model ontology and an ontology formally specifying the learning path to be followed by a student, then advanced levels of learning content personalization can be achieved [4], as well. Furthermore, the relevancy of the integrated use of these different kinds of e-learning ontologies for providing online educators with reliable, fine grained, and semantically rich feedback about the learning process [5], [6], [7]. Finally, in most recent research, use of ontologies and Linked Data paradigm to develop a personal learning environment for collaborative learning of software design patterns is made [8]. The main problem with all the approaches that make use of ontologies to offer advanced TEL (Technology Enhanced Learning) solutions is their assumption that the required ontologies are available or easy to develop. However, based on the experience of the TEL research community [1], [2] this assumption is not realistic. The major obstacle for widespread use of ontologies in e-learning systems lies in the complexity of the ontology development process, especially when considered from the perspective of educators who are typically unaware of ontology existence and its relevancy altogether. Although, recently, the Semantic Web research has shown a constantly increasing interest in automating ontology development and thus reducing the required human effort [9]

fully automatic ontology development is still in the distant future [10]. In this regard, domain ontologies, i.e., ontologies formally specifying concepts and relationships of a specific subject domain (e.g., a learning course), are the most challenging. Some researchers in the field have proven that unlike other kinds of ontologies relevant for e-learning, domain ontology often cannot be directly reused even within the same subject domain. This is due to different requirements of the systems that make use of domain ontologies. Therefore, domain ontologies often have to be either created a new or adapted to the needs of the specific context of their intended use. Another feature that makes domain ontologies distinct from other aforementioned ontologies relevant for e-learning is the need for their constant evolvement, so that the semantics they capture do not lag behind the courses they are aimed to support. Accordingly, a significant topic in current research is to investigate how to reduce the efforts required for creating domain ontologies , and thus implicitly enable easier and wider acceptance of ontology-based systems. First step toward achieving this goal is to explore the existing approaches for ontology development. Having done that, three general approaches are distinguished:

1) Handcrafting ontologies from scratch.
2) (semi)automatic ontology development using ontology learning tools.
3) search and retrieval of ontologies from online ontology libraries and Linked Open Data (LOD) cloud.

The second step towards this is to define the requirements that domain ontology should satisfy in order to serve as a foundation for the development and functioning of an advanced e-learning system. This requirement should be derived from previous research work in the area of personalized learning supported by Semantic Web technologies (e.g., [3], [4], [5], [7], [8]) as well as the work of other researchers in this area.

2 LITERATURE SURVEY

The ontology comprises all relevant concepts of the corresponding subject domain (i.e., the study course). The more detailed the ontology (i.e., the more domain specific concepts it has), the better, since an e-learning system would be able to provide better (more accurate) personalization [5]. The ontology comprises relationships between domain concepts. The ontology provides good coverage of the entire course, meaning that it does not cover just one part of the course, while other parts are just barely covered or not covered at all. High semantic richness (in terms of ontology axioms) is a desirable but not a mandatory characteristic since advanced functionalities of an e-learning system (e.g., recommendation of learning resources, and provision of feedback) can be achieved with a rather simple domain ontology [3], [4], [6]. In the study, that is the focus of this paper we investigated the (semi)automatic ontology development.

Even after a comprehensive literature review (including, for example, [9]), any research is not identified which aim at evaluating the level of adoption of tools for (semi)automatic ontology development among end users and identifying the requirements for enabling their widespread use. The closest work was the evaluation study done by Park et al. [10]. It was based on a comprehensive framework the authors proposed for the evaluation of ontology extraction tools. The experiment consisted of two parts: first, using the proposed evaluation framework, the authors themselves assessed four ontology extraction tools selected for the study; in the second part, four expert users assessed and ranked the considered tools using the Analytic Hierarchy Process method. However, the proposed evaluation framework and the experiment did not cater for the specific requirements of e-learning systems and their users. Furthermore, the assessment of the tools was done by ontology researchers and not real end users (i.e., educators who are typically unaware of ontologies). Accordingly, this paper evaluates ontology extraction tools taking into account both the requirements of e-learning practitioners and constraints imposed on ontologies by advanced e-learning systems.

3 THEORITICAL BACKGROUND

3.1 Ontology Tools

Aiming to eliminate one of the major obstacles to wider acceptance and deployment of ontology based software solutions, the Semantic Web community has devoted a significant attention to development of tools that would facilitate and scaffold the ontology development task.

Following three categories are there for the available tool support:

1. Tools for handcrafting ontologies.
2. Tools supporting reuse of existing ontologies, and
3. Tools for (semi)automatic development of ontologies.

3.1.1 Tools for Handcrafting Ontologies

Manual creation of ontologies using a specialized ontology editor is still the predominant approach in ontology development. Its main drawback is the fact that the majority of ontology editors are suitable only for experts in the field of ontology development.

3.1.2 Tools Supporting Reuse of Existing Ontologies

Supporting tools are needed to facilitate the searching process and evaluation of the retrieved ontologies.

3.1.3 Tools Supporting (Semi) Automatic Ontology Development

These tools aim at reducing the human intervention to supervision of the development process and refinement of the results. They offer users with suggestions for ontology elements, typically based on the analysis of existing domain documents.

3.2 Ontology

Ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts.

3.2.1 Use of Ontology

Ontologies are used as a form of knowledge representation about the world or some part of it. As ontology defines the concepts and relationships within a domain, it provides a standardized vocabulary for that domain and the relationships between those concepts. Ontologies are now central to many applications such as scientific knowledge portals, information management and integration systems, electronic commerce, and semantic web services.

3.3 TYPES OF ONTOLOGIES

3.3.1 Upper Ontology

Concepts supporting development of an ontonlogy (meta-ontology)

3.3.2 Domain Ontology

Concepts relevant to a particular topic or area of interest, for example, information technology or computer languages, or particular branches of science.

3.3.3 Interface Ontology

Concepts relevant to the juncture of two disciplines.

3.3.4 Process Ontology

Inputs, outputs, constraints, sequencing information, involved in business or engineering processes.

4 METHODOLOGIES

4.1 Generate and Purge Approach

The Generate and Purge Approach is exemplified by the Text2Onto tool [11] (Fig. 1). In this approach, a user first selects and loads a set of input documents (area 1). This results in a list of terms being extracted from the documents that may or may not be shown to the user (area 2). Next, one or more algorithm(s) is/are selected for extracting a particular ontological feature(s) (in area 3 of Fig. 1, algorithms for Concept and Relation algorithms are selected). After running the algorithms, the selected terms (or phrases) are presented to the user as suggestions, typically accompanied by some additional information such as the weights indicating the relevancy of the terms (area 4). The user is then asked to accept or reject these suggestions (area 5) to narrow them down to those to be included in the resulting ontology. Another well-known tool that belongs to this category is OntoLT which was developed as a plug-in the for the popular Prote´ge´ ontology development tool. It makes use of mapping rules to extract concepts and properties from linguistically annotated text collections. However, OntoLT does not offer support for the linguistic annotation; for that, it depends on another proprietary tool. Fig. 1. Mapping rules are included in the tool, and users can create new ones (provided that they master the precondition language that OntoLT uses for defining mapping rules). Onto Builder relies on the Generate and Purge approach to enable the extraction of ontology concepts and properties from WebPages. The extraction is based on the heuristic rules that the tool learns from a training set of HTML documents. After being given the URL of a webpage, Onto Builder identifies the elements of the page, and then generates a dictionary of terms by extracting labels and fieldnames from the webpage; it also recognizes relations among the extracted terms.

4.2 Build Incrementally Approach

In the Build Incrementally approach, a system helps users to build an ontology step by step by suggesting concepts and relations that can expand the selected node of the ontology based on the analysis of the underlying document corpus. Onto Gen tool [12] (Fig. 2) implements this approach. The user starts by loading the documents into the tool. The starting point for building ontology is the root node. In the unsupervised approach, the user can specify how many concepts the tool should suggest (area 1) as sub nodes of a selected node (area 2). The tool generates suggestions and for each one presents some information (e.g., keywords describing a suggested concept and the number of documents from the corpus that contain the respective concept) that should help the user to make a decision about whether to include the concept or not. In Fig. 2, four suggestions for the selected (highlighted) node have been accepted by the user and included in the ontology graph. In this way, starting from the root node, the user iteratively expands the nodes until he/she is satisfied with the ontology. DODDLE-OWL is another example of an ontology extraction tool. Although the tool initially offers users a list of extracted concepts, similar to the Generate and Purge approach, it actually relies on the Build Incrementally approach. After extracting a set of terms from the given domain specific corpus, the tool requires from the user to select relevant terms and identify their meaning by mapping each term to the corresponding concept in Word Net. Then, based on the user's input and by referring to the reference ontologies and documents, the tool generates an initial concept hierarchy and a set of concept pairs. The user then refines the initial ontology through the interactive support offered by the tool.

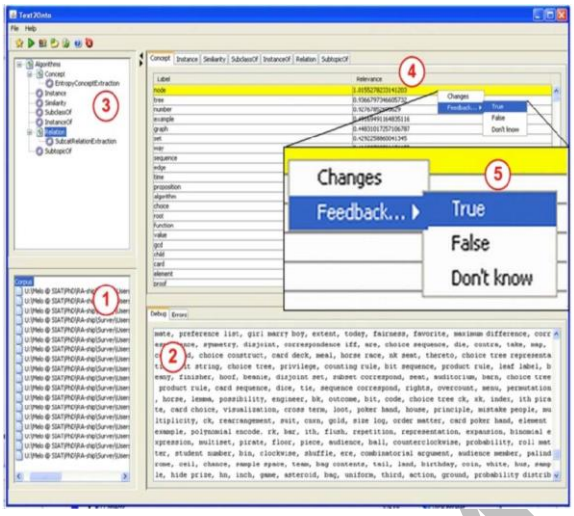

Fig.1. Generate and Purge Approach in Text2Onto tool

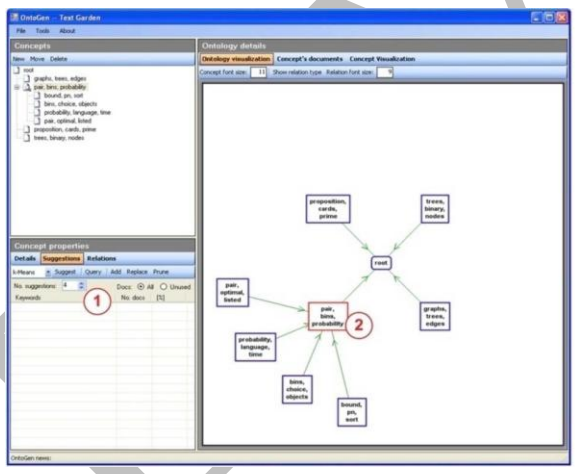

Fig.2. Build Incrementally approach in onto Gen Tool

4.3 Methodologies for Ontology Development

To produce high-quality result (i.e., ontology), an ontology development process has to be driven by a sound methodology. The Semantic Web research community has put a significant effort in the development of methodologies for ontology development; Corcho et al. [28] give a nice overview of the developed methodologies, whereas Simperl et al. [13] report on their actual usage in ontology development projects, based on an empirical study of a large number of both academic and industry projects. In the context of this paper, the most relevant is the methodology proposed in [14] since it targets the ontology development process based on the use of ontology extraction tools. The proposed methodology recognizes several phases in the ontology development process and for each phase gives a detailed description of input and output elements, activities to be performed, supporting tools to be used, and decisions to be made. It thus provides domain experts with detailed guidance for selecting and preparing information sources, applying ontology learning tools, and evaluating the learned ontology.

5 CONCLUSIONS

This study was driven for the following three main reasons:

1. To perceive the utility of the existing ontology extraction tools: first identify the elements that contribute to the user's overall perception of the utility of an ontology extraction tool. Then, investigate the perceived utility of the tools.
2. To perceive the usability of the existing ontology extraction tools.
3. To perceive the quality of the resulting ontologies because the perceived quality of the resulting ontologies indirectly indicates the perceived usefulness of the tool for the end users.

This paper presents two ontology extraction tools to build domain ontologies for their courses from their course materials. The two tools implement the two most widely explored and adopted approaches to ontology generation from documents, namely, Generate-and-Purge (Text2Onto) and Build Incrementally (OntoGen). Several conclusions can be made and suggestions given with respect to the explored approaches and desirable features of the tools. First, there is an appeal for the approach that generates a large number of suggestions for ontology concepts and relationships that are then "weeded out" by the user. This approach, applied by the Text2Onto tool, was especially favored by the non-IT group. However, having examined the produced ontologies from the perspective of the requirements of advanced e-learning technologies, their utility is found to be rather minimal. The requirements for visualization and editing of ontologies generated using ontology extraction tools, point to the need for comprehensive ontology development environments that would offer all those functionalities to the end users. Finally, an ontology building tool should have built-in evaluation metrics for the quality of the ontology being developed, so that it can provide users with immediate feedback on how good their ontology is and some guidance for improving it. In this study, the definition of ontology quality and the metrics for assessing it on the findings of software quality research for similar type of software artifacts was based. These findings will be beneficial not only for future research but also for the future work of other researchers in the area.

REFERENCES

1. V. Devedzic, Semantic Web and Education. Springer, 2006.
2. Semantic Web Technologies for E-Learning, D. Dicheva, R. Mizoguchi, and J. Greer, eds. IOS Press, 2009.
3. Z. Jeremic, J. Jovanovic, and D. Ga_sevi_c, "Personal Learning Environments on Social Semantic Web," Semantic Web J., accepted, http://www.semantic-web-journal.net/sites/default/ files/swj183.pdf, 2012.
4. J. Jovanovic, D. Ga_sevi_c, and V. Devedzic, "TANGRAM for Personalized Learning Using Semantic Web Technologies," J. Emerging Technologies in Web Intelligence, vol. 1, no. 1, pp. 6-21, 2009.
5. J. Jovanovic et al., "Using Semantic Web Technologies for the Analysis of Learning Content," IEEE Internet Computing, vol. 11, no. 5, pp. 45-53, Sept./Oct. 2007.
6. J. Jovanovic, C. Knight, D. Ga_sevi_c, and G. Richards, "Ontologies for Effective Use of Context in e-Learning Settings," Educational Technology & Soc., vol. 10, no. 3, pp. 47-59, 2007.
7. L. Ali, M. Hatala, D. Ga_sevi_c, and J. Jovanovic, "A Qualitative Evaluation of Evolution of a Learning Analytics Tool," Computers & Education, vol. 58, no.1, pp. 470-489, 2012.
8. Z. Jeremic, J. Jovanovic, and D. Ga_sevi_c, "An Environment for Project-Based Collaborative Learning of Software Design Patterns," Int'l J. Eng. Education, vol. 27, no. 1, pp. 41-51, 2011.
9. P. Cimiano, J. Vo¨lker, and R. Studer, "Ontologies on Demand? – A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text Information," Wissenschaft und Praxis, vol. 57, nos. 6/7, pp. 315-320, 2006.
10. J. Park, W. Cho, and S. Rho, "Evaluating Ontology Extraction Tools Using a Comprehensive Evaluation Framework," Data & Knowledge Eng., vol. 69, pp. 1043-1061, 2010.
11. P. Cimiano and J. Voelker, "Text2onto-A Framework for Ontology Learning and Data-Driven Change Discovery," Proc. 10th Int'l Conf. Applications of Natural Language to Information Systems, pp. 227-238, 2005.
12. B. Fortuna, M. Grobelnik, and D. Mladenic, OntoGen: Semi- Automatic Ontology Editor. HCI Int'l, July 2007.38] M. Siadaty, J. Jovanovic, K. Pata, T. Holocher-Ertl, D. Ga_sevi_c, and N. Milikic, "A Semantic Web-Enabled Tool for Self-Regulated Learning in the Workplace," Proc. 11th IEEE Int'l Conf. Advanced Learning Technologies, pp. 66-70, 2011.
13. E. Simperl, M. Mochol, and T. Burger, "Achieving Maturity: the State of Practice in Ontology Engineering in 2009," Int'l J. Computer Science and Applications, vol. 7, no. 1, pp. 45-65, 2010.
14. E. Simperl, C. Tempich, and D. Vrande_ci_c, "A Methodology for Ontology Learning," Proc. Conf. Ontology Learning and Population: Bridging the Gap between Text and Knowledge, pp. 225-249, 2008.