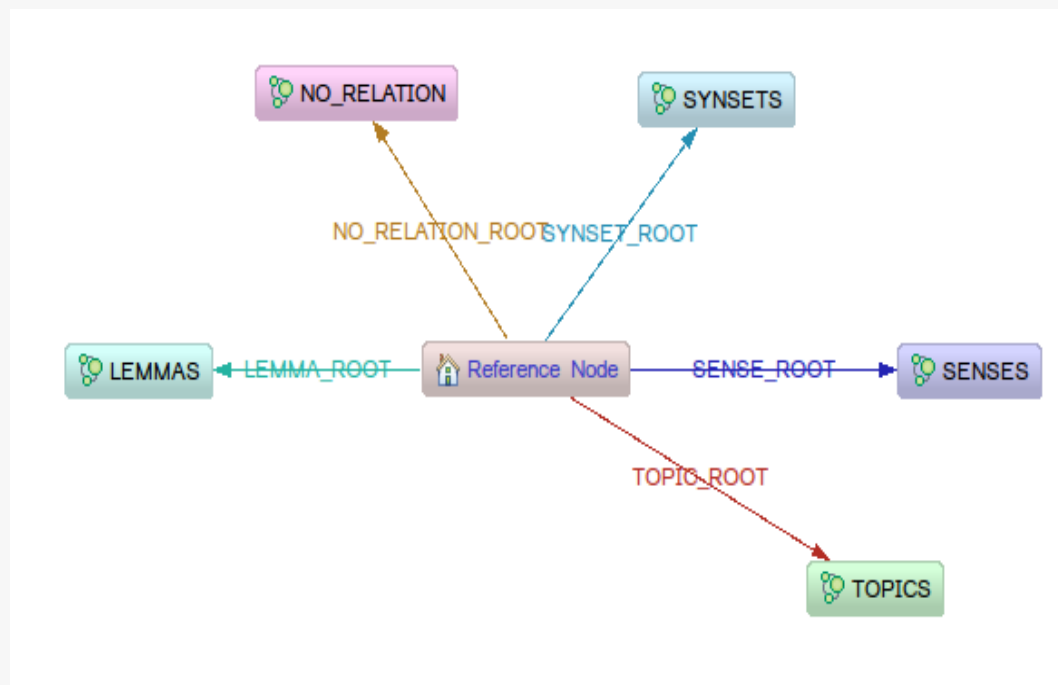


Thai Web Forum Topic Suggestion Using Thai WordNet Graph Semantic Relations



Seksan Poltree, Kanda Saikaew
Department of Computer Engineering
Faculty of Engineering, Khon Kean University
seksan.poltree@gmail.com, krunapon@kku.ac.th

Outline

- Why web forum and what's the problem?
- Related works
- System overview
- Experiment and evaluation result
- Discussion
- Conclusion and future work

Why Web Forum?

- Question answering method: real life used
- Web forum is a popular searching / sharing knowledge information
- Most of web forum contains existing information
- It has questions and its answers
- Topic \sim Question
Reply \sim Answer



What's the Problem?

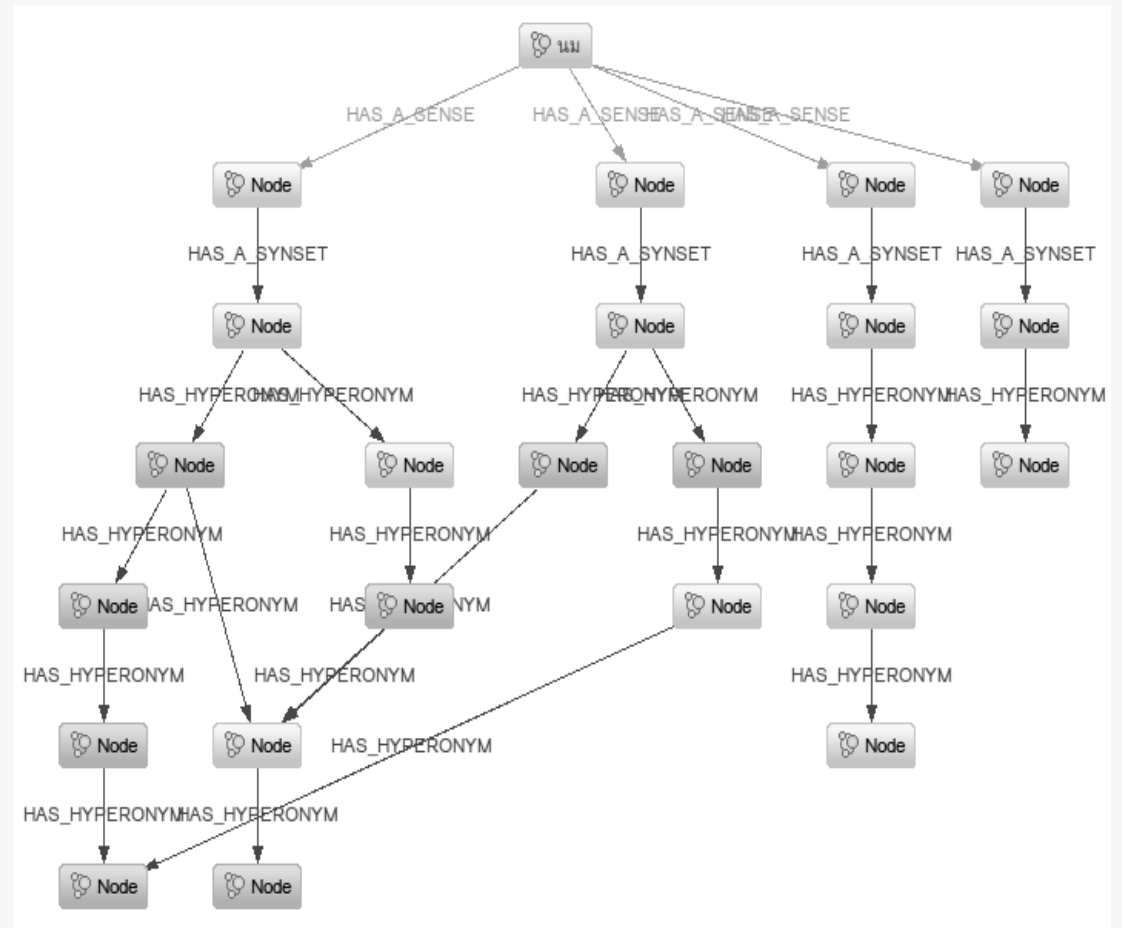
- Ambiguity of Thai words, sentence forms
- **Topic duplication** : A user asked the question that has already been answered repeatedly
- A large web forum has duplicated searching results, unuseable results



How to increase searching topic result relevancy to avoid duplication?

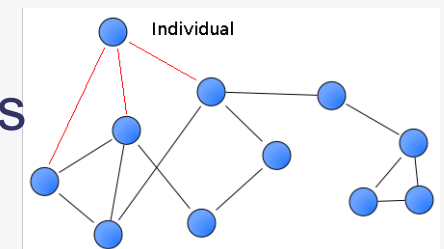
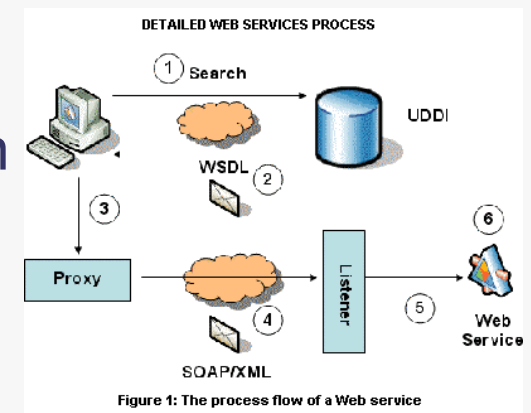
Thai WordNet

- A part of Asian Wordnet
- LMF : XML based Format
- A Translation of Princeton wordnet
- Word relations as a synsets graph (eg. Hyperonyms/ Hyponyms/ Meronyms/Holonyms)
- We suppress unused information and rebuild the relation graph
- We propose to reuse this relation information

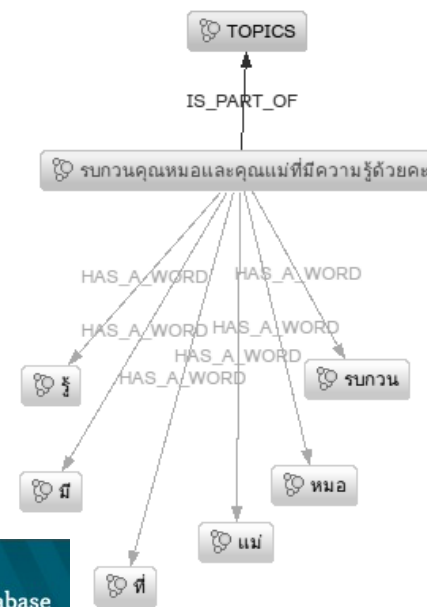
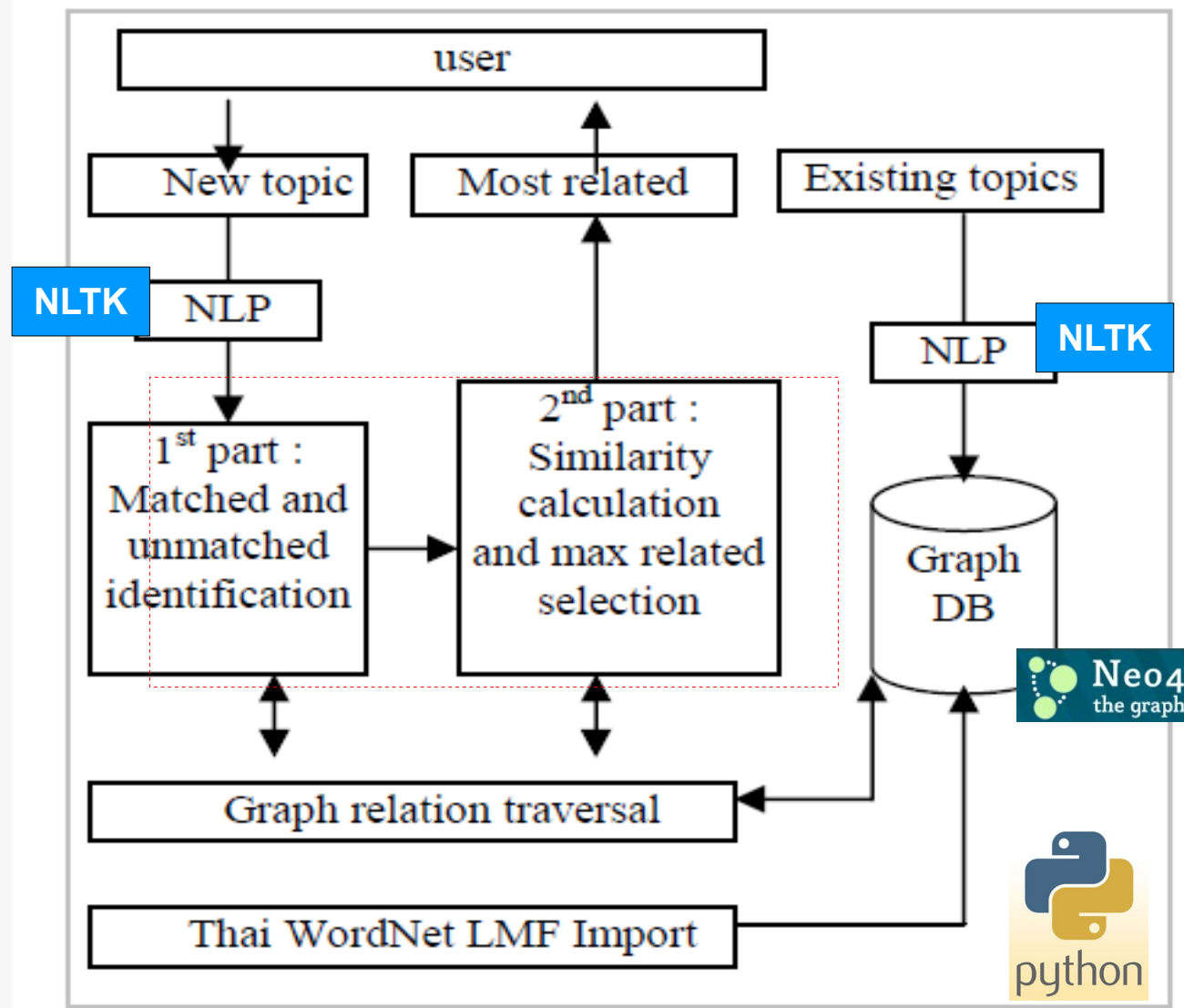


Related works

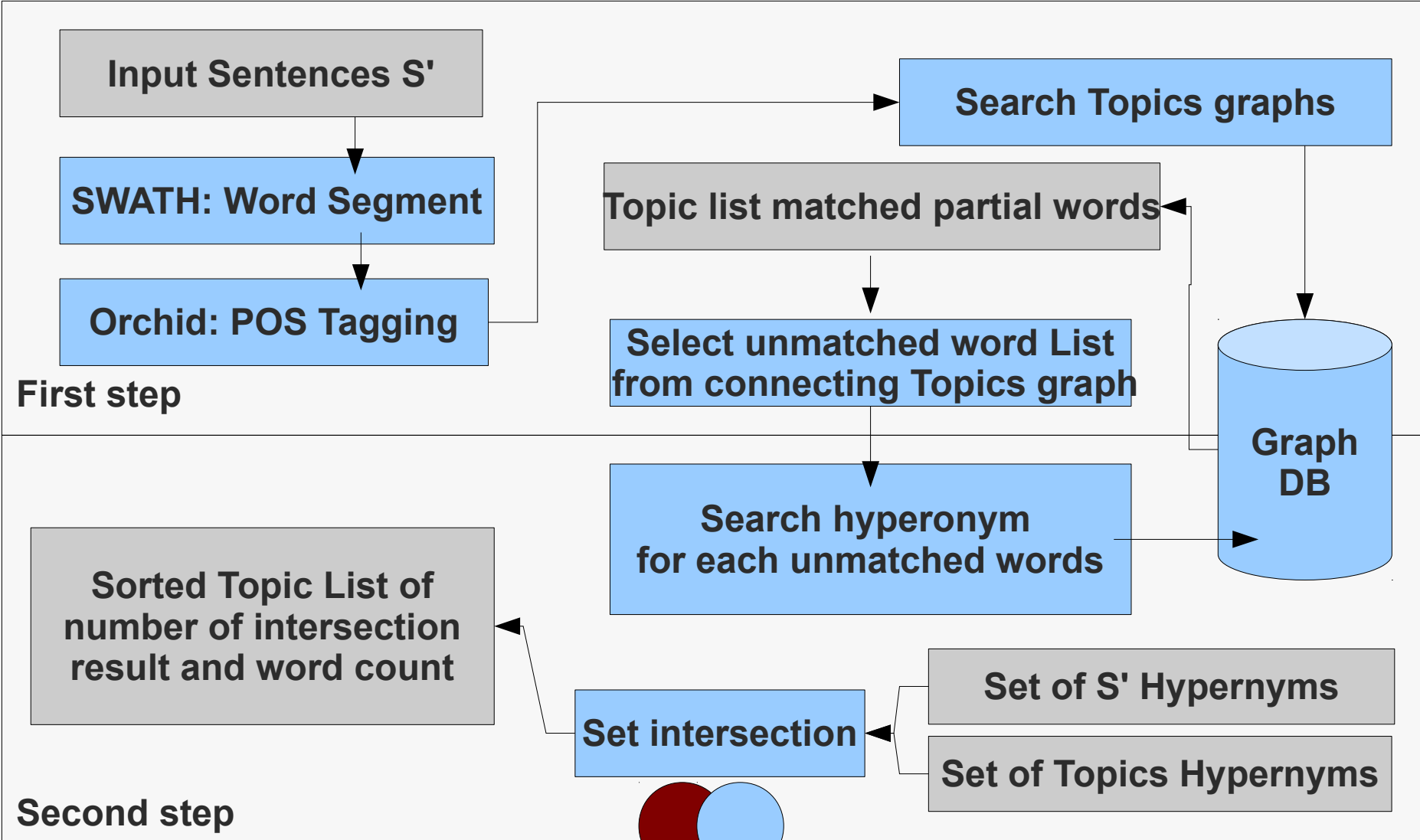
- There is an experience database built by its source, expericer, and indexes its information for searching using NLP.
- There is a web service with predefined domain searching through web services template based on extracted input question.
- Another one is a web service searching through semantic web services for answer generator.
- Searching pre-generated ontology with graph based is an exist solution.



System overview



Similarity and max related calculation



Experiment

- Random select exist 7,000 random topics as input by length*
 - Short input : sentence length < 25 words
 - Medium input: sentence length > 25 word and < 80 words
 - Long input: sentence length > 80 word (80 - ~200 words)
- Manually insert topic
 - Non related topics

* average topic length is about 100 words

** more topics take very long calculation time : current average 30s per topic result

Evaluation result

Random Input	Result	Example
Short Input < 25 words	<ul style="list-style-type: none"> • Very short topic, 5 words • Most of non-related topic shown • More unmatched words calculation time	รบกวนคุณหมอมและคุณแม่ที่มีความรู้ด้วยค่ะ - ขอเสียดคุณแม่ที่ให้นมแม่ล้วนไม่เสริมนมใดจนลูกเข้าอนุบาลเลยหนอยจ้า [187] - ถ้าตั้งครรถ์ลูกคนที่ 2 สามารถให้นมแม่กับลูกคนแรก [181] - ควรเสริมนมผง คุณแม่ตอนลูกอายุเท่าใดคะ [152] - เรียนถามคุณหมอมถึงเคล็ดลัการให้นมลูกถึงปัจจุบัน [149] - ยาแก้ท้องอืดและแก้อาเจียนกับการให้นมลูก [141] - ถ้าลูกน้ำหนักไม่ขึ้นหมายความว่าน้ำหนักแม่ไม่พอหรือเปลาค่ะ [137] - รายงานผล...ปฏิบัติการกุ่มแม่ [89]
Medium Input 25>i>80	<ul style="list-style-type: none"> • Better result if it has some matched words, • Re-ranking improve the results • First 5 topics 	ลูกชายอายุหกเดือนกับสามวันคะ ฟีนน้ำหนักเริ่มขึ้นร้องงแงมาก ๆ ช่วงกลางคืน ถ้าจะให้เค้ากินยาแก้ปวดลดไข้ได้รีปาวคะ ยี่ห้อไหนปราคะ 48131 อยากทราบทำให้ลูกกินน้ำลูกพรุนทุกวันได้ไหมคะ 114 11865 เราจะให้ลูกกินได้ตอนกี่เดือนคะ 108 38180 ลูกสามเดือนขึ้นไปจะร้องกินนมน้อยลงจริงหรือคะ 98 72807 ปรีกษาคุณหมอมเริ่มท้อ จะให้นมแม่จากขวด ลูก 19 วัน 92 58957 เป็นหวัดกินยาจะให้นมลูกได้หรือไม่ 90
Long input > 80 words	<ul style="list-style-type: none"> • More slow if there is more unmatched words 	ไม่ได้ต้องเสียนะคะ ไม่รู้ว่าจากสาเหตุอะไร ตอนแรกว่าอาจเกิดจากผ้าอ้อมอับชื้น ที่จิมหนูก็เป็นแดงๆ แต่ก็พยายามตลอดทั้งจิมทั้งกัน จิมเริ่มดีบ้าง หายบ้าง เป็นอีกสลับไป แต่ก็ไม่หายซะที่แดงตลอด ไม่รู้เค้าจะเจ็บหรือเปล่านะ จะเกี่ยวกับการแพ้อะไรได้บ้างมั้ยคะ ที่ผ่านทางนมแม่คะขอบคุณมากคะ 44216 ขอเคล็ดวิชาที่ทำให้ลูกได้ดื่มนมแม่ได้นานที่สุดทีค parameter 133 51315 ส่งนมแม่ทางไปรษณีย์ได้ไหมคะ parameter 126 53632 อาหารแสลงสำหรับแม่ที่กำลังให้นมลูกทางช่อง 3 parameter 124 35880 น้ำแม่ที่ลูกอายุ10เดือนเอาให้เด็กแรกเกิดกินได้ไม้ parameter 116 13584 จะเก็บนมแม่ที่ลูกกินไม่หมดไว้มี้อต่อไปได้อย่างไร parameter 109

Example result

- **TOPIC** : นมที่แช่ในช่องแข็ง นำมาไว้ในตู้เย็นช่องปกติและละลายแล้ว แต่ยังไม่ได้อุ่นจะนำกลับไปแช่ในช่องแข็งเหมือนเดิมได้หรือไม่ (About Freezing milk)
- **Found Partial matched** : 29
- **Re-ranking result:**
 - 40764 ชื่อสื่อที่สามารถเปิดช่องให้นมแม่ได้เลยที่ไหนกันค parameter 85
 - 34618 นมแม่ที่ฟรีสแล้ว แช่ต่อช่องธรรมดาได้อีกกี่วัน parameter 73
 - 72919 นมบี้มเสร็จรอใส่รวมกัน แช่ไว้ในช่องฟรีชได้มั้ยคะ parameter 69
 - 56651 การขนนมจากช่องแช่เย็นธรรมดาที่บ้านไปฟรีชที่ทำงาน parameter 62
 - 53685 สามารถนำนมที่ละลายแล้วไว้นอกตู้เย็นได้นานเท่าไรคะ parameter 60
 - 56330 นมแช่ฟรีชที่ละลายแล้ว เอามาฟรีชใหม่ได้มั้ยคะ parameter 55
 - 63853 นมที่ตู้แช่แข็งละลาย นำมา ฟรีชใหม่ได้ไหมคะ parameter 51
 - 56400 นมฟรีชที่ละลายแล้ว เก็บไว้ได้นานแค่ไหนคะ parameter 46

Discussion

- More unmatched words takes more times.
- Short input sentence has a poor result
 - because the relevancy information cannot be found.
- Thai Wordnet is a translation based system. Some words cannot be replaced by Thai word.
- So, Try to use only nouns and verbs in the topic has closely the same as using all topic words
- For longer topics, it could have some inexact match keywords (nouns/verbs), and the results could be better.
- Current result, acceptable relevancy topic about **60%** by observation
 - It might be because the topic is very short or its ambiguity.

Conclusion and future work

- We use the hyperonym relation in the graph database to calculate the additional weight for max relevancy sentences
- It has better result if there are some related inexact word for incoming post, bad result if no relevance or very short topic
- Future propose works
 - Need to speed up graph traversal / topic searching or indexing
 - Latent Semantic Indexing: LSI
 - Try to use Clustering
 - Try to use other relations, eg. meronym/holonym

Reference

- [1] Google Inc, “Google trends”, <http://www.google.com/trends>
 - access on May 15, 2011
- [2] Google Inc, “Google search”, <http://www.google.com/>
 - access on May 15, 2011.
- [3] Google Inc, “Google guru”<http://guru.google.co.th/guru/>
 - access on May 15, 2011
- [4] Yahoo Inc, “Yahoo answers”, <http://answers.yahoo.com/>
 - access on on May 15, 2011
- [5] K. Inui, S. Abe, K. Zuo, K. Morita and C. Sao, “Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents”,
 - IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.
- [6] G. W. Chen and M. M. Chiu, “Online Discussion Process: Effects of Earlier Messages Evolutions, Knowledge Content, Social Cues and Personal Information on Later Messages”,
 - In preceeding of Sixth International Conference on Advanced Learning Technologies, 2006.
- [7] Zhe Chen and Dunwei Wen, “A New Web-Service-Based Architecture for Question Answering”, In preceeding of
 - IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005.
- [8] M. Jang, J. Sohn, and H. Cho, “Automated Question Answering using Semantic Web Services”, IEEE Asia-Pacific Services Computer Conference, 2007.
- [9] N. Duhan, A. K. Shama, and K. K. Bhatia, “Page Rank Algorithms: A Survey”, IEEE International Adcaned Computing Conference (IACC), 2009.
- [10] Sareewan Thoongsup, Thatsanee Charoenporn, Kergrit Robkop, Tan Sinthurahat, Chumpol Mokrat, Virach Sornlertlamvanich and Hitoshi Isahara., “Thai WordNet Construction”, Proceedings of The 7th Workshop on Asian Language Resources (ALR7), Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics (ACL) and the 4th International Joint Conference on Natural Language Processing (IJCNLP), Suntec, Singapore, August 6-7, 2009.
- [11] G. van Rossum, “Python tutorial, Technical Report CS-R9526”, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.
- [12] Neo Database AB, “The Neo Database - A Technology Introduction”, 2006.
- [13] S. Bird and E. Loper, “NLTK: The Natural Language Toolkit”, Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, 2002, pp. 62-69.
- [14] T. Karoonboonyanan, “libthai Library”,
 - <http://libthai.sourceforge.net/> access on May 15, 2011