

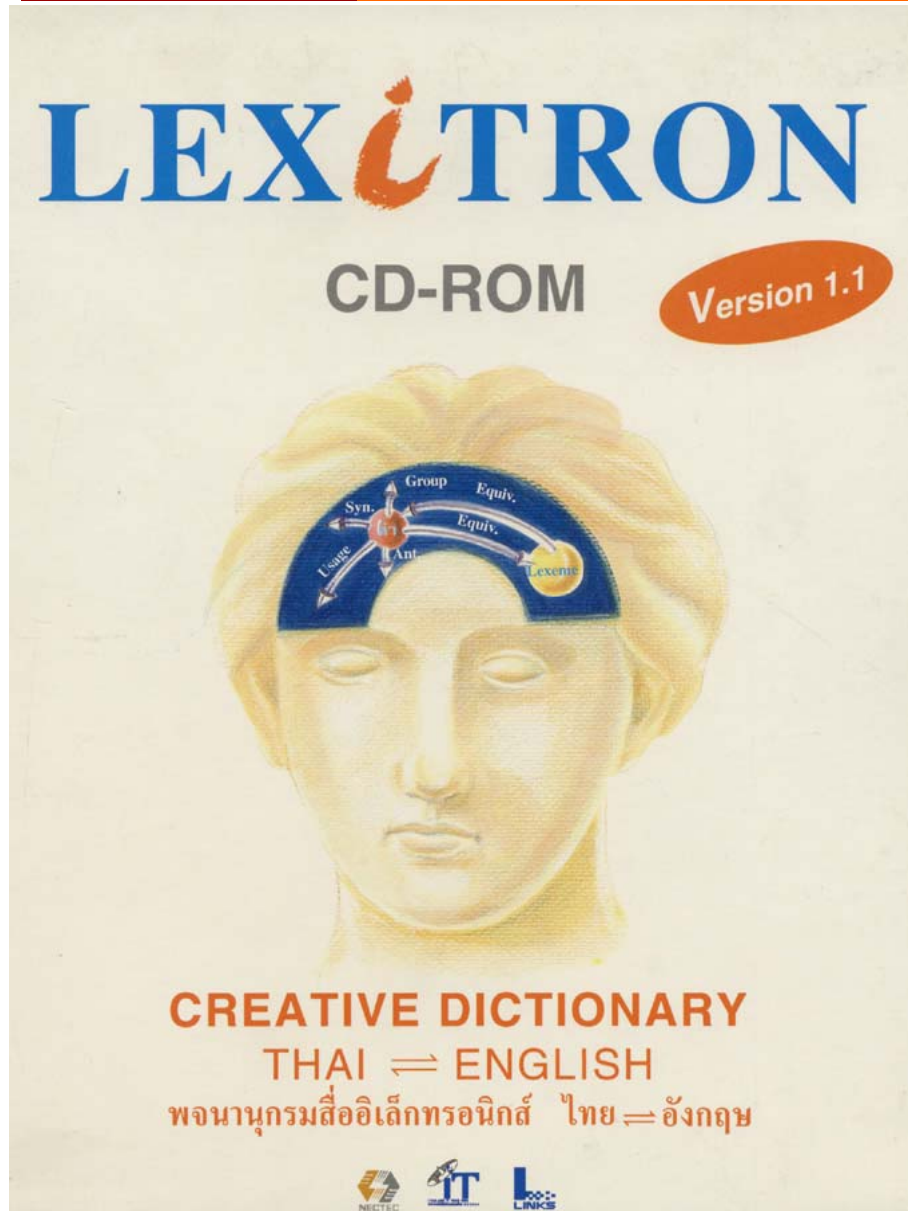
From LEXiTRON to Asian WordNet on Collaborative Development Platform



Virach Sornlertlamvanich

National Electronics and Computer Technology (NECTEC) NSTDA,
Thailand

virach@tcllab.org

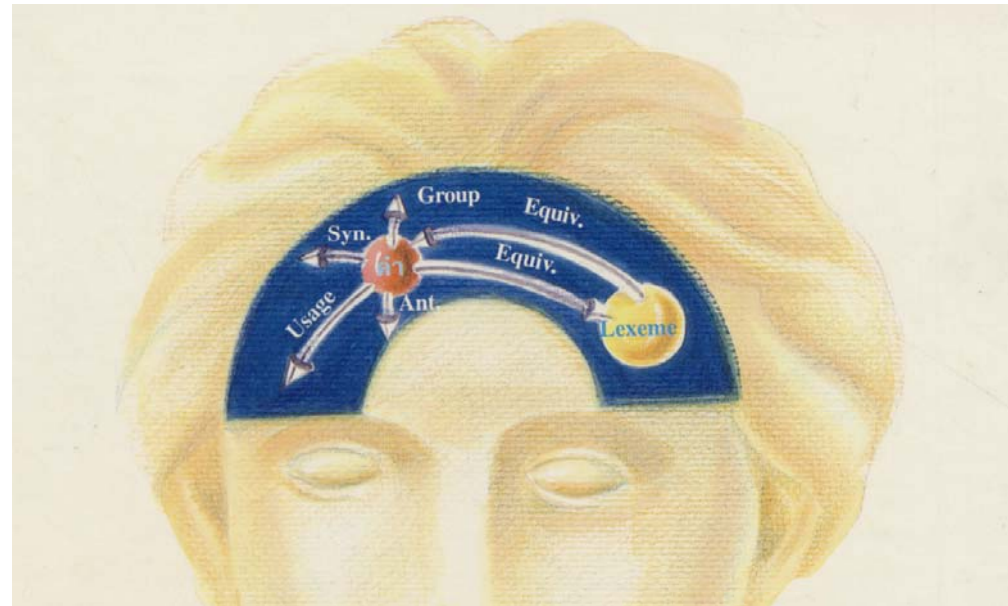


LEXiTRON version 1.1

- Corpus-based dictionary
- Dictionary for writing
- เผยแพร่ในปี 2538
- CD-ROM สำหรับ Windows 3.1 Thai Edition
- ไทย 11,000 คำ; อังกฤษ 9,000 คำ
- 6 พจนานุกรมในหนึ่งเดียว
 - 1) พจนานุกรมไทยทั่วไป
 - 2) พจนานุกรมการใช้ภาษาไทย
 - 3) พจนานุกรมคำเหมือนคำตรงข้าม
 - 4) พจนานุกรมไทยอังกฤษ
 - 5) พจนานุกรมกลุ่มคำไทย

Corpus-based Dictionary and Dictionary for Writing

- การเข้าถึงคำ
 - คำเหมือน (synonym)
 - คำตรงข้าม (antonym)
 - ตัวอย่างประโยค (usage)
 - กลุ่มคำ (word group)
 - คำแปล (equivalent)



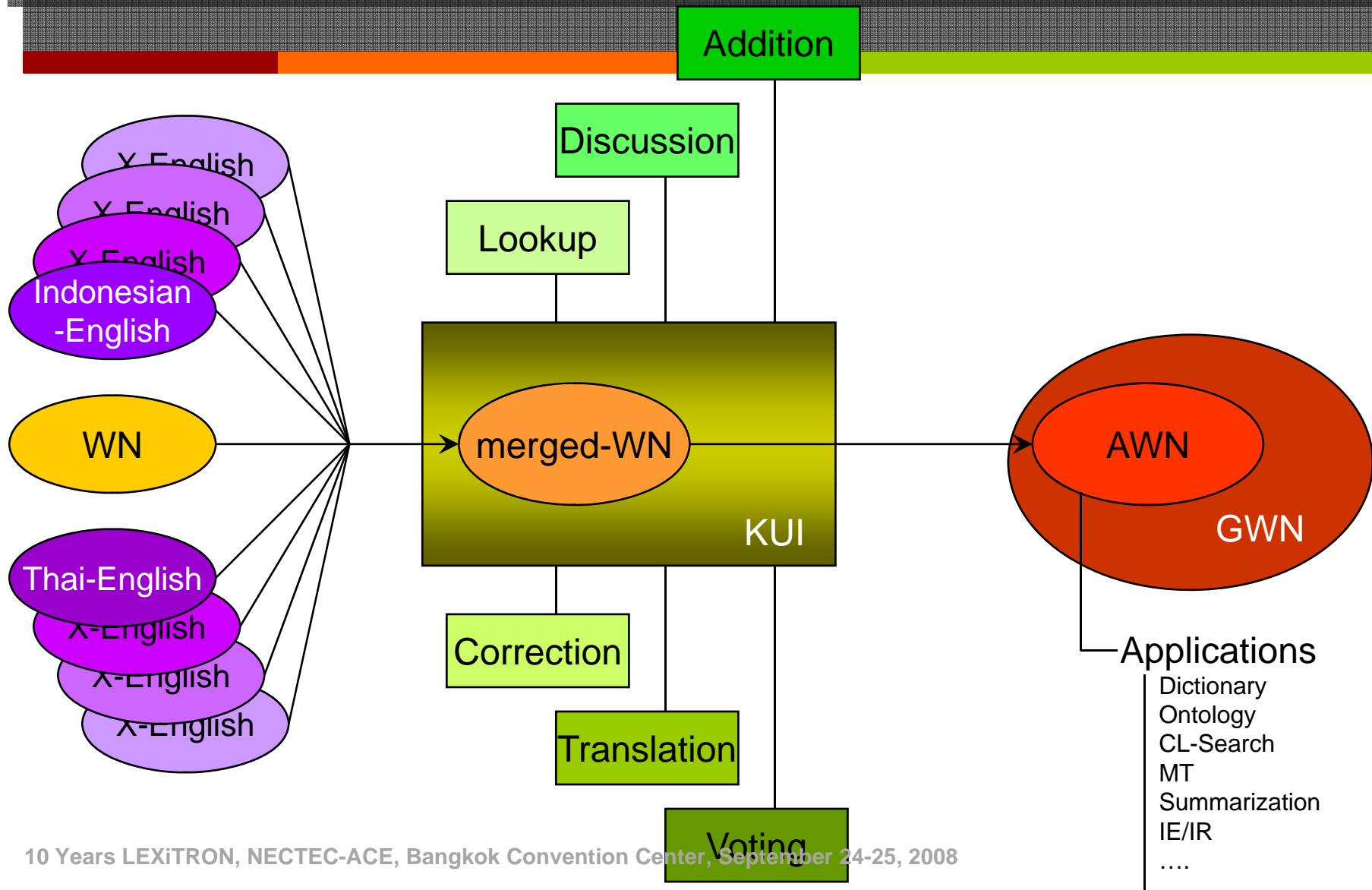
Design of LEXiTRON

- สร้างจากพจนานุกรมสำหรับระบบแปลภาษา 30,000 คำ
- ข้อมูลของคำ
 - คำ
 - คำอ่าน
 - ประเภทของคำ (หลัก 14, ย่อย 45)
 - คำลักษณนาม
 - Verb pattern (12 -> 9 VPs)
 - คำเหมือน
 - คำตรงข้าม
 - ตัวอย่างประโยค
 - คำแปลภาษาอังกฤษ
 - กลุ่มความหมาย
- โครงสร้างคำ
 - คำเดี่ยว
 - คำประสม
 - Prefix
 - Suffix

Synset Assignment via English Surface

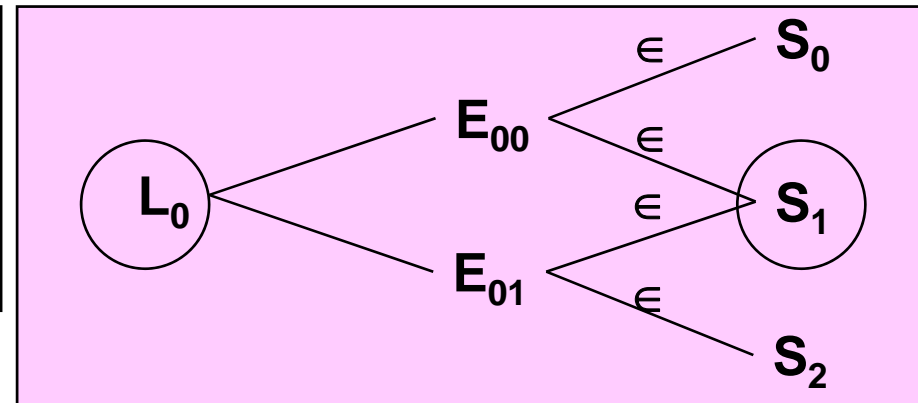
- Use English equivalents to link the existing dictionary to WordNet
- POS (n, v, adv, adj), English equivalent, and English equivalent of synonym of the target language are used to pinpoint the appropriate link
- Number of matched English equivalents in the Synset confirms the appropriate link
- Experiment on Thai-English, Indonesian-English and Mongolian-English dictionaries

Asian WordNet Development



Synset Assignment (CS=4)

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



Example:

L0: เป้าหมาย

E0: aim

E1: target

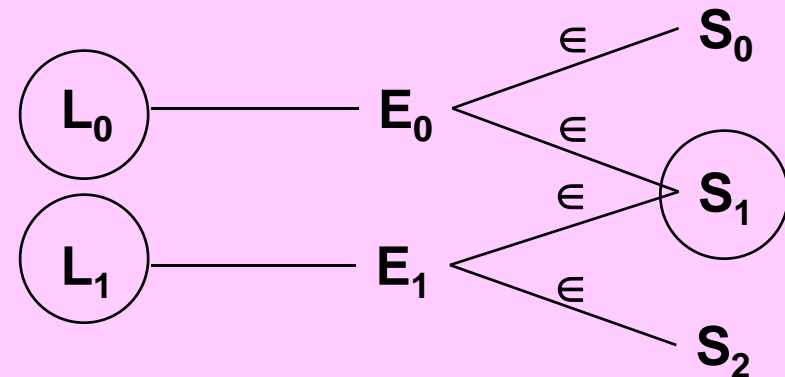
S0: purpose, intent, intention, aim, design

S1: aim, object, objective, target

S2: aim

Synset Assignment (CS=3)

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



Example:

L0: จ้อง

L1: เพ่งมอง

} Synonym

E0: stare

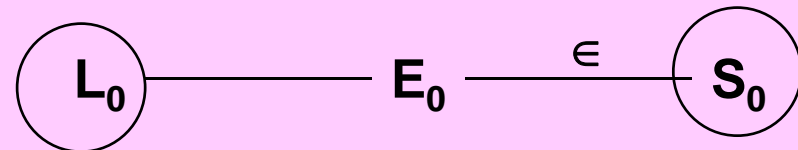
E1: gaze

S0: stare

S1: gaze, stare

Synset Assignment (CS=2)

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



Example:

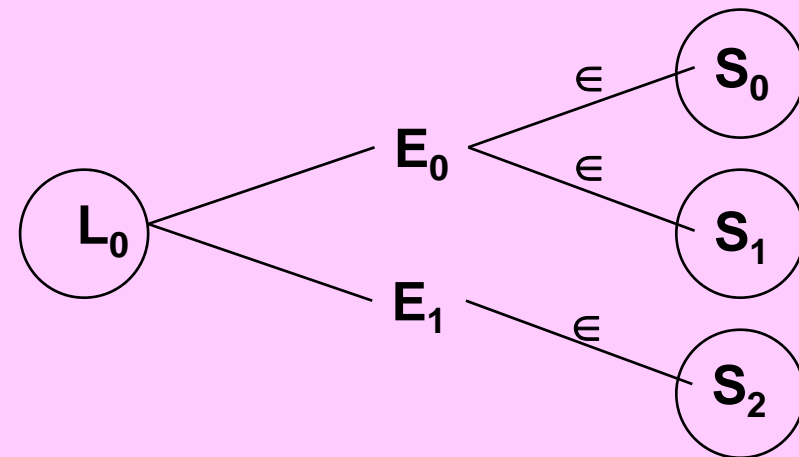
L₀: สูติแพทย์

E₀: obstetrician

S₀: obstetrician, accoucheur

Synset Assignment (CS=1)

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



Example:

L0: ช่อง

E0: hole

E1: canal

S0: hole, hollow

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

S2: canal, duct, epithelial duct, channel

Quantitative Evaluation for T-E

	WordNet (synset)		T-E Dict (entry)	
	total	assigned	total	assigned
Noun	145,103	18,353 (13%)	43,072	11,867 (28%)
Verb	24,884	1,333 (5%)	17,669	2,298 (13%)
Adjective	31,302	4,034 (13%)	18,448	3,722 (20%)
Adverb	5,721	737 (13%)	3,008	1,519 (51%)
total	207,010	24,457 (12%)	82,197	19,406 (24%)

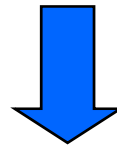
Qualitative Evaluation for T-E

	CS=4	CS=3	CS=2	CS=1	total
Noun	5 (71.4%)	306 (63.9%)	34 (53.1%)	55 (20.2%)	400 (48.7%)
Verb		23 (52.3%)	6 (8.0%)	4 (13.8%)	33 (22.3%)
Adjective		2 (8.0%)			2 (3.4%)
Adverb	7 (100%)	4 (100%)	4 (100%)	1 (100%)	16 (100%)
total	12 (80.0%)	335 (60.7%)	44 (30.8%)	60 (18%)	451 (43.2%)

Improvement by Consulting Dictionaries from Multiple Sources

MMT T-E Dictionary

	CS=4	CS=3	CS=2	CS=1	total
Total	12 (80.0%)	335 (60.7%)	44 (30.8%)	60 (18%)	451 (43.2%)



MMT and LEXiTRON T-E Dictionary

	CS=4	CS=3	CS=2	CS=1	total
Total	14 (93.3%)	337 (61.1%)	72 (50.3%)	93 (27.8%)	516 (49.4%)

Translating ... Convert the existing knowledge to your own language community.

Show/Hide

- [Introduction](#) |
 [Task List](#) |
 [Lookup](#) |
 [Forum](#) |
 [Participation](#) |
 [New Task](#)

Remark
 active non-act expired
 inert private

- Task List
 Active tasks only
- Asian WordNet (ADJECTIV ..
 - Asian WordNet (ADVERB)
 - Asian WordNet (VERB)
 - Asian WordNet (frequent ..**
 - Asian WordNet (other NO ..
 - Let's Latin!
 - Medical Translation
 - NICT's Japanese-English ..
 - OSS Glossary
 - OpenOffice Glossary
 - SWADESH LIST
 - Sample English-Nepali
 - Technical Term Dictiona ..
 - tcl-ALRN web site gloss ..
 - tcl-KUI web site glossa ..

- Item List
 Item Search
- 1-100 Of 200 Item
- 1) car
 - 2) car
 - 3) car
 - 4) car
 - 5) car
 - 6) care
 - 7) care
 - 8) care
 - 9) care
 - 10) care
- *green:opinions exist
- Go2 Item

Task Title : Asian WordNet (frequent NOUN)

Task group :

Description : Let's build Asian WordNet together. Starting from Princeton's WordNet 2.1, we generate a list of word entries and provide some information on POS, Synset and Gloss for understanding the exact MEANING of the word entry. Participants are to read carefully to understand the MEANING and post only an appropriate WORD (avoid phrase and explanation).

Quantity : 28,421 Items

[\[Edit task\]](#)

car

« Previous | Next »

car

Hint Optional
 POS : NOUN
 Synset : [car_elevator_car](#)
 Gloss : where passengers ride up and down;

List of opinion

Sort By Date

Sort By Vote

« 1-4 From 4 Result »

No	Result	Score
1	VOTE ลิฟท์	100
2	VOTE รถกระเช้า	1
3	VOTE エレベーター	1
4	VOTE mobil	1

Post your opinion

Description : You can post your translation for Translating; opinion for Polling; or suggestion for Public Hearing in the following Opinion box then click Post.

* **Opinion :** (2048 Chars)

- Contributor
- pakapril (8019)
 - n4n4 (2360)
 - ttan (1764)
 - nipa (1522)
 - t_saree (435)
 - zarin44 (403)
 - Purev (299)
 - virach (55)
 - shruti (37)
 - nyanbobo (18)

User online

Chat / Remark

[0 users] on the last 5 minutes

Submit

modules by © grochat

Task in progress

Japanese	20424	71.9%
Thai	14356	50.5%
Indonesian	6071	21.4%
Bengali	403	1.4%
Mongolian	172	0.6%

Translating ... Convert the existing knowledge to your own language community.

Show/Hide

[Introduction](#)

[Task List](#)

[Lookup](#)

[Forum](#)

[Participation](#)

[New Task](#)

Remark
 active non-act expired
 inert private

Task List
 Active tasks only

- Asian WordNet (ADJECTIV ..
- Asian WordNet (ADVERB)
- Asian WordNet (VERB)
- Asian WordNet (frequent ..**
- Asian WordNet (other NO ..
- Let's Latin!
- Medical Translation
- NICT's Japanese-English ..
- OSS Glossary
- OpenOffice Glossary
- SWADESH LIST
- Sample English-Nepali
- Technical Term Dictiona ..
- tcl-ALRN web site gloss ..
- tcl-KUI web site glossa ..

Task Title : Asian WordNet (frequent NOUN)

Task group :

Description : Let's build Asian WordNet together. Starting from Princeton's WordNet 2.1, we generate a list of word entries and provide some information on POS, Synset and Gloss for understanding the exact MEANING of the word entry. Participants are to read carefully to understand the MEANING and post only an appropriate WORD (avoid phrase and explanation).

Quantity : 28,421 Item

Keyword : [Advanced Search](#)

Type : Exact Partial

◀ 1-5 From 5 Result ▶

No	Result
1	<p>car (cable car, car) <i>a conveyance for passengers or freight on a cable railway;</i></p> <p>Japanese : ケーブルカー</p> <p>Thai : รถกระเช้า</p> <p>Indonesian : mobil</p>
2	<p>car (car, auto, automobile, machine, motorcar) <i>a motor vehicle with four wheels; usually propelled by an internal combustion engine;</i></p> <p>Japanese : 自動車</p> <p>Thai : รถยนต์</p> <p>Indonesian : mobil</p>
3	<p>car (car, elevator car) <i>where passengers ride up and down;</i></p> <p>Japanese : エレベーター</p> <p>Thai : ลิฟท์</p> <p>Indonesian : mobil</p>
4	<p>car (car, gondola) <i>the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant</i></p> <p>Japanese : ゴンドラ</p> <p>Thai : รถกระเช้า</p>

Language

User online

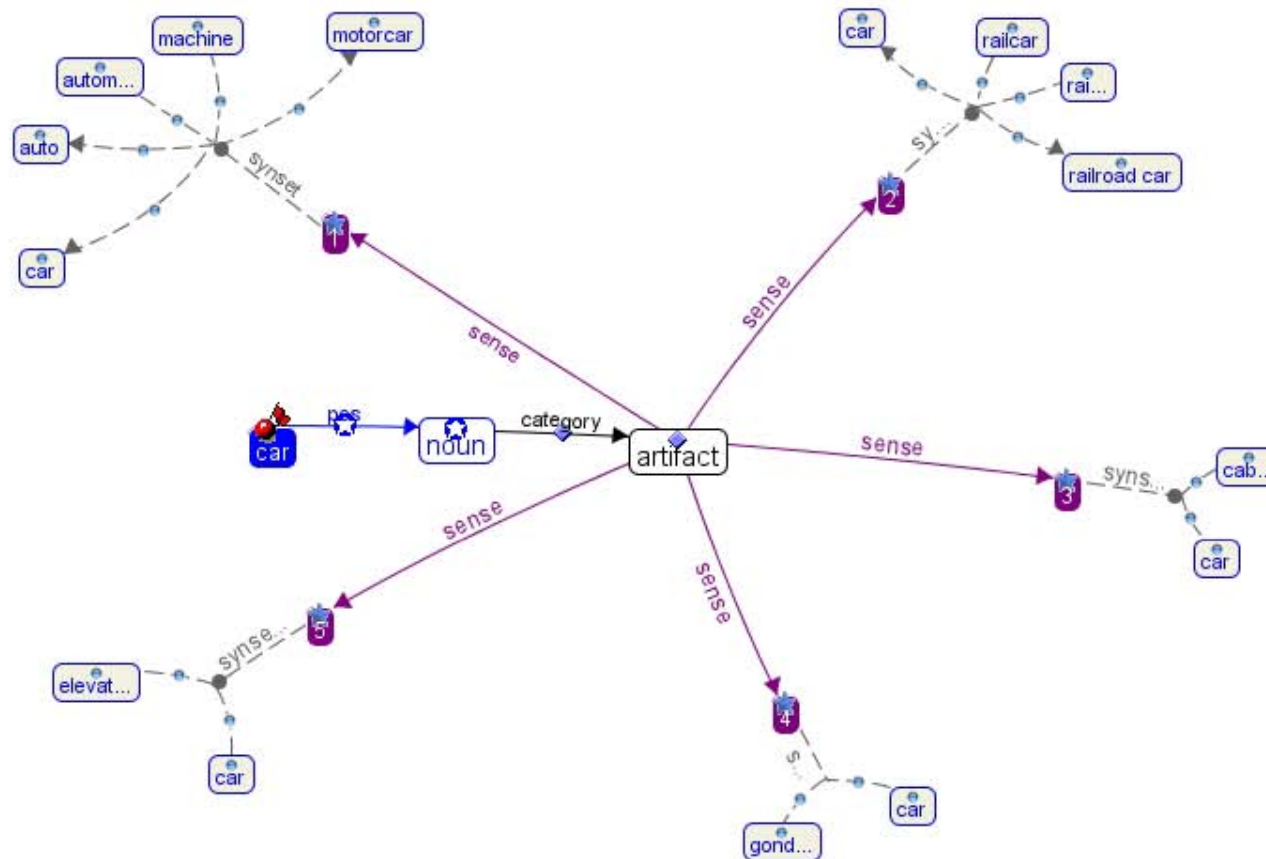
Task in progress

Japanese	20424	71.9%
Thai	14356	50.5%
Indonesian	6071	21.4%
Bengali	403	1.4%
Mongolian	172	0.6%
Nepali	39	0.1%
Burmese	18	0.1%
Arabic	7	0.0%
Khmer	6	0.0%
Vietnamese	3	0.0%
Chinese, Mandarin	2	0.0%
French	1	0.0%
English	1	0.0%

Powered By

English-English

KUI - Knowledge Unifying Initiator | TCLLAB.ORG



Search

From:

English - English

To:

English - English

Keyword :

car

Go

Help

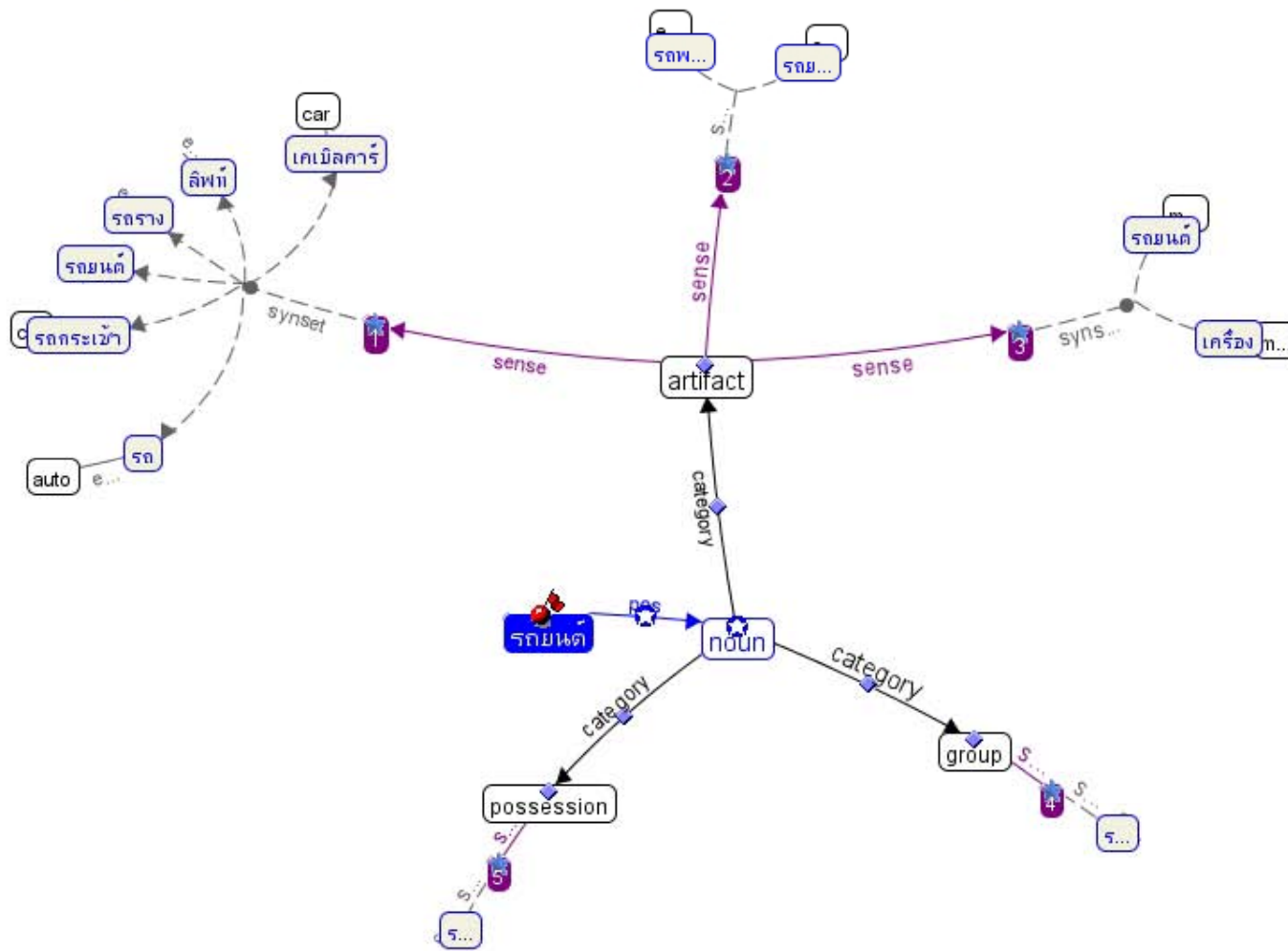
1. Drag a mouse when you want to focus a word.
2. Double click on word when you want to go that word visualization.
3. Try to search any word if you want to know.

Powered by

NECTEC
a member of NECTEC
NICT TCL

Thai-English

KUI - Knowledge Unifying Initiator | TCLLAB.ORG



Search

From:

To:

Keyword :

Help

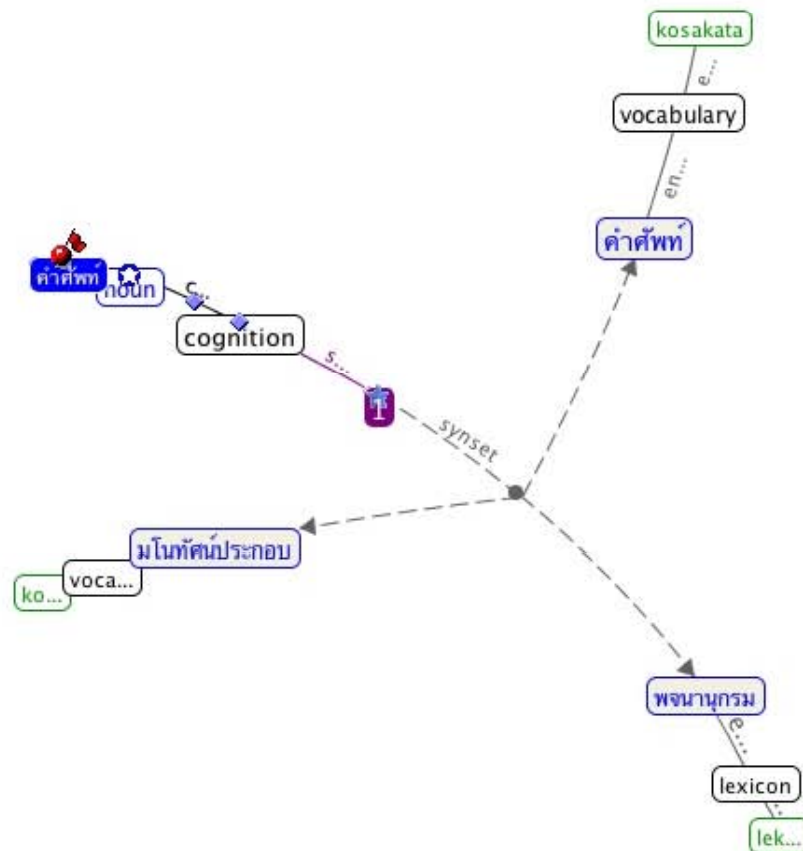
1. Drag a mouse when you want to focus a word.
2. Double click on word when you want to go that word visualization.
3. Try to search any word if you want to know.

Powered by



Thai-Indonesian

KUI - Knowledge Unifying Initiator | TCLLAB.ORG



Search

From:

ภาษาไทย - Thai

To:

bahasa Indonesia - Indonesian

Keyword :

คำศัพท์

Go

Help

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.

Powered by

NECTEC
a member of NSTDA
NICT TCL

Future Work

- Asian WordNet Community
- Language resource conversion and alignment
- Language technology sharing
- Collaborative development platform

AsianWordnet
(www.tcclab.org/kui -> www.asianwordnet.org)