

Methodology of Knowledge Representation from Natural Language

D. N. Koggalahewa, S. P. B. Athauda, S. U. Pilapitiya, K. G. D. Tharanga
and O. A. R. K. Fonseka

Department of Information Technology, Sri Lanka Institute of Information Technology

Information available in different formats cannot be understood by a computer or a machine due to lack of a proper knowledge representation mechanism. It always requires more human effort in feeding the knowledge to the computers or the knowledge base. XML covers the basic level of knowledge representation, but is incapable of utilizing the concepts and semantics in a proper way. Onto_X is an effort made to automate the process of ontology construction from an annotated xml file or database.

The annotation process is done by any natural language processing tool (apart from the system). The system requires an xml file as the input and converts it into ontology in owl format. The system is capable of generating the semantics over annotated content with owl components. Xml entities will be automatically mapped into the owl components such as classes, sub classes, instances and relationships. The conversion mechanism is totally automated inside the Onto_X since it assures all the co-relationships over the annotated content. The conversion process identifies the xml properties and assigns semantics with the integration of word-net 2.1 and owl properties over the parsed content. The system uses the protégé libraries for the conversion process. The most special feature in the conversion process is that it uses its own inference, without just mapping xml properties to owl. The system is capable of visualizing the mapped owl ontology and it allows the user to refine the content of the constructed ontology. The final outcome of the system is ontology in owl format, which is mapped from the xml file or a database. The research ensures a better knowledge representation mechanism and it will assure the creation of domain knowledge from the xml file. The expandability is high since it takes information from the base level.

Key words: Information, Knowledge, Onto_X, Natural language