

World WordNet Database Structure (WWDS)

An Efficient Schema for Storing Information of the WordNets of the World

By
Hanumant Redkar
Sudha Bhingardive
Diptesh Kanojia
and
Pushpak Bhattacharyya

Indian Institute of Technology Bombay, Mumbai, India

*29th Association for the Advancement of Artificial Intelligence (AAAI) Conference
25th to 30th January, 2015, Austin, Texas, United States*

Outline

- Introduction
- WordNet Storage Mechanisms
- Motivation
- World WordNet Database Structure
- Demonstration
- Summary
- Future Work
- References

Introduction

- WordNet – an Online Lexical Resource expresses unique concept in a language.
- Various WordNets are developed for different languages in the world.
 - **Individual WordNets:** Princeton WordNet, GermaNet, Japanese WordNet, Hindi WordNet, etc.
 - **Multi-lingual WordNets:** IndoWordNet, EuroWordNet, MultiWordNet, etc.

WordNet Storage Mechanisms

- WordNets use various data storage methods:
 - Princeton WordNet uses text files to store WordNet data (Miller, 1990).
 - GermaNet uses relational database structure (Henrich et al., 2010).
 - MultiWordNet uses multiple databases (Pianta et al., 2002).
 - WOLF uses XML file structure (Sagot et al., 2008).

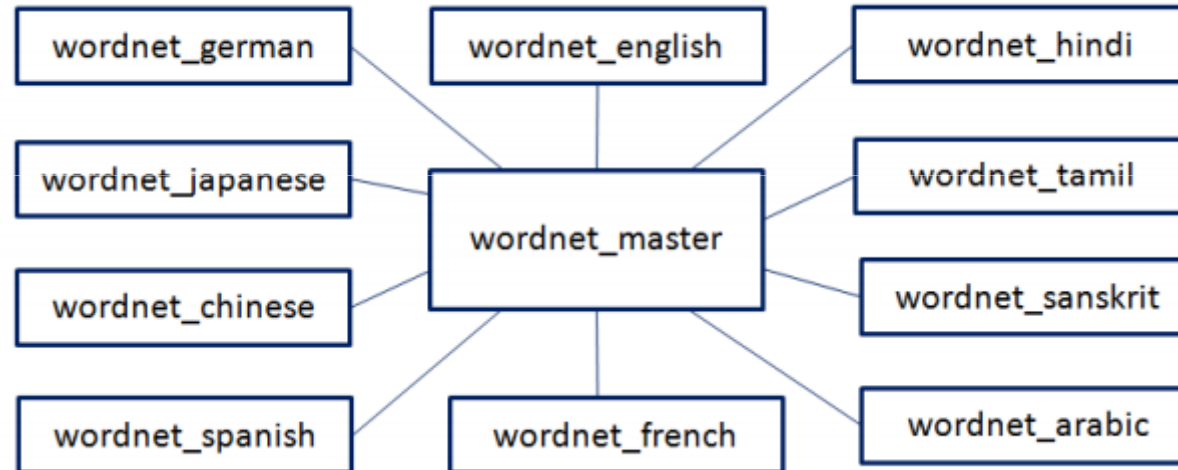
Motivation

- Challenge to store the WordNet data
- Storage Mechanisms used so far:
 - Text Files.
 - Relational Databases.
 - XML Files.
 - Others.

Motivation (Contd...)

- Problems with Text File system
 - Inefficient in case of selected field look-up.
 - Leads to increase in processing time.
 - Difficulty in data manipulation and data retrieval.
 - Need to parse different formats for different languages.
- Problems with XML storage
 - XML parsers need to be used for look up.
 - XML parsers need to be written depending on distinct attributes.
 - Adding new attributes in XML structure needs a lot of modification.

World WordNet Database Structure (WWDS)



Block diagram of WWDS

WWDS (Contd...)

- wordnet_master
 - Language independent database where all the common data and relations are stored.
- wordnet_<language>
 - Language dependent database where the language specific data and relations are stored.
 - For instance, for English and Hindi languages the language database would be wordnet_english and wordnet_hindi respectively.

WWDS (Contd...)

- The master database and all the language databases are connected through a single inter-lingual synset id.
- Using this inter-lingual synset id we can retrieve cross-lingual synset information of all the available languages.

WWDS (Contd...)

- Demonstration (WWDS Schema)

WWDS (Contd...)

wordnet_hindi (15)	
<input type="text"/>	X
<input type="checkbox"/> wn_fileholder	
<input type="checkbox"/> wn_lexical_relations	
<input type="checkbox"/> wn_ontology_nodes	
<input type="checkbox"/> wn_rel_adverb_derived_from_verb	
<input type="checkbox"/> wn_rel_antonymy	
<input type="checkbox"/> wn_rel_compounding	
<input type="checkbox"/> wn_rel_conjunction	
<input type="checkbox"/> wn_rel_gradation	
<input type="checkbox"/> wn_rel_verb_derived_from_noun	
<input type="checkbox"/> wn_source	<input type="checkbox"/> wn_source
<input type="checkbox"/> wn_synset	<input type="checkbox"/> wn_synset
<input type="checkbox"/> wn_synset_example	<input type="checkbox"/> wn_synset_example
<input type="checkbox"/> wn_synset_file	<input type="checkbox"/> wn_synset_file
<input type="checkbox"/> wn_synset_words	<input type="checkbox"/> wn_synset_words
<input type="checkbox"/> wn_word	<input type="checkbox"/> wn_word
15 table(s)	
<input type="checkbox"/> Check All / Uncheck All	

Snapshot of **wordnet_hindi** database

WWDS (Contd...)

wordnet_master (22)	
<input type="checkbox"/>	wn_master_domain
<input type="checkbox"/>	wn_master_interlingual_synset
<input type="checkbox"/>	wn_master_language
<input type="checkbox"/>	wn_master_ontology_nodes
<input type="checkbox"/>	wn_master_ontology_synset_map
<input type="checkbox"/>	wn_master_ontology_tree
<input type="checkbox"/>	wn_master_property
<input type="checkbox"/>	wn_master_relation_types
<input type="checkbox"/>	wn_master_rel_adjective_modifies_noun
<input type="checkbox"/>	wn_master_rel_adverb_modifies_verb
<input type="checkbox"/>	wn_master_rel_also_see
<input type="checkbox"/>	wn_master_rel_hyponymy_hyponym
<input type="checkbox"/>	wn_master_rel_meronymy_holonymy
<input type="checkbox"/>	wn_master_rel_near_synsets
<input type="checkbox"/>	wn_master_rel_noun_adjective_attril
<input type="checkbox"/>	wn_master_rel_noun_verb_link
<input type="checkbox"/>	wn_master_rel_similar
<input type="checkbox"/>	wn_master_rel_troponymy
<input type="checkbox"/>	wn_master_semantic_relations

Snapshot of **wordnet_master** database

WWDS (Contd...)

	WWDS	Princeton WordNet	WOLF	Euro WordNet	GermaNet	Multi-WordNet
WordNet Type	Multi-lingual	Mono-lingual	Mono-lingual	Multi-lingual	Mono-lingual	Multi-lingual
Languages Involved	All languages	English	French	European languages	German	Italian and English
Storage Structure	Relational database	File system	XML system	File system	Relational database	Relational database
Synset Identified By	Isynset id	File offset	Synset id	File offset	Synset id	Synset id
Data Management	Database query processing	Text file processing	Individual parsers required	Text file processing	Database query processing	Database query processing
Applications	Cross lingual NLP systems	-	-	Cross lingual NLP systems	-	Cross lingual NLP systems

Comparative Study of Existing Storage Mechanism with WWDS

WWDS (Contd...)

- Advantages of WWDS
 - Centralized database system.
 - Easier and faster to process, manipulate and retrieve the WordNet data.
 - Can be used in applications like Cross Lingual Information Retrieval, Word Sense Disambiguation, Machine Translation, etc.

Summary

- Challenge to store WordNet data is addressed by WWDS.
- WWDS is an efficient and normalized database schema for storing, manipulating and retrieving multi-lingual WordNet information.
- WWDS is a centralized storage structure.
- One can easily adapt this structure to store their WordNet data.

Future Work

- Study the semantic and lexical relations of all the WordNets and revise the WWDS.
- Plan to study features of individual languages and add to WWDS.
- To provide WWDS APIs to the WordNet users for easy manipulation and retrieval of WWDS data.

References

- Bhattacharyya, P. 2010. IndoWordNet. *Proc. of LREC-10*, Malta.
- Henrich, V., & Hinrichs, E. W. 2010. GernEdiT-The GermaNet Editing Tool. *In ACL (System Demo)*, Malta. (pp. 19-24).
- Miller, George A., R., Fellbaum, C., Gross, D., & Miller, K. J. 1990. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, OUP. (pp. 3.4: 235-244).
- Pianta, E., Bentivogli, L., & Girardi, C. 2002. Developing an aligned multilingual database. *In Proc. 1st GWC*, Mysore, India.

References (Contd...)

- Prabhu, V., Desai, S., Redkar, H., Prabhugaonkar, N., Nagvenkar, A., & Karmali, R. 2012. An Efficient Database Design for IndoWordNet Development Using Hybrid Approach. *COLING 2012*, Mumbai, India. (pp. 229).
- Prabhugaonkar, N., Nagvenkar, A., & Karmali, Ramdas N. 2012. IndoWordNet Application Programming Interfaces. *COLING 2012*, Mumbai, India. (pp. 237 - 244).
- Sagot, B., & Fišer, D. 2008. Building a free French wordnet from multilingual resources. *In OntoLex 2008*, Morocco.
- Vossen, P. 1997. EuroWordNet: A multilingual database for information retrieval. *DELOS*, Zurich. (pp. 5-7).

Acknowledgments

- Center for Indian Language Technology (CFILT), Indian Institute of Technology Bombay (IITB), Mumbai, India.
- Industrial Research and Consultancy Center (IRCC), IITB, Mumbai, India.
- Microsoft Research Outreach India, Bangalore, India.
- Association for the Advancement of Artificial Intelligence (AAAI), California, United States.

Thank You!!!