

Towards a system for constructing Arabic Ontology based on natural text

Dr. Nada Ghneim*

Eng. Waseem Safi**

Eng. Moayad Al Said Ali***

Abstract

This paper presents ArOntoLearn, a Framework for Arabic Ontology learning from textual resources. Supporting Arabic language and using domain knowledge in the learning process are the main features of our framework. Besides it represents the learned ontology in Probabilistic Ontology Model (POM), which can be translated into any knowledge representation formalism, and implements data-driven change discovery. Therefore it updates the POM according to the corpus changes only, and allows user to trace the evolution of the ontology with respect to the changes in the underlying corpus. Our framework analyses Arabic textual resources, and matches them to Arabic Lexico-syntactic patterns in order to learn new Concepts and Relations.

Supporting Arabic language is not that easy task, because current linguistic analysis tools are not efficient enough to process unvocalized Arabic corpora that rarely contain appropriate punctuation. So we tried to build a flexible and freely configured framework whereas any linguistic analysis tool can be replaced by more sophisticated one whenever it is available.

Keywords: Ontology, Ontology Learning, Knowledge Acquisition, Arabic Natural Language Processing.

For the Paper in Arabic see pages (185-192)

* Informatics Department, HIAST, Damascus, Syria, nada.ghneim@hiast.edu.sy

** Informatics Department, HIAST, Damascus, Syria, waseemm2005@hotmail.com

*** AI Department, IT Faculty, Damascus University, Damascus, Syria, moayad@live.com

References:

1. VerbNet, from the site:
<http://verbs.colorado.edu/~mpalmer/projects/verbnet/downloads.html>
2. John. son, C., Fillmore, C., Petruck, M. Baker, C., Ellsworth, M., Ruppenhofer, J., and Wood, E. 2002. FrameNet: Theory and Practice, from <http://www.icsi.Berkeley.edu/framenet>.
3. Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Jan Scheffczyk. "Frame Net II :Extended Theory and Practice", 2006.
4. WordNet. Retrieved June 2009, from <http://www.globalwordnet.org>
5. P. Cimiano and J. Volker. Text2onto – a framework for ontology learning and data-driven change discovery, Proc. NLDB 2005, Lecture Notes in Computer Science, vol. 3513, Springer, Alicante, 2005, pp. 227-238.
6. H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications, 2002, In Proceedings of the 40th Annual Meeting of the ACL.
7. R. Sonbol, N. Ghneim, M. Desouki, Arabic Morphological Analysis: a new approach, ICTTA 2008, Damascus, Syria.
8. G. Bisson, c. Nedellec, and L. canamero. Designing clustering methods for ontology building – The Mo'K workbench. In proceedings of the ECAI ontology Learning Workshop, pages 13-19, 2000.
9. Arabic Stanford parser, Retrieved October 2, 2008, from <http://www.nlp.stanford.edu/software/parser-faq.shtml>.
10. P. Buitelaar, D. Olejnik, and M. Sintek. OntoLT: A protégé plug-in for ontology extraction from text, 2003, In Proceeding of the international Semantic Web Conference, (ISWC).
11. P. Velardi, R. Navigli, A. Cuchiarelli, and F. Neri. Evaluation of ontolearn, a methodology for automatic population of domain Ontologies, 2005, In P. Buitelaar.
12. G. Salton. Developments in automatic text retrieval. 253:974–979, 1991.
13. R. M. Gray. Entropy and Information Theory. Springer, 1990.
14. J. Tsuji K. Frantzi, S. Ananiadou. The c-value/nc-value method of automatic recognition for multi -word terms. In Proceedings of the ECDL, pages 585–604, 1998.
15. G. Miller. WordNet Hyponym Classification: A lexical database for English. Communications of the ACM, 38(11):39–41, 1995.