

World WordNet Database Structure: An Efficient Schema for Storing Information of WordNets of the World

Hanumant Redkar, Sudha Bhingardive, Diptesh Kanojia, Pushpak Bhattacharyya

Center for Indian Language Technology, Indian Institute of Technology Bombay, Mumbai, India

hanumantredkar@iitb.ac.in, sudha@cse.iitb.ac.in, diptesh@cse.iitb.ac.in, pb@cse.iitb.ac.in

Abstract

WordNet is an online lexical resource which expresses unique concepts in a language. English WordNet is the first WordNet which was developed at Princeton University. Over a period of time, many language WordNets were developed by various organizations all over the world. It has always been a challenge to store the WordNet data. Some WordNets are stored using file system and some WordNets are stored using different database models. In this paper, we present the World WordNet Database Structure which can be used to efficiently store the WordNet information of all languages of the World. This design can be adapted by most language WordNets to store information such as synset data, semantic and lexical relations, ontology details, language specific features, linguistic information, etc. An attempt is made to develop Application Programming Interfaces to manipulate the data from these databases. This database structure can help in various Natural Language Processing applications like Multilingual Information Retrieval, Word Sense Disambiguation, Machine Translation, etc.

Introduction

The Princeton WordNet or the English WordNet (Miller, 1990) was the first WordNet which was developed at Princeton University. This WordNet inspired many countries and organizations in the world to develop WordNet in their own languages. Over a period of time, many individual language as well as multi-lingual WordNets evolved. Some of the individual language WordNets are GermaNet, Japanese WordNet, etc. and some of the multi-lingual WordNets are EuroWordNet (Vossen et al., 1997), IndoWordNet (Bhattacharyya, 2010), etc. All these WordNets use various methods to store their WordNet data. For example, Princeton WordNet uses text files to store WordNet data, whereas GermaNet uses relational database structure (Henrich et al., 2010). Also, some WordNets such as MultiWordNet uses multiple databases (Pianta et al., 2002) while WOLF (Sagot et al., 2008) uses XML file structure to store their data. All these storage methods are good in their own respects, but they

also have some limitations – File system uses flat files which may not be efficient to store, manipulate and retrieve the relevant WordNet data. The database system uses relational databases to store data, but these databases do not capture all the WordNet and related data. XML also uses file system to store information.

We propose a World WordNet Database Structure (WWDS) which can possibly be made standard to store global WordNet data. It can be used to efficiently store and manipulate multi-lingual WordNet data. It uses multiple databases wherein the language independent information is stored in a single master database and language dependent information is stored in multiple language specific databases. The detailed description of WWDS is given in the following section. Further, the design and the database schema of WWDS are depicted. In the subsequent sections, the advantages and future work of WWDS are mentioned.

World WordNet Database Structure

World WordNet Database Structure or WWDS is designed based on IndoWordNet database structure (Prabhu et al., 2012). It consists of a single master database and multiple language specific databases. The master database is named as *wordnet_master*, which contains language independent data such as semantic relations, ontology details, etc., which are common across all the languages. The language specific database is named as *wordnet_<language>*, which contains language dependent data such as synsets, words, lexical relations, etc., which are specific to a particular language in consideration. For example, *wordnet_german*, *wordnet_hindi*, *wordnet_english* are language specific WordNet databases for German, Hindi, English languages respectively. In this way, we can make use of WWDS to store multi-lingual WordNets. In *wordnet_master* database,

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

