Abstract

In this paper, we are presenting a model for implementation of ontology through invoking a semi-automatic process from web data tables. The singularity of this method is the implementation of cosine similarity for filtering the documents. We present the framework for the development of generic method concerning data integration. The data warehouse is composed of several data tables extracted from the web and it has been supplemented by the existing local data sources. Documents pertaining to different domains can be supplemented as input to the data sources which are stored in data warehouse and then are converted into XML/RDF data. We have used cosine similarities to measure the similarity of two documents which are likely to be same in terms of their subjects. It is a semi-automatic method to extract and integration of web documents by using ontology.

References

- Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu, and Lydie Soler; Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource; IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4,
APRIL 2013.
- JCGM 200:2008 International vocabulary of metrology — Basic and general concepts and associated terms (VIM)
- Mark van Assem, Hajo Rijgersberg, Mari Wigham and Jan Top; Converting and Annotating Quantitative Data Tables in 2009.
- S´ebastien Destercke and Patrice Buche and Brigitte Charnomordic; Data reliability assessment in a data warehouse opened on the Web; IEEE Transactons on 2011.
- P. Cimiano a; P. Buitelaar b, J. McCrae a, M. Sintek; LexInfo: A Declarative Model for the Lexicon-Ontology Interface in 2009.

Index Terms
Computer Science
Web Technology

Keywords
Ontology XML/RDF SPARQL Cosine similarity.