

Asian WordNet Framework: Its Web Service and Collaborative Platform

Virach Sornlertlamvanich

virach@gmail.com

SIIT, Thammasat University, Thailand



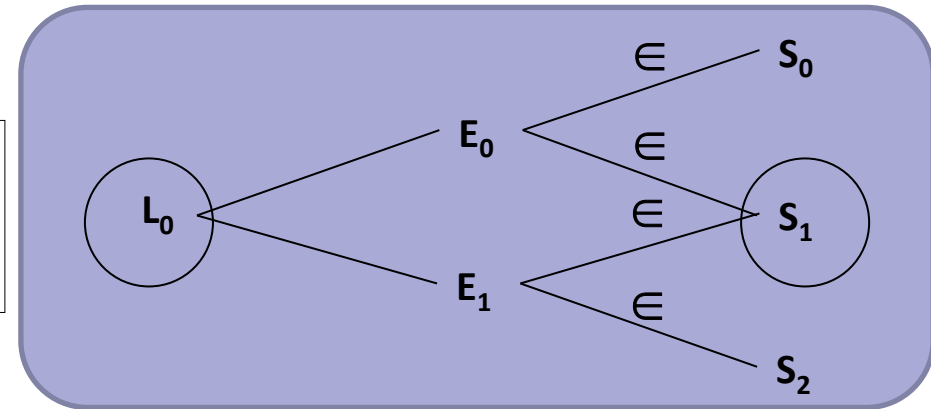


Outline

- Asian WordNet Development
 - Translation approach
 - Use of the existing bilingual dictionaries
 - Synset assignment
 - KUI for collaborative editing
- WNMS (WordNet Management System)
 - Distributed WordNet service
 - Service for cross language WordNet retrieval

Synset Assignment Algorithm (CS=4)

- Accept the Synset that includes more than one English Equivalent with confidence score 4.



Example:

L0: เป้าหมาย

E0: aim

E1: target

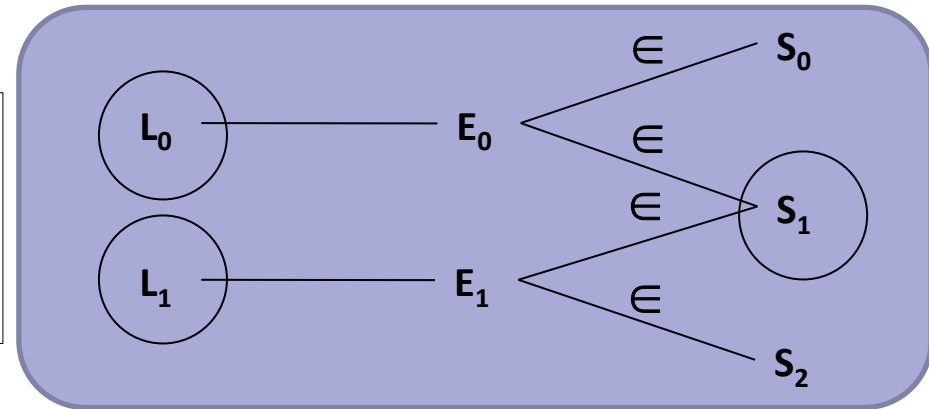
S0: purpose, intent, intention, aim, design

S1: aim, object, objective, target

S2: aim

Synset Assignment Algorithm (CS=3)

- Accept the Synset that includes more than one English Equivalent from the synonym of the target language with confidence score 3.



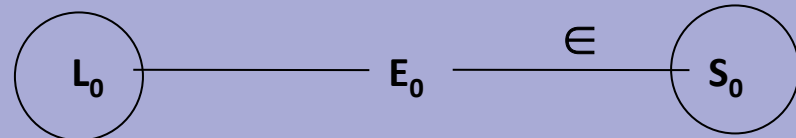
Example:

L0: จ้อง	} Synonym
L1: เพ่งมอง	
E0: stare	
E1: gaze	
S0: stare	
S1: gaze, stare	

A blue arrow points from the S1 row back to the L0 and L1 rows, indicating the selection of S1 based on the algorithm's criteria.

Synset Assignment Algorithm (CS=2)

- Accept the only Synset that includes the English Equivalent with confidence score 2.



Example:

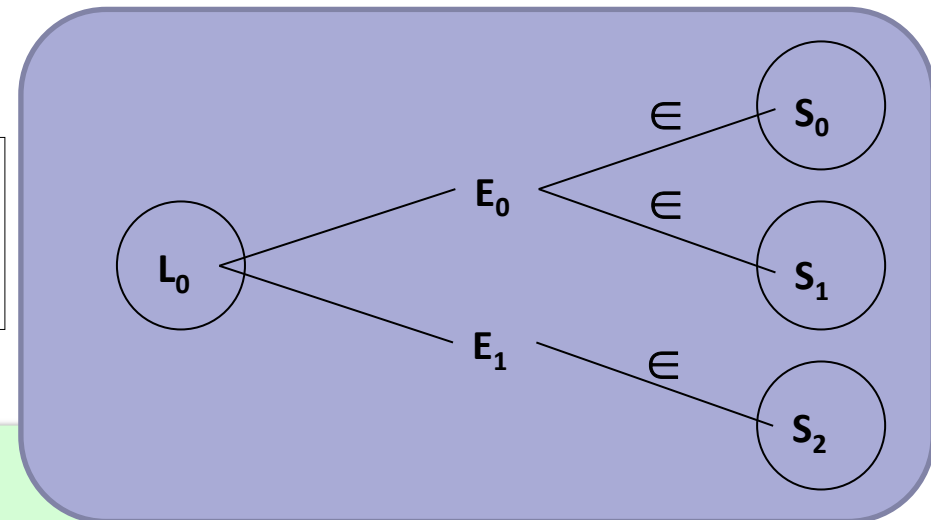
L0: สูติแพทย์

E0: obstetrician

S0: obstetrician, accoucheur

Synset Assignment Algorithm (CS=1)

- Accept more than one Synset that includes each of the English Equivalent with confidence score 1.



Example:

L0: ช่อง

E0: hole

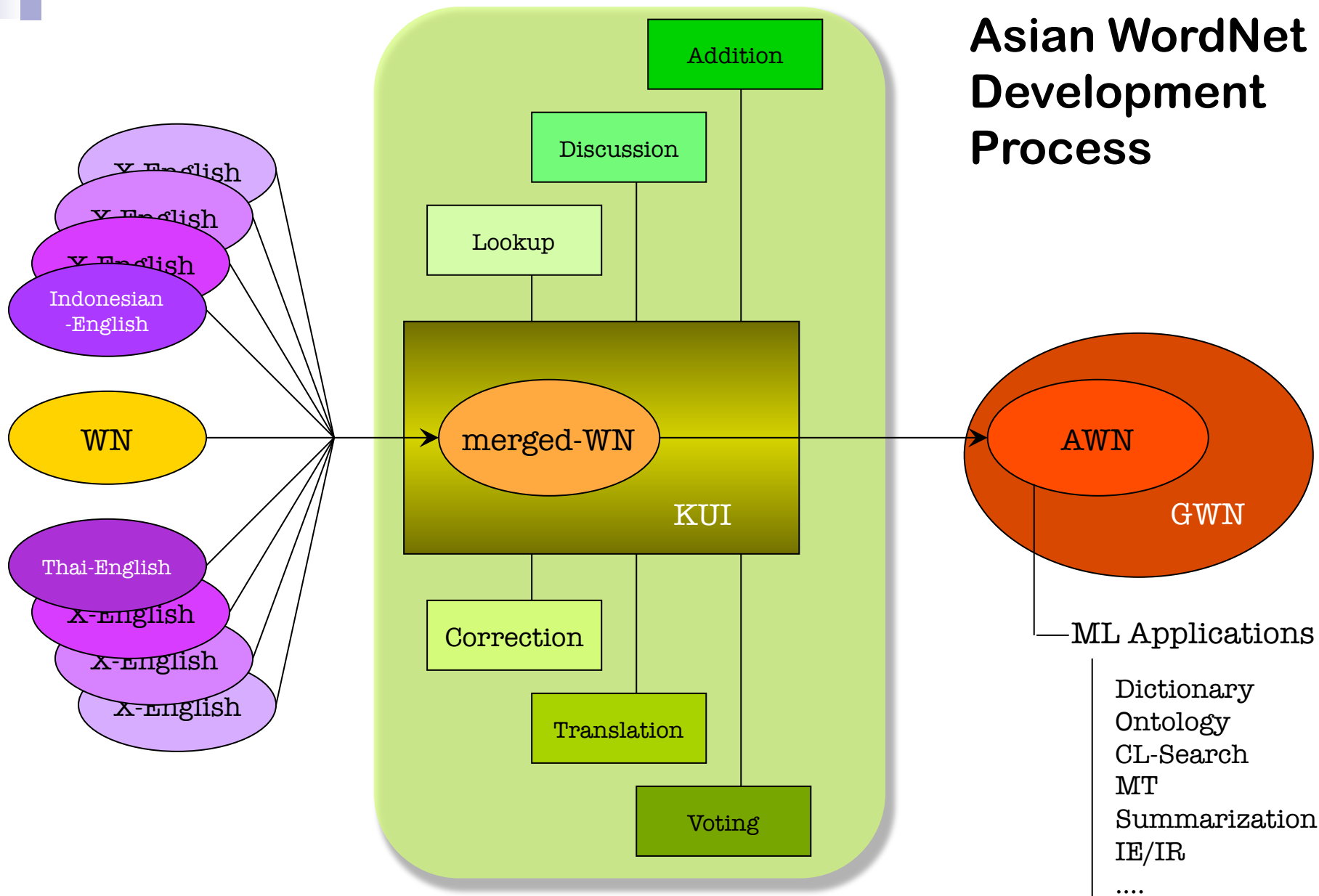
E1: canal

S0: hole, hollow

S1: hole, trap, cakehole, maw, yap, gap

S2: canal, duct, epithelial duct, channel

Asian WordNet Development Process





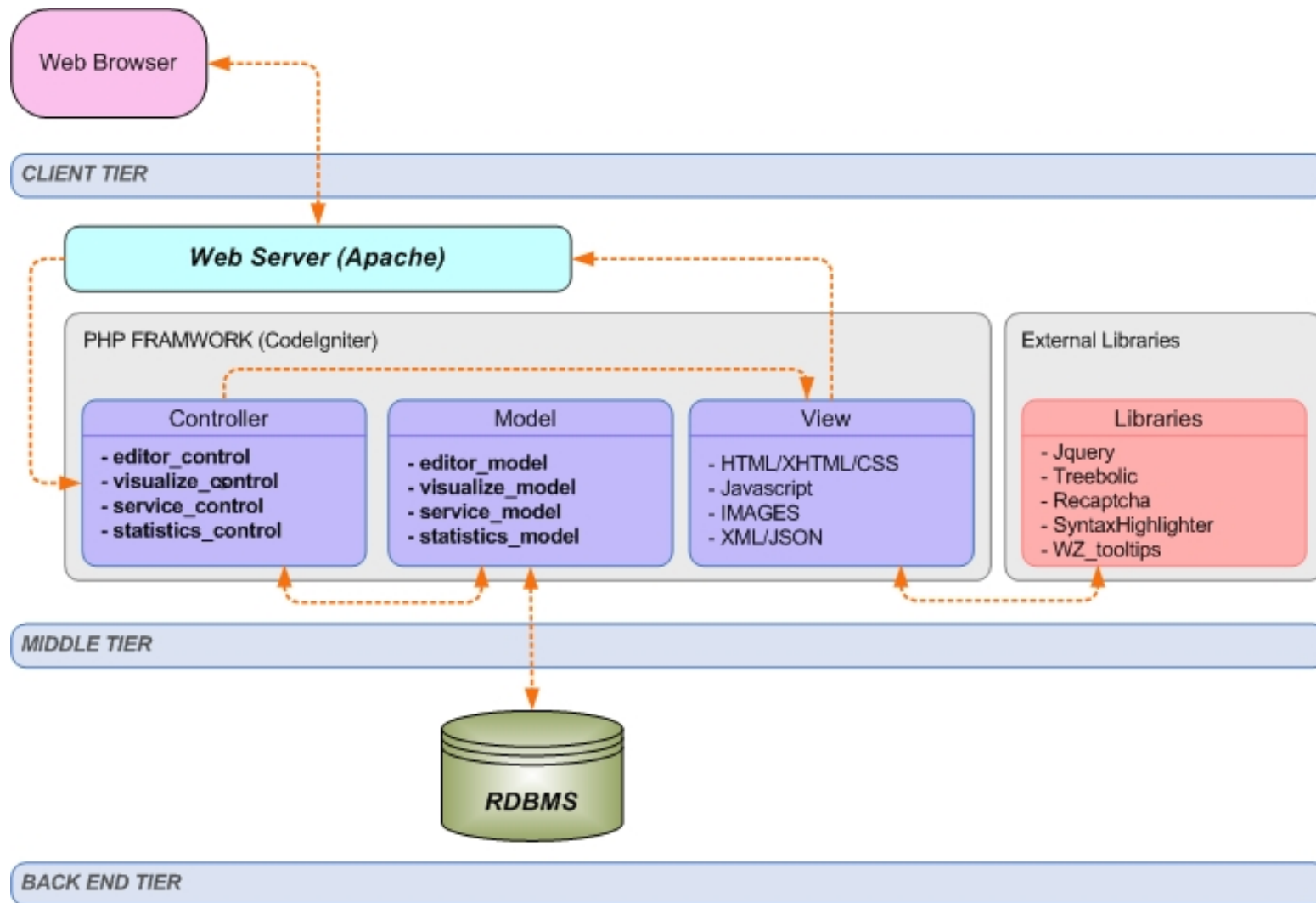
WNMS for AWN

- **WNMS (WordNet Management System)**
 - Sense based translation rather word based translation
 - Show the relation between senses
 - System is fully distributed connected through a standard Open API
 - Collaborative editing tools based on KUI concept

WMS for BalkaNet, GWC2004

WordNet Exploitation through a Distributed Network of Servers, I. D. Koutsoubos, and et. al.

WNMS (WordNet Management System)



Participation (Translate)

- Input a word to search
- Input a translated word, and select degree of confidence
- Input comment or memo if have

Delete

The screenshot shows the Lexitron web interface for the word "travel". At the top, there are navigation tabs: "By Category (Sense)", "By Search", and "Bookmark". Below this, a search bar contains the word "travel" and a "Search Word" button. A "Go to:" section shows page navigation (1-9). The main content area displays the word "travel" with its Thai translation "เดินทาง" and a definition. Below the definition, there are two user comments. At the bottom, there is a "Translate" section with an "Add new translate" input field containing "เดินทาง" and a "Confidence score" dropdown menu. The dropdown menu is open, showing options: 5)-Very good, 4)-Good, 3)-Average, 2)-Poor, and 1)-Very poor. A red box highlights the "Add new translate" input field and the "Confidence score" dropdown menu. A blue box highlights the "Add comment" field. A green box highlights the search input field. Numbered callouts 1, 2, and 3 point to these elements respectively.

Participation (Vote)

Read the comment or memo



Vote



vote up



vote down

2

1



ExpertScore (score for voting)

- Each member is provided an initial value of voting score as 1
- The voting score is changed according to ExpertScore, which is estimated by the value of
 - Expertise
 - Contribution
 - Continuityof the participation history of each member

ExpertScore

- **Expertise:** a composite score of the accuracy of opinion and vote

$$Expertise = \alpha \frac{\text{count}(\text{BestOpinion})}{\text{count}(\text{Opinion})} + \beta \frac{\text{count}(\text{BestVote})}{\text{count}(\text{Vote})} \quad (1)$$

- **Contribution:** a composite score of the ratio of opinion and vote posting comparing to the total

$$Contribution = \gamma \frac{\text{count}(\text{Opinion})}{\text{count}(\text{TotalOpinion})} + \rho \frac{\text{count}(\text{Vote})}{\text{count}(\text{TotalVote})} \quad (2)$$

- **Continuity:** a regressive function based on the assumption that the absence of participation of a member will gradually decrease its ExpertScore to one after a 365 days of the absence

$$Continuity = 1 - \left(\frac{D}{365} \right)^4 \quad (3)$$



ExpertScore

$$\text{ExpertScore} = f(\text{Expertise}, \text{Contribution}, \text{Continuity})$$

$$\text{Expertise} = \alpha \frac{\text{count}(\text{BestOpinion})}{\text{count}(\text{Opinion})} + \beta \frac{\text{count}(\text{BestVote})}{\text{count}(\text{Vote})}$$

$$\text{Contribution} = \gamma \frac{\text{count}(\text{Opinion})}{\text{count}(\text{TotalOpinion})} + \rho \frac{\text{count}(\text{Vote})}{\text{count}(\text{TotalVote})}$$

$$\text{Continuity} = 1 - \left(\frac{D}{365} \right)^4$$

where

$$\alpha + \beta + \gamma + \rho = 1$$

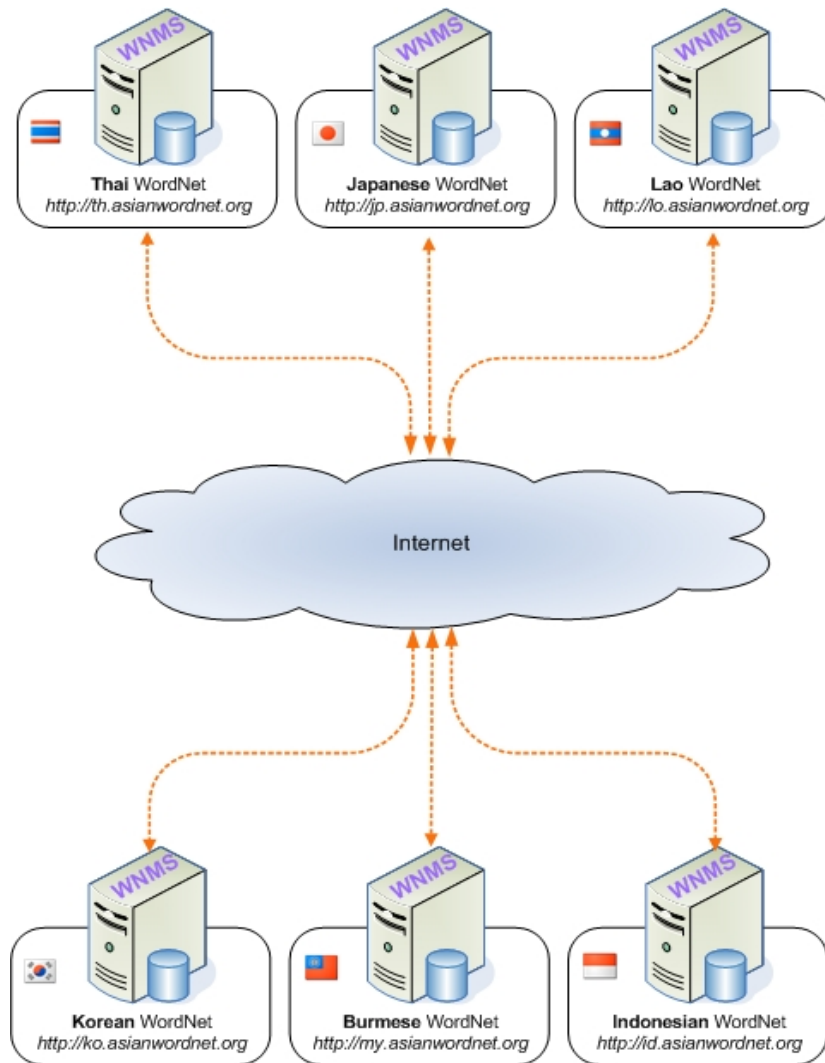
D = number of absent date

ExpertScore

$$ExpertScore = \left(1 - \left(\frac{D}{365} \right)^4 \right) \times \left\{ \begin{array}{l} \alpha \frac{count(BestOpinion)}{count(Opinion)} + \beta \frac{count(BestVote)}{count(Vote)} \\ + \gamma \frac{count(Opinion)}{count(TotalOpinion)} + \rho \frac{count(Vote)}{count(TotalVote)} \end{array} \right\}$$

- Range of the value of ExpertScore is 1 to 365 according to the accuracy and the rate of contribution of each member
 - Reliable members are rewarded better score for each vote
 - The expertise of the member is decreased by the discontinuity of the participation

Distributed WordNet Service

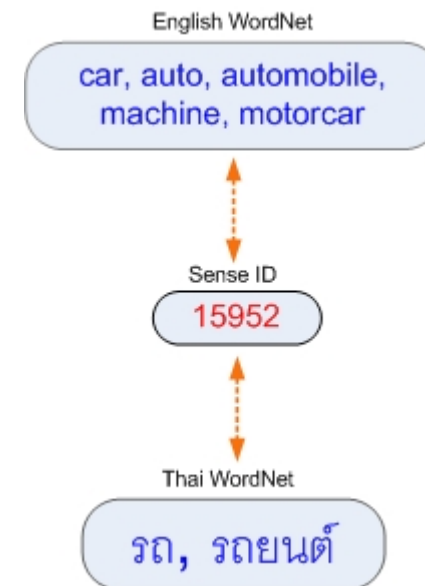


- Distribute the WordNet service node
- Service node can be locally maintained
- Synset ID (or Synset Offset) is the key to link between nodes

Representation of Synset Translation

wn30_data_pos	
sense_id	15952
synset_offset	02958343
lex_filenum	06
ss_type	n
w_cnt	5
words	car, auto, automobile, machine, motorcar
gloss	a motor vehicle with four wheels; usually propelled by an internal combustion engine; "he needs a car to get to work"

contribute		
contribute_id	40725	40808
sense_id	15952	15952
lex_filenum	06	06
message	รถ	รถยนต์
contributed_by	215296	215296
contributed_date	2007-06-12 09:50:42	2008-04-03 15:32:52
confidence_score	3	3
vote_score	101	101



Types of Services 'sense'

■ Thai Sense (Get word translation by POS and SYNSET_OFFSET)

Service URI : [http://th.asianwordnet.org/services/sense/output/\[callback\]/pos/synset_offset](http://th.asianwordnet.org/services/sense/output/[callback]/pos/synset_offset)

Service Name : sense

Parameter :

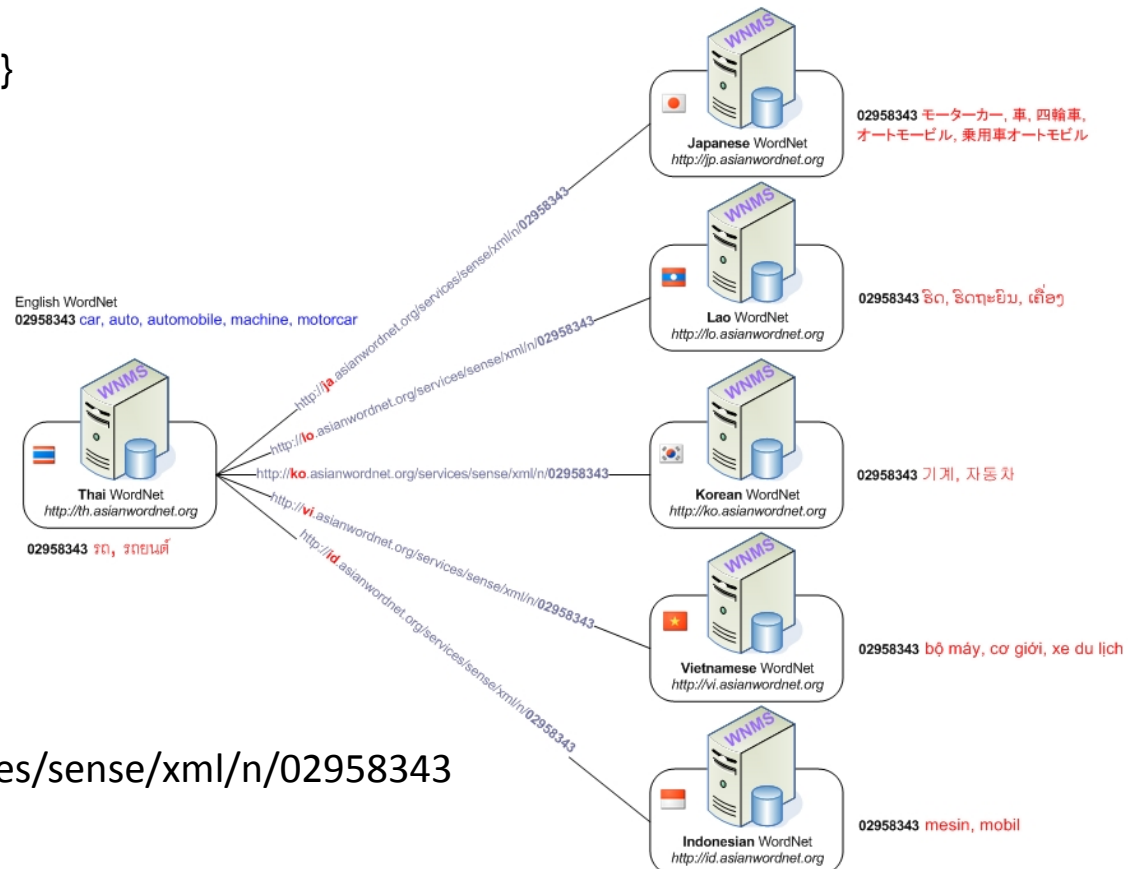
pos = PartOfSpeech {n,v,r,s}

synset_offset is an English

Princeton WordNet v.3.0

offset, represented

in 8 digits



<http://th.asianwordnet.org/services/sense/xml/n/02958343>

Types of Services 'dictionary'

- **E-Dictionary (Get word translation by word entry)**

Service URI :

[http://th.asianwordnet.org/services/dictionary/output/\[callback\]/type_of_dict/search_word](http://th.asianwordnet.org/services/dictionary/output/[callback]/type_of_dict/search_word)

Service Name : dictionary

Parameter : type_of_dict = {en2th, th2en}, search_word is a word you want to search

Bi-lingual dictionary
Let to search Asian WordNet dictionary.

from to

Result
Found : 3 item(s)

- [noun.phenomenon] {กระแสลม} = {breath} = {n/a}
- [noun.phenomenon] {กระแสลม} = {draft, draught} = {透間風, 隙間風}
- [noun.phenomenon] {กระแสลม} = {wind, air_current, current_of_air} = {風, 気流}



Types of Services

- **Auto complete (Get a list of words existing in WordNet by prefix auto completion)**

Service URI :

[http://th.asianwordnet.org/services/autocomplete/output/\[callback\]/language/search_word](http://th.asianwordnet.org/services/autocomplete/output/[callback]/language/search_word)

Service Name : autocomplete

Parameter : language = {en,th}, search_word is a word you want to get autocomplete (Result:limit 50 records found)

- **WN-Browser (Browse WordNet and its semantic relations)**

Service URI :

[http://th.asianwordnet.org/services/browse/output/\[callback\]/language/search_word](http://th.asianwordnet.org/services/browse/output/[callback]/language/search_word)

Service Name : browse

Parameter : language = {en,th}, search_word is a word you want to get all relations

Visualization of AWN

(<http://www.asianwordnet.org/>)

KUI - Knowledge Unifying Initiator | TCLLAB.ORG

Thai->Indonesian

The screenshot displays the Asian WordNet (AWN) interface. On the right, a search interface is shown with the following fields:

- Search: [Search]
- From: ภาษาไทย - Thai
- To: bahasa Indonesia - Indonesian
- Keyword: คำศัพท์
- Go button

Below the search interface is a "Help" section:

1. Drag a mouse when you want to focus a word.
2. Double click on word when you want to go that word visualization.
3. Try to search any word if you want to know.

At the bottom right, there is a logo for "Powered by NECTEC NICT TCL".

The main area shows two network visualizations of word relationships:

- Left visualization:** A network centered on the Thai word "รถยนต์" (car). It is connected to "รถยนต์" (car) and "รถจักรยานยนต์" (motorcycle). The Thai word "รถยนต์" is linked to the English word "car" via a "synset" relationship. Other related Thai words include "รถราง", "รถจักรยาน", "รถจักรยานยนต์", "รถกระบะ", and "รถ".
- Right visualization:** A network centered on the Thai word "คำศัพท์" (vocabulary). It is connected to "คำศัพท์" (vocabulary) and "คำศัพท์" (vocabulary). The Thai word "คำศัพท์" is linked to the English word "vocabulary" via a "synset" relationship. Other related Thai words include "คำศัพท์" (vocabulary) and "คำศัพท์" (vocabulary).



Guideline in WordNet Translation

- Word entry must be translated into the appropriate WORD(s) by avoiding phrase and meaning explanation.
- Words in a Synset must be interchangeable in a sentence.



Translational Issues

- There are many cases that a gloss need to be expressed in a phrase or explanation, especially in the case of technical terms and scientific vocabulary.

Ex. **Chaperon**

POS	Noun
Synset	chaperon, chaperone
Gloss	one who accompanies and supervises a young woman or gatherings of young people
Thai	ผู้ตามควบคุมหญิงสาว

- The concept is not common in the Thai language



Translational Issues (cont.)

- A gloss can be expressed by two or more Thai words. These words have the core meaning but occur in different context. Should it be divided into more specific concept?

Ex. **Appear**

POS Verb

Synset appear, come out

Gloss be issued or published; "Did your latest book appear yet?"; "The new Woody Allen film hasn't come out yet"

Thai T1 = ตีพิมพ์; T2 = ออกฉาย

- T1 occurs in the context of printed matter
- T2 occurs in the context of film or movie



■ For missing Synset

- => Computational model for the compound Synset (Synset composition/decomposition)
- => Adding new Synset with language/domain specific Synset tag

WNMS 2.0 (Langrid Plug-in)

The screenshot shows the homepage of the Language Grid website. At the top left is the Language Grid logo with the NICT acronym and a red square containing Japanese characters. To the right is a Google search bar with the text 'Search in site' and a 'Search' button. Below the search bar are links for 'English | Japanese' and a 'Sitemap' link. The main content area is divided into several sections: 'Trial site' with a description of the Language Grid Toolbox and an 'Enter' button; 'What's New' with a list of recent events and a 'Newsletter' link; 'What's Language Grid' with a description of the platform and a list of statistics and news items; 'Use Cases' with a 'Toolbox User List' containing various institutions; and 'Activity' with a table listing 'Service Grid', 'Service Grid Initiative', and 'Wikipedia Translation'. The bottom of the page features a navigation bar with these three categories.

Language Grid
NICT

Google Search in site Search
English | Japanese
Sitemap

Trial site
Language Grid Toolbox supports communication in a multilingual community. We offer a hosting service for user organizations.
[Language Grid Toolbox](#) Enter
At the Language Grid Playground, you can try a variety of language services registered in the Language Grid through a Web browser.
[Language Grid Playground](#) Enter

What's New [Newsletter](#)
2010/10/21 [Language Grid Bangkok Operation Center](#)
2010/10/21 [Report on Japan-Korea Webcam Exchange Event](#)
2010/10/21 [Wikipedia Translation Project Launched](#)
2010/10/21 [Language Grid Toolbox ver. 2.4 Released](#)

What's Language Grid
Language Grid is an online multilingual service platform which enables easy registration and sharing of language services such as online dictionaries, bilingual corpora, and machine translations. The Language Grid protects Intellectual properties of language resources and can be used for non-profit and research purpose.
131 organizations from 18 countries have participated in the Language Grid (2010/08)
The Language Grid is providing 91 language services (2010/08)
Federated operation starts between Kyoto University and NECTEC in summer 2010
Service Grid Open Source activity has started globally
Language Grid is aiming at helping translate Wikipedia articles
EU/NSF projects will use the source code of the Language Grid
Research papers on Language Grid are accepted by ICSOC, SCC, ICWS, etc
For more information about the Language Grid, please visit [NICT Language Grid Project page](#).

Use Cases
Toolbox User List
Kyoto University Global 30
NPO Pangaea
Center for Multilingual Multicultural Education and Research, Tokyo University of Foreign Studies
IT University of Copenhagen
House Saison
NPO Center for Multicultural Society Kyoto
Ritsumeikan University
Kyoto Information Card System (KICS-LLC)
Kodai-ji Zen Temple
Advanced Scientific Technology & Management Research Institute of KYOTO (ASTEM)

The Language Grid (Language Grid Operation Center)
The Language Grid is in

Service Grid	Service Grid Initiative	Wikipedia Translation
---------------------	--------------------------------	------------------------------



Summary

- Translation approach for Asian language Wordnet development
- Algorithm for synset alignment is proposed
- ExpertScore for weighting the vote
- Langrid plug-in