

Introduction to Malayalam WordNet

Project Coordinator
Dr. Sumam Mary Idicula

Team Members
Drishya Gopinath
Varghese K. Aniyam
Department of Computer Science
CUSAT

Table Of Contents

1. WordNet
2. History of WordNet
3. Applications of WordNet
4. Semantic Relationships in Malayalam WordNet
5. Features of Malayalam WordNet
6. Useful Links
7. Acknowledgement

- WordNet is a semantic dictionary that was designed as a network following the idea that representing words and concepts as an interrelated system
- WordNet groups words into sets of synonyms and provides short definitions and usage examples, also records a number of relations among these synonym sets or their members.
- WordNet can thus be seen as a combination of dictionary and thesaurus.
- Malayalam WordNet is a lexical database for the Malayalam language.

History of WordNet

WordNet	Reference	Details
Princeton WordNet (English) 1985	<p>Introduction to WordNet: An On-line Lexical Database George A. Miller, Richard Beckwith Christiane Fellbaum, Derek Gross, and Katherine Miller International Journal of Lexicography , Volume 3, Issue 4(1990)Pp. 235-244p http://WordNet.princeton.edu/</p>	<p>First wordnet to be developed Headed by G.A.Miller Nouns, verbs, adjectives and adverbs are organized into sets of synonyms, each representing a lexicalized concept Contains about 2 lakhs of words Uses text files to store WordNet data</p>
EuroWordNet (Dutch, Italian, Spanish, Czech, French, German, Estonian) 1996-1999	<p>EuroWordNet: a multilingual database for information retrieval Piek Vossen, University of Amsterdam Delos Workshop on Cross-language Information Retrieval, Chicago, 1997</p>	<p>First multilingual wordnet to be developed WordNet for European languages Each language has unique language-internal system of lexicalizations WordNets are linked to an Inter-Lingual-Index, based on the Princeton WordNet Uses text files to store WordNet data</p>

Table 1 : Princeton and Euro WordNet

History of WordNet

WordNet	Reference	Details
IndoWordNet	<p>IndoWordNet Dr. Pushpak Bhattacharyya, IIT Bombay Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)</p> <p>Building a WordNet for Dravidian Languages S. Rajendran (Tamil University) G. Shivapratap, V. Dhanalakshmi, KP. Soman (Amrita Vishwa Vidyapeeth) Fifth International Conference of the Global WordNet Association (GWC-2010)</p>	<p>Linked WordNet of major Indian languages Composite of North East WordNet, Dravidian WordNet and Indradhanush project Hindi WordNet is the first Indian language WordNet Project was guided by Dr. Pushpak Bhattacharyya, IIT Bombay.</p> <p>It uses relational database for storing the data The project was funded by Ministry of Communication and Information technology, and also partially from Ministry of Human Resource Development, India.</p>

Table 2 : IndoWordNet

Project Name	Languages	Developed By
Dravidian WordNets	Telugu, Tamil, Kannada, Malayalam	IIT Bombay (Leader), Tamil University, Amrita Viswa Vidyapeetham, University of Mysore, Dravidian University
Indradhanush WordNets	Bengali, Gujarati, Kashmiri, Konkani, Odia, Punjabi, Urdu	Goa University(Leader), IIT Bombay, Indian Statistical Institute, University of Kashmir, University of Hyderabad, Punjabi University, Thapar University, Dharmsinh Desai University, Jawaharlal Nehru University
North-East WordNets	Assamese, Bodo, Nepali, Manipuri	Gauhati University, Manipur University, Assam University

Table 3 : IndoWordNet Structure

Applications of WordNet

- Information retrieval and extraction
- Document categorisation
- Audio and video retrieval
- Language teaching and translation applications
- Machine Translation
- Automatic text summarization
- Word Sense Disambiguation
- Semantic Parsing
- Opinion Mining

Synset, Gloss, Position in Ontology, Example

- **Synset**
 - It is a set of synonymous words
- **Gloss**
 - It explains the concept denoted by the synset
- **Position in Ontology**
 - Each synset is mapped into some places in ontology such as noun,verb,adjective and adverb
- **Example**
 - An example sentence containing a word in the synset is also provided

Hypernymy and Hyponymy

- Hyponymy/Hypernymy relation is a subordination/superordination, subset/superset relationship
- A concept represented by the synset ($x_1, x_2, x_3 \dots$) is said to be a hyponym of the concept represented by the synset ($y_1, y_2, y_3 \dots$) if an x is a kind of y
- Hypernymy is the reverse of hyponymy

Meronymy, Holonymy and Antonymy

- If the concepts A and B are related in such a manner that A is one of the constituent of B, then A is the meronym of B and B is the holonym of A
- Meronymy consists of Part of, Member of and Substance of relationships
- Antonymy is a relation that hold with two words that express opposite meanings

Features of Malayalam WordNet

- Searching and browsing of words
- Creating new Synset
- Edit existing Synset
- Adding new Semantic Relationships
- Malayalam WordNet also supports REST Services

Searching and browsing

- It will allow the user to search for a specific word
- All the input fields are transliteration enabled
- Example

Creating new Synset

The procedures for adding new word is as follows

- Select the Create Synset from User Option tab
- Enter the word(s) and press proceed
- Now a list of available synsets (if any) related to the given synset will be shown
- If we need to add to the existing one select Edit this and provide the new synonym
- else select create new synset option and enter details

Example

The procedures for editing synset is as follows

- Select the Edit Synset from User Option tab
- Enter the word(s) and press proceed
- Now a list of available synsets (if any) related to the given synset will be shown
- Select the synset we want to edit and edit the details

Example

Adding new Semantic Relationships

The procedures for adding new relationship as follows

- Select the Add Relationship from User Option tab
- Enter the word(s) and press proceed
- Now a list of available synsets (if any) related to the given synset will be shown
- Select the synset to which relationship to be added and click add to this
- Then provide the necessary details
- The relationship will be updated on both base and related synset

Example



WordNet REST Services - Structure of Output Value

Variable	Details
Search_Result	It returns the available synsets. The below fields will be included if it contains data. Else it will be empty
sid	Gives the id for the Synset
synset	Gives the synset. It list all synonyms[array form]
meaning	Gives the meaning of the Synset
example	Gives an example statement for the synset
pos	Gives the POS of the word
hypernym	Gives the list of allrelated Hypernyms. Hypernyms will be represented as a subobject containing {sid,synset,meaning example andpos}
hyponym	Gives the list of allrelated Hyponyms. Hyponyms will be represented as a subobject containing {sid,synset,meaning example andpos}
meronym	Gives the list of allrelated Meronyms. Meronyms will be represented as a subobject containing {sid,synset,meaning example andpos}
holonym	Gives the list of allrelated Holonyms. Holonyms will be represented as a subobject containing {sid,synset,meaning example andpos}
antonym	Gives the list of allrelated Antonyms. Antonyms will be represented as a subobject containing {sid,synset,meaning example andpos}

Table 4 : Structure of Output Value

WordNet REST Services - URL Pattern

- **Access by Word**

- Input Parameter : queryword - The word to be searched

- Url Format for getting XML :

- ```
http://malayalamwordnet.cusat.ac.in/searchByWordXml.do?
queryword=<word>
```

- Url Format for getting Json :

- ```
http://malayalamwordnet.cusat.ac.in/searchByWordJson.do?  
queryword=<word>
```

- XML

- JSON

WordNet REST Services - URL Pattern

- **Access by Synset Id**

- Input Parameter : sid - The synset id to be searched

- Url Format for getting XML :

- [http://malayalamwordnet.cusat.ac.in/searchByWordXml.do?
sid=<synset id>](http://malayalamwordnet.cusat.ac.in/searchByWordXml.do?sid=<synset id>)

- Url Format for getting Json :

- [http://malayalamwordnet.cusat.ac.in/searchByWordJson.do?
sid=<synset id>](http://malayalamwordnet.cusat.ac.in/searchByWordJson.do?sid=<synset id>)

- XML

- JSON

WordNet REST Services - URL Pattern

- **Access by Word and Position in Ontology**

- Input Parameter : queryword - The word to be searched
& sid - The synset id to be searched

- Url Format for getting XML :

[http://malayalamwordnet.cusat.ac.in/searchByWordXml.do?
queryword=<word>& pos=<pos>](http://malayalamwordnet.cusat.ac.in/searchByWordXml.do?queryword=<word>&pos=<pos>)

- Url Format for getting Json :

[http://malayalamwordnet.cusat.ac.in/searchByWordJson.do?
queryword=<word>& pos=<pos>](http://malayalamwordnet.cusat.ac.in/searchByWordJson.do?queryword=<word>&pos=<pos>)

- XML
- JSON

Useful Links

- WordNet Site :
<http://malayalamwordnet.cusat.ac.in>
- Wikipedia Page :
https://en.wikipedia.org/wiki/Malayalam_WordNet
- Facebook Page :
<https://www.facebook.com/MalayalamWordnet>

Acknowledgement

- CFILT - Center for Indian Language Technology
- TDIL - Technology Development for Indian Languages
- C-DAC - Centre for Development of Advanced Computing
- Dr. Pushpak Bhattacharyya (Director, IIT Patna)
- Dr. K.P. Soman (Head and Professor, Department of Computational Engineering and Networking, Amrita Vishwa Vidyapeetham)
- Dr. S. Rajendran (Professor of Linguistics, Department of Computational Engineering and Networking, Amrita Vishwa Vidyapeetham)

Thank You