Language Resource Management System for Asian WordNet Collaboration and Its Web Service Application

Virach Sornlertlamvanich (1,2), Thatsanee Charoenporn (1,2), Hitoshi Isahara (3)
(1) Thai Computational Linguistics Laboratory, NICT Asia Research Center, Thailand
(2) National Electronics and Computer Technology Center, Pathumthani, Thailand
(3) National Institute of Information and Communications Technology, Japan
E-mail: virach@tcllab.org, thatsanee@tcllab.org, isahara@nict.go.jp

This paper presents the language resource management system for the development and dissemination of Asian WordNet (AWN) and its web service application. We develop the platform to establish a network for the cross language WordNet development. Each node of the network is designed for maintaining the WordNet for a language. Via the table that maps between each language WordNet and the Princeton WordNet (PWN), the Asian WordNet is realized to visualize the cross language WordNet between the Asian languages. We propose a language resource management system, called WordNet Management System (WNMS), as a distributed management system that allows the server to perform the cross language WordNet retrieval, including the fundamental web service applications for editing, visualizing and language processing. The WNMS is implemented on a web service protocol therefore each node can be independently maintained, and the service of each language WordNet can be called directly through the web service API. In case of cross language implementation, the synset ID (or synset offset) defined by PWN is used to determined the linkage between the languages.

Motivation

- Need of a computational ontology
  - Implementation
- Quick start approach
  - Reusability
  - Less language resource
- Online collaborative environment
  - Social networking
  - Multilingual development
- Cross language web service
  - Sharing
  - Interoperability
  - Evaluation
  - Dissemination

Figure 1 Asian WordNet Classified by Category

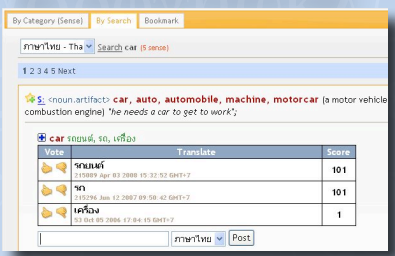

Figure 2 Asian WordNet Editing and Voting

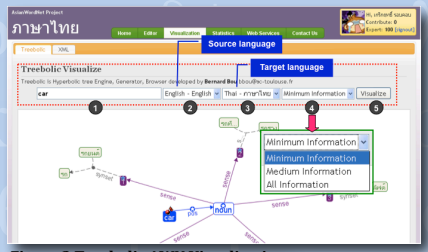

Figure 3 Treebolic AWN Visualization

AWN lookup (Thai to Japanese):

1. When receiving the word entry, Web Service API will look for the attaching senses of Thai word in Thai WordNet database.
2. Following information of Thai word entry will be retrieved from the database.
   a. Synset of the Thai word
   b. Synset offsets of English WordNet
   c. POS with category of base type
   d. Synset of English word
3. The synset offsets of English WordNet will be submitted to Japanese server to look for the information of Japanese WordNet.
4. The corresponding synsets of Japanese word will be sent back to Thai server.
5. The visualization tool presents the WordNet structure of the retrieved Thai and Japanese WordNet.

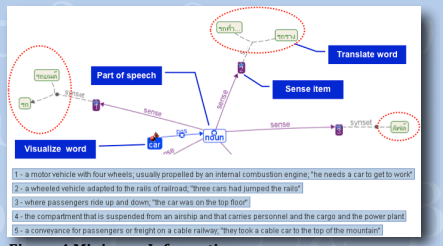

Figure 4 Minimum Information

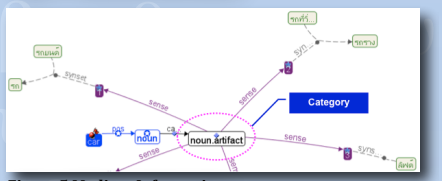

Figure 5 Medium Information

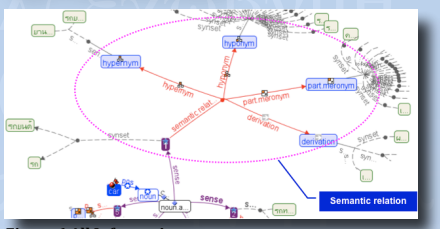

Figure 6 All Information

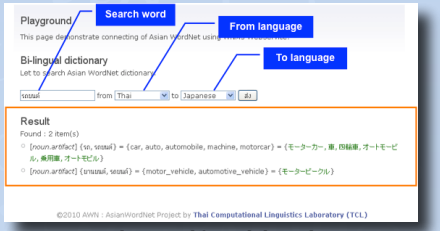

Figure 9 AWN Playground for Multilingual Dictionary

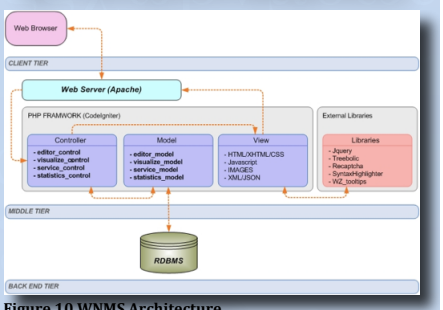

Figure 10 WNMS Architecture

Figure 7 Cross Language AWN through Web Service

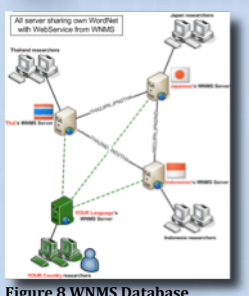

Figure 8 WNMS Database Sharing through Web Service Protocol

AWN Partners and Progress (as of May 3, 2010)

Language/Organization	Sense	Unique word
Burmese MCF (Myanmar Computer Federation)	19949	11006
Indonesian BPPT (Badan Pengkajian Dan Penerapan Teknologi)	26175	24398
Japanese NICT (National Institute of Information and Communications Technology)	58447	64678
Korean TCL/NICT (collaborator wanted)	42274	26009
Lao NAST (National Authority of Science and Technology)	38890	44032
Mongolian NUM (National University of Mongolia)	1624	1574
Nepali MPP (Madan Puraskar Pustakalaya)	41	42
Sinhala UCSC (University of Colombo School of Computing)	268	119
Sudanese TCL/NICT (collaborator wanted)	69	52
Thai TCL (Thai Computational Linguistics Laboratory), NICT	71139	69998
Vietnamese VAST (Vietnamese Academy of Science and Technology)	12234	4916

http://www.asianwordnet.org/