GENERALIZING FRAMEWORK FOR ONTOLOGY LEARNING

Orobinska O.O.
Laboratoire ERIC de l'université Lyon-2 Lumière
5 avenue Pierre Mendès-France 69676 Bron cedex,
tel. : +33 (0)4 78 77 23 76 e-mail: Olena.Orobinska@univ-lyon2.fr

The techniques of ontology engineering have got new impact during the last decade with the occurrence of the different kinds of digital resources (such as MRDs, lexicons, thesaurus etc.). This has permitted to apply widely linguistic methods towards the tasks of ontology building and has generated new domain of ontology engineering such as ontology learning.

With in reference to M. Bergman [1] in 2011 among the fastest growing categories of “sweet tools”[1] have been all ones related to ontologies with Java as the dominant language of these tools.

But for all the successes in ontology engineering there is neither universal procedure generally accepted nor a fortiori any consistent suites of tools allowing to conceive in a traceable, explicit way an domain ontology from group of informational resources relevant to this domain [2].

We propose a general framework for ontology learning based on the related researches [3] and our point of view in ontology learning. According our approach the different components that are involved in ontology learning process is integrated into the framework to generate an ontology.

This framework consists of three backbone stages: preliminary information extraction, ontology discovery, and ontology organization.

Preliminary information extraction. A variety of data can be exploited in ontology learning, including raw-text documents (e.g., scientific articles, Web pages, book chapters, newspaper), (semi-)structural data (e.g., Web site structure lexicons), and usage data (e.g., the log of user navigation and search queries). Information extraction pre-processes and recognizes information in a variety of forms and converts them into the forms that can be used for ontology discovery. In particular, text documents are processed via content analysis by employing a variety of natural language processing techniques, ranging from tokenization, to part-of-speech tagging, phrase structure and/or grammatical function parsing, semantic and discourse analyses.

Ontology componrnts discovery. Supervised and unsupervised learning algorithms have been applied to discover the concepts and the relations from the extracted information. Approaches to relation learning vary in terms of the scope of co-occurrence, the metrics for the significance of co-occurrence, the criteria for selecting candidate concepts, and the thresholds for extracting potential relations. Some learning approaches require the assistance of domain-specific resources such as thesauri.

[1] semantic Web and -related tools by the AI3

Ontology implementation. Given the large number of possible ontological concepts and relations extracted from the learning process, an issue arises as to how to improve the usability of the discovered knowledge. Ontology implementation seeks to achieve the above goal via the following steps:

• clustering synonymous terms and their relations;

• discovering local centers of concepts. A concept cloud is a group of concepts that are closely related among themselves but marginally related to concepts outside the group. If connect any two concepts *i* and *j* that are directly related within a cloud via path p_{ij}, a network will be constructed. The length of *p* is the total number of concepts on *p* minus 1. Cloud center C_c can be found by looking for the concept(s) that has the shortest total path connecting to the rest of the concepts on the same cloud (the total is M), as shown in (1):

$$C_C = \min \sum_{i=1, i \neq c}^{M} p_{ci} \qquad (1);$$

• deriving inverse relations. We can derive inverse relationships from the ones that have been discovered;

• building higher-level ontology. The local centers discovered in the previous step can serve as the top-level ontology of the target domain. The process for identifying local centers in a concept cloud is repeated to find concepts for lower-level ontology by selecting concepts that have the next shortest length of path. This process continues to the desired level where concepts constitute the leaf nodes of an ontology hierarchy. It is possible to construct such a hierarchy for each relation separately.

It is desirable to automate all the components in the framework for ontology learning by developing techniques.

References

1. *Bergman M.* AI3's Inaugural State of Tooling for Semantic Technologies/ http://www.mkbergman.com/991/the-state-of-tooling-for-semantic-technologies.

2. *Charlet, J.,* at al. Apport des outils de TAL a la construction d'ontologies : propositions au sein de la plateforme DaFOE/ Charlet, J., at al dans TALN 2009, Senlis, 24–26 juin 2009

3. *De Nicola, A.,* & all. A software engineering approach to ontology building/ De Nicola, *A.,* et al In Information Systems 34 (2009) 258–275.