

การแปลคำสั่งสอบถามในการค้นคืนข้ามภาษา
ภาษาไทย-ภาษาอังกฤษ โดยใช้คอร์ปัสแบบขนาน

THAI-ENGLISH QUERY TRANSLATION
IN CROSS-LANGUAGE INFORMATION RETRIEVAL
USING PARALLEL CORPUS



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2547

ISBN 974-9700-77-2

การแปลคำสั่งสอบถามในการค้นคืนข้ามภาษา
ภาษาไทย-ภาษาอังกฤษ โดยใช้คอร์ปัสแบบขนาน

THAI-ENGLISH QUERY TRANSLATION
IN CROSS-LANGUAGE INFORMATION RETRIEVAL
USING PARALLEL CORPUS



ณัฐนันท์ ศรีสะอาด
NATHANAN SRISA-ARD

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

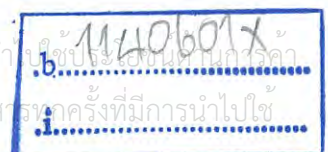
พ.ศ. 2547

เลขหมู่.....

เลขทะเบียน..... 51621

วัน,เดือน,ปี..... 2.6 ก.ค. 2547

ISBN 974-9709-77-2



THAI-ENGLISH QUERY TRANSLATION
IN CROSS-LANGUAGE INFORMATION RETRIEVAL
USING PARALLEL CORPUS



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
KING MONGKUTS INSTITUTE OF TECHNOLOGY LADKRABANG

2004

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้เฉพาะที่ออกจากร้านเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ISBN 974-9709-77-2



COPYRIGHT 2004

SCHOOL OF GRADUATE STUDIES

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับบุคลากรในแผนกที่เอกสารนี้ถูกส่งมอบเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์

การแปลคำสั่งสอบถามในการค้นคืนข้ามภาษา
ภาษาไทย-ภาษาอังกฤษ โดยใช้คอร์ปัสแบบ
ขนาน

นักศึกษา

นายณัฐนันท์ ศรีสะอาด

รหัสประจำตัว

42067075

ปริญญา

วิทยาศาสตร์มหาบัณฑิต

สาขาวิชา

เทคโนโลยีสารสนเทศ

พ.ศ.

2547

อาจารย์ผู้ควบคุมวิทยานิพนธ์

ผศ.ดร.โชติพัชร์ ภรณ์วลัย

บทคัดย่อ

การค้นคืนสารสนเทศข้ามภาษา (Cross-Language Information Retrieval : CLIR) คือวิธีการที่ทำให้สามารถใช้คำสั่งสอบถาม(Query) ภาษาเดียวในการค้นคืนเอกสารจากภาษาอื่น เราทำการศึกษาการค้นคืนสารสนเทศข้ามภาษา ทั้งจากภาษาไทยไปภาษาอังกฤษ และจากภาษาอังกฤษไปภาษาไทย ด้วยวิธีแปลคำสั่งสอบถาม(Query Translation) ซึ่งสามารถทำได้โดย การแปลด้วยพจนานุกรม การใช้คลังข้อมูลภาษาแบบขนาน (Parallel Corpus) ร่วมกับวิธี Pseudo-Relevance Feedback (PRF) และการแปลด้วยระบบแปลภาษา ผลการทดลองพบว่าการแปลคำสั่งสอบถามโดยใช้คลังข้อมูลภาษาแบบขนานร่วมกับวิธี PRF ให้ประสิทธิภาพการค้นคืนที่ดีที่สุด นอกจากนี้เรายังได้ชุดข้อมูลทดสอบ (Test Collection) สำหรับการศึกษาค้นคืนสารสนเทศภาษาไทยอีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Thai-English Query Translation in Cross-Language Information Retrieval Using Parallel Corpus
Student	Mr. Nathanan Srisa-ard
Student ID.	42067075
Degree	Master of Science
Programme	Information Technology
Year	2004
Thesis Advisor	Asst. Prof. Chotipat Pornavalai

ABSTRACT

Cross-Language Information Retrieval (CLIR) is the approach to retrieve document from other language using one language query. We study CLIR using Thai to English and English to Thai query translation. The queries are translated using translation resources bilingual dictionary, parallel corpus and machine translation. The Pseudo-Relevance Feedback (PRF) method for parallel corpus performs best in query translation. We also introduce the new test collection for Thai Information retrieval.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลงได้ด้วยข้อมูลอันมีค่าจากแหล่งข้อมูลได้แก่ ฐานข้อมูลวิทยานิพนธ์ไทย ข้อมูลจากพจนานุกรมของ สอ เสถบุตร ผู้ล่วงลับ พจนานุกรมของ อ. เขียวชัย เขี่ยมวรเมธ และพจนานุกรมเลขচিতรอนของฝ่ายวิจัยและพัฒนาสาขาสารสนเทศ เนคเทค

ขอขอบพระคุณ ผศ.ดร.โชติพัทธ์ ภรณ์วลัย ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์ ผู้เป็นแรงผลักดันให้มีความคืบหน้าตลอดระยะเวลาหลายปี ขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณบริษัทเออาร์ อินฟอร์เมชันแอนด์พับลิเคชันผู้ให้ใช้เครื่องมือช่วยสำหรับการทดลอง ขอขอบพระคุณบริษัทเอเน็ต อินเทอร์เน็ต และ บริษัทชลบุรีอินเทอร์เน็ต จำกัด ผู้ให้ใช้แบนด์วิดท์เครือข่าย

ขอขอบคุณ คุณสิริมนต์ เสตะพงษ์สัตย์ และคุณรัชดา เสตะพงษ์สัตย์ ผู้ที่ช่วยแปลงหนังสือให้เป็นไฟล์เอกสาร

ขอขอบคุณ คุณสุดคณิง ศรีสะอาด ภรรยาของข้าพเจ้า ที่ช่วยตรวจทานเอกสารและให้กำลังใจข้าพเจ้าตลอดมา

ณัฐนันท์ ศรีสะอาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป	VIII
บทที่ 1 บทนำ	1
1.1. ความเป็นมาและความสำคัญของปัญหา	1
1.2. ความมุ่งหมายและวัตถุประสงค์ของการศึกษา	3
1.3. สมมุติฐานของการศึกษา	3
1.4. คำจำกัดความ	3
1.5. ทฤษฎีหรือแนวคิดที่ใช้ในการศึกษา	4
1.6 ขอบเขตของการศึกษา	5
1.7 งานวิจัยที่เกี่ยวข้อง	6
1.8 ขั้นตอนการศึกษา	6
บทที่ 2 ขั้นตอนการเตรียมข้อมูล	8
2.1 ทรัพยากรทางภาษาสำหรับการศึกษา CLIR	8
2.2 ทรัพยากรทางภาษาภาษาไทย-ภาษาอังกฤษ	10
2.3 การจัดเตรียมพจนานุกรม	12
2.4 การจัดเตรียมฐานข้อมูลข้อความแบบขนาน (Parallel Corpus)	14
2.5 การจัดเตรียมชุดข้อมูลทดสอบ (Test Collection)	16
บทที่ 3 ระบบค้นคืนสารสนเทศสำหรับการค้นคืนข้ามภาษา	22
3.1 ระบบค้นคืนสารสนเทศ	22
3.2 Vector Space Model	23
3.2 ระบบค้นคืนสารสนเทศสำหรับการศึกษา CLIR ภาษาไทย-อังกฤษ	26
3.3 ทดสอบประสิทธิภาพระบบค้นคืนสารสนเทศ	27
3.3 สรุปการทดสอบประสิทธิภาพระบบค้นคืนสารสนเทศ	29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4 การแปลคำสั่งสอบถาม.....	30
4.1. การแปลคำสั่งสอบถามโดยใช้คอร์ปัสแบบขนาน.....	30
4.2. การแปลคำสั่งสอบถาม ด้วยพจนานุกรม.....	34
4.3. การแปลคำสั่งสอบถาม ด้วยระบบแปลภาษา.....	37
4.4. การทดลองวัดประสิทธิภาพระบบค้นคืนข้ามภาษา.....	38
บทที่ 5 การทดลองวัดประสิทธิภาพระบบค้นคืนข้ามภาษา.....	40
5.1. การทดลองแปลคำสั่งสอบถาม.....	40
5.2. ทดสอบการค้นคืนสารสนเทศ (IR Evaluation).....	41
5.3. ผลการทดลอง.....	42
5.4. การทดลองวัดค่าความแม่นยำเมื่อค่า Recall คงที่.....	75
บทที่ 6 สรุปและอภิปรายผลการทดลอง.....	76
เอกสารอ้างอิง.....	80
ภาคผนวก.....	82
ประวัติผู้เขียน.....	107

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ทรรศนคติทางภาษาประเภทฐานความรู้	11
2.2 ทรรศนคติทางภาษาประเภทคลังข้อมูลทางภาษา	12
2.3 ชื่อเรื่องวิทยานิพนธ์ที่นำมาใช้เป็น Query จำนวน 10 เรื่อง	20
2.4 หมายเลขเอกสารที่ Relevance กับ Query	21
3.1 เปรียบเทียบการสร้างตรรกะจากเอกสารภาษาไทยและภาษาอังกฤษ	26
3.2 ค่าความแม่นยำของการค้นคืนด้วยคำสั่งสอบถามภาษาไทย	28
3.3 ค่าความแม่นยำของการค้นคืนด้วยคำสั่งสอบถามภาษาอังกฤษ	28
5.1 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ	42
5.2 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย	43
5.3 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query1	45
5.4 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query1	46
5.5 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query2	48
5.6 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query2	49
5.7 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query3	52
5.8 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query3	53
5.9 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query4	55
5.10 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query4	56
5.11 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query5	58
5.12 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query5	59
5.13 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query6	61
5.14 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query6	62
5.15 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query7	64
5.16 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query7	65
5.17 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query8	67
5.18 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query8	68
5.19 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query9	70
5.20 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query9	71
5.21 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query10	73
5.22 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query10	74

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง(ต่อ)

ตารางที่	หน้า
5.23 ค่าความแม่นยำเมื่อค่า Recall คงที่ของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ วิธี PRF โดยมีความยาว Query 1.2 เท่า.....	75
5.24 ค่าความแม่นยำเมื่อค่า Recall คงที่ของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย วิธี PRF โดยมีความยาว Query 1.4 เท่า.....	75
6.1 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ.....	76
6.2 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย.....	77



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
1.1 สัดส่วนของภาษาพูดของผู้ใช้และภาษาที่ใช้ในอินเทอร์เน็ต.....	1
1.2 ตัวอย่างการค้นหาด้วย Altavista.com ซึ่งเป็น Multilingual IR.....	2
1.3 แนวทางการศึกษาเรื่องการค้นคืนข้ามภาษา.....	5
1.4 การค้นคืนสารสนเทศข้ามภาษาโดยการแปลคำสั่งสอบถาม.....	6
2.1 ชนิดของคลังข้อมูลทางภาษา.....	9
2.2 ตัวอย่างคลังข้อมูลภาษาแบบประโยคตรงกัน.....	9
2.3. ตัวอย่างคลังข้อมูลภาษาแบบขนานที่มีการกำกับ.....	10
2.4 พจนานุกรมที่แปลงให้อยู่ในรูปตาราง.....	13
2.5 ตัวอย่างบทคัดย่อจากฐานข้อมูลวิทยานิพนธ์ไทย.....	15
2.6 การแสดงค่าความแม่นยำที่การเรียกคืนระดับต่างๆ.....	16
2.7 ตัวอย่างแสดงการสร้างชุดข้อมูลทดสอบ.....	19
3.1 CLIR Block Diagram.....	22
3.2 แสดงตัวอย่างผลการค้นของ MLIR จาก Altavista.com.....	23
3.3 เอกสาร Doc1, Doc2, Doc3, Doc4.....	23
3.4 ดรรชนีย้อนกลับ (Inverted Index).....	24
3.5 แสดงเวกเตอร์ของเอกสาร Doc1, Doc2, Doc3 และ Doc4.....	25
3.6 แสดงเปรียบเทียบความเหมือนด้วย Cosine Similarity.....	25
3.6 การสร้าง Inverted Index.....	27
3.7 การค้นคืนจาก Inverted Index.....	27
3.8 กราฟความแม่นยำต่อการเรียกคืนของการค้นคืนภาษาเดียว.....	29
4.3 ขั้นตอนการแปลคำสั่งสอบถาม ด้วย PRF.....	31
4.4 เอกสารจำนวน 10 เอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม.....	32
4.5 คำที่มีค่าน้ำหนักสูงสุด จากตัวอย่างการแปลคำสั่งสอบถาม.....	33
4.7 ตัวอย่างการแปลศัพท์คำเดียว.....	35
4.8 ตัวอย่างการแปลศัพท์คำเดียว.....	35
4.9 ตัวอย่างการศัพท์ที่ไม่พบในพจนานุกรม ด้วยการตัดที่ตัดส่วนเติมท้าย (Suffix).....	36
4.10 ตัวอย่างการแปล คำต่อคำ (Word by word) โดยละทิ้งคำหยุด.....	36
4.11 ตัวอย่างการแปลคำสั่งสอบถาม ด้วย "ภาชิต".....	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่	หน้า
5.1 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ.....	42
5.2 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย.....	43
5.3 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query1	45
5.4 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query1	46
5.5 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query2	48
5.6 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query2	49
5.7 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query3	52
5.8 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query3	53
5.9 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query4	55
5.10 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query4	56
5.11 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query5	58
5.12 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query5	59
5.13 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query6	61
5.14 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query6	62
5.15 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query7	64
5.16 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query7	65
5.17 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query8	67
5.18 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query8	68
5.19 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query9	70
5.20 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query9	71
5.21 ค่าความแม่นยำของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ Query10	73
5.22 ค่าความแม่นยำของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย Query10	74
6.1 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ.....	76
6.2 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย.....	77

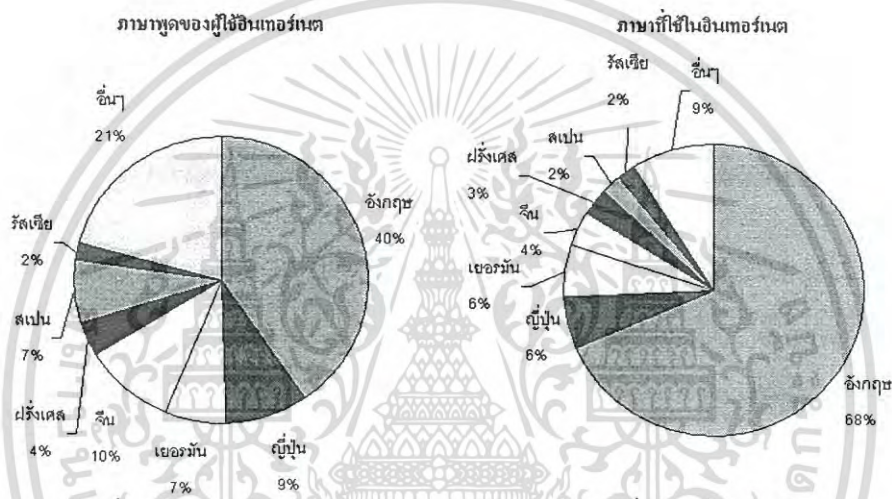
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

เนื่องมาจากเทคโนโลยีการสื่อสารข้อมูลที่พัฒนาขึ้น ทำให้การเผยแพร่เอกสารทำได้โดยสะดวก ดังจะเห็นได้จากจำนวนเอกสารบนเครือข่ายอินเทอร์เน็ตที่มีจำนวนมหาศาลและมีหลายภาษา จากผลการวิจัยของโกลบอลรีทซ์[1] ในปี 2546 รูป 1.1 เอกสารในเว็บเพจส่วนใหญ่ คือ ประมาณร้อยละ 68 ใช้ภาษาอังกฤษ และมีภาษาไทยไม่ถึงร้อยละ 1



รูปที่ 1.1 สัดส่วนของภาษาพูดของผู้ใช้และภาษาที่ใช้ในอินเทอร์เน็ต

นักวิจัยทางการค้นคืนสารสนเทศ (Information Retrieval) กำลังให้ความสนใจกับการค้นคืนสารสนเทศข้ามภาษา (Cross-Language Information Retrieval หรือ CLIR) ซึ่งหมายถึงการค้นคืนโดยใช้คำสั่งสอบถาม (Query) ในภาษาหนึ่งแล้วได้ผลการค้นคืนเป็นเอกสารภาษาอื่นๆ[2] จากนั้นจึงแปลเอกสารด้วยระบบแปลภาษา หรือแปลด้วยคน

1.1.1. การค้นหาข้อมูลโดยผู้ใช้ที่อ่านเขียนได้ภาษาเดียว

แม้ว่าการพัฒนาระบบแปลภาษาจะสามารถทำให้ผู้ใช้ที่ไม่ชำนาญภาษาต่างประเทศเข้าถึงข้อมูลในภาษาต่างประเทศสะดวกขึ้น แต่ระบบแปลภาษาต้องใช้เวลาการประมวลผลอย่างมาก จึงไม่มีทางเป็นไปได้ที่จะแปลเอกสารทั้งหมดที่มีอยู่ในเว็บรวมทั้งเอกสารที่กำลังเกิดขึ้นใหม่ได้ทัน ดังนั้นจึงเหมาะสมกว่าถ้าหากคัดเลือกเอกสารจำนวนหนึ่งโดยวิธีการค้นคืนสารสนเทศ จากนั้นจึงนำเอกสารที่ค้นได้เข้าสู่กระบวนการแปลภาษา ขอยกตัวอย่างให้เห็นชัดเจนเช่น จำนวนเว็บเพจที่ถูกทำดรรชนีโดย Google.com เมื่อเดือนมกราคมปี 2547 มีจำนวนประมาณ 4,200 ล้านหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การค้นคืนให้ได้เอกสารที่ต้องการจำนวนเพียง 10 หน้า แล้วจึงนำเข้าสู่กระบวนการแปลภาษา ย่อมเป็นการอำนวยความสะดวกแก่ผู้ใช้เป็นอย่างมาก

1.1.2. การค้นหาข้อมูลโดยผู้ใช้ที่สามารถอ่านเขียนได้หลายภาษา

การสร้างคำสั่งสอบถามในภาษาที่ผู้ใช้ไม่ถนัดเป็นอุปสรรคสำคัญอย่างหนึ่ง ตัวอย่างเช่น การค้นคืนเอกสารภาษาอังกฤษโดยผู้ใช้ชาวไทย ซึ่งอาจจะมีความรู้ในภาษาอังกฤษในระดับกลาง คุณภาพของคำสั่งสอบถามภาษาอังกฤษที่ป้อนให้กลไกค้นหาข้อมูล อาจจะไม่ดีพอ ทำให้ไม่ได้เอกสารที่ตรงกับที่ต้องการ

การสร้างคำสั่งสอบถาม อาจจะต้องใช้ความรู้ในศาสตร์ที่เกี่ยวข้อง นอกเหนือจากความรู้ เรื่องความหมายศัพท์ ตัวอย่างคำสั่งสอบถามที่ซับซ้อนเช่น

“ต้องการทราบ ความสัมพันธ์ระหว่างความรู้ความเข้าใจ ทักษะการคิดคำนวณ และทักษะการแก้ ใจทย์ปัญหาในวิชาคณิตศาสตร์ ในระดับชั้น ประถมศึกษา”

ย่อมต้องใช้ความรู้ในสาขาวิชาที่เกี่ยวข้องนอกเหนือไปจากความรู้เรื่องภาษา

1.1.3. แนวโน้มการใช้เอกสารจากภาษาที่ไม่ใช่ภาษาอังกฤษอื่นกำลังเพิ่มขึ้น

ดังจะเห็นได้จากบริการค้นคืนสารสนเทศสำหรับเครือข่ายอินเทอร์เน็ต (Search Engine) มีการปรับปรุงให้มีความสามารถในการค้นหาเอกสารหลายภาษาได้พร้อมกัน (Multilingual) และ มีความสามารถในการแปลภาษา หากว่าเอกสารที่ค้นหาได้ไม่ใช่ภาษาอังกฤษ ดังตัวอย่างในรูปที่ 1.2 เป็นการค้นหาคำว่า “Japanese in Thailand” จากเว็บ Altavista.com โดยมีการทำการเชื่อมโยง (Link) ไประบบแปลภาษาเมื่อพบเว็บเพจภาษาญี่ปุ่น

AltaVista found 448,241 results [About](#)

[Thailand.com](#) are the leaders for Thai Exports!

[Thailand.com](#) has the most comprehensive, detailed ... quality products from [Thailand](#) a leading in rubber and plastic footwear Rubber and plastic footwear [Japanese](#) snack food,Norimaki,Rice ...

[www.thailand.com](#) • Refreshed in past 48 hours • [Related Pages](#)

[More pages from www.thailand.com](#)

[タイ国日本人会は](#)

... 又は皆様にご紹介したい情報は下記にどうぞ

[Japanese Association in Thailand](#)

1st Fl. Sathorn

Thani Bldg. II, 92/2 North Sathorn Rd. Bangrak Bangkok 10500 [Thailand](#) Tel:0-2236-1201(代 ...

[www.jat.or.th](#) • Refreshed in past 48 hours • [Related Pages](#) • [Translate](#)

[More pages from www.jat.or.th](#)

[Thailand: Phuket hotels, diving, maps and travel guide](#)

[Thailand: Phuket hotels, diving, maps and travel information guides and Asian Newspapers ... Le Guide en Francais Hotel guide in \[Japanese\]\(#\) Travel and Tours \(in \[Japanese\]\(#\)\) The yellow pages of ...](#)

[www.thaiwave.com](#) • Refreshed in past 24 hours • [Related Pages](#)

[More pages from www.thaiwave.com](#)

รูปที่ 1.2 ตัวอย่างการค้นหาด้วย Altavista.com ซึ่งเป็น Multilingual IR

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถึงแม้ปัจจุบัน ยังไม่มีบริการค้นคืนสารสนเทศสำหรับเครือข่ายอินเทอร์เน็ตที่ให้บริการค้นคืนข้ามภาษา อาจเนื่องมาจากผู้ใช้บริการส่วนใหญ่สามารถใช้ภาษาอังกฤษได้ ซึ่งในประเทศที่ไม่ได้ใช้ภาษาอังกฤษเป็นภาษาหลักอย่างประเทศไทย บริการค้นคืนข้ามภาษา น่าจะเป็นประโยชน์กับผู้ใช้จำนวนมาก

1.2. ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

วัตถุประสงค์ของการศึกษานี้คือ

- 1.2.1. ศึกษาวิธีการค้นคืนเอกสารภาษาอังกฤษโดยใช้คำสั่งสอบถามภาษาไทย เพื่อให้คนไทยเข้าถึงข้อมูลในภาษาอื่นได้มากขึ้น
- 1.2.2. ศึกษาวิธีการค้นคืนเอกสารภาษาไทยโดยใช้คำสั่งสอบถามภาษาอังกฤษ เพื่อเปิดโอกาสในการเผยแพร่ข้อมูลภาษาไทยแก่ชาวต่างชาติ

1.3. สมมุติฐานของการศึกษา

สมมุติฐานของการศึกษาคือ

- 1.3.1. สามารถทำการค้นคืนสารสนเทศข้ามภาษาระหว่างภาษาไทยและภาษาอังกฤษได้โดยใช้ทรัพยากรทางภาษาที่มีอยู่ในปัจจุบัน ได้แก่พจนานุกรม ระบบแปลภาษา หรือทรัพยากรทางภาษาอื่นๆ ที่สามารถใช้ในการแปลคำสั่งสอบถามได้
- 1.3.2. สามารถทำการการค้นคืนข้ามภาษาจากภาษาไทยไปภาษาอังกฤษ และกลับกันคือ จากภาษาอังกฤษไปภาษาไทย โดยหลักการเดียวกัน

1.4. คำจำกัดความ (DEFINITION)

เนื่องจากศัพท์ที่ใช้สำหรับการศึกษาค้นคืนสารสนเทศ ในทางปฏิบัติเรานิยมเรียกทับศัพท์ สาเหตุหนึ่งนอกจากความคุ้นเคย คือไม่ได้มีการบัญญัติศัพท์ไว้ สำหรับการเรียบเรียงบทความภาษาไทยในวิทยานิพนธ์นี้ ศัพท์ใดที่ใช้บ่อย เราได้ค้นหาคำแปลภาษาไทยจากเอกสารวิชาการ โดยเราได้นำมาอธิบายความหมายโดยละเอียดไว้ตรงนี้ เพื่อความเข้าใจที่ตรงกันตลอด

- 1.4.1. การค้นคืนสารสนเทศ (Information Retrieval) คือการดึงสารสนเทศที่ต้องการออกมาจากสารสนเทศทั้งหมด ซึ่งมีปริมาณมากเกินกว่าที่จะเปิดดูที่ละอัน (Browse) ในบางครั้งเราใช้คำว่าสืบค้นสารสนเทศแทนกันได้ เช่นการสืบค้นหนังสือจากห้องสมุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1.4.2. การค้นหา (Search) ผู้ใช้ทั่วไปใช้คำว่า "ค้นหา" แทนความหมายทั้ง Search หรือ Retrieval ส่วนในงานวิจัยเช่นวิทยานิพนธ์ภาษาไทยที่กล่าวถึงการค้นคืนสารสนเทศมากกว่าครึ่งหนึ่ง ใช้คำว่าค้นคืน เมื่อต้องการหมายถึง Retrieval
- 1.4.3. ทรัพยากรทางภาษา (Linguistics resource) คือข้อมูลที่เป็นประโยชน์สำหรับการศึกษารื่องภาษา เช่นพจนานุกรม
- 1.4.4. คลังข้อมูลภาษา (Corpus) คือข้อความ (Text) ในรูปแบบอิเล็กทรอนิกส์ อาจจะมีการกำกับ (Mark Up) หรือไม่ก็ได้ จัดเป็นทรัพยากรทางภาษาที่สำคัญอย่างหนึ่งเนื่องจากให้ข้อมูลทางภาษาแบบที่ใช้จริง
- 1.4.5. ชุดเอกสาร (Collection) หรือบางที่เรียกว่า ชุดข้อมูล หมายถึงกลุ่มของเอกสารการจัดกลุ่มนี้เพื่อประโยชน์อย่างใดอย่างหนึ่ง เช่นชุดข้อมูลสำหรับทดสอบประสิทธิภาพ (Test Collection)
- 1.4.6. การค้นคืนการค้นคืนสารสนเทศข้ามภาษา (Cross-Language Information Retrieval หรือ CLIR) คำนี้ถูกนำมาใช้เมื่อการประชุมวิชาการ TREC (Text Retrieval Conference) ครั้งที่ 2 [3] โดยก่อนหน้านี้มีการใช้คำว่า Translingual Information Retrieval แทนความหมายเดียวกัน เนื่องจากศัพท์คำนี้มีความยาวมาก เราจึงขอเขียนย่อ เป็น "การค้นคืนข้ามภาษา" ตลอดทั้งวิทยานิพนธ์
- 1.4.7. การค้นคืนการค้นคืนสารสนเทศหลายภาษา (Multilingual Information Retrieval) หรือ MLIR คือการค้นคืนจากเอกสารหลายภาษา
- 1.4.8. คำสั่งสอบถาม (Query) คือคำประโยคหรือวลีที่ผู้ใช้ป้อนเข้าสู่ระบบค้นคืนสารสนเทศเพื่อแสดงว่าต้องการเอกสารใด

1.5. ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

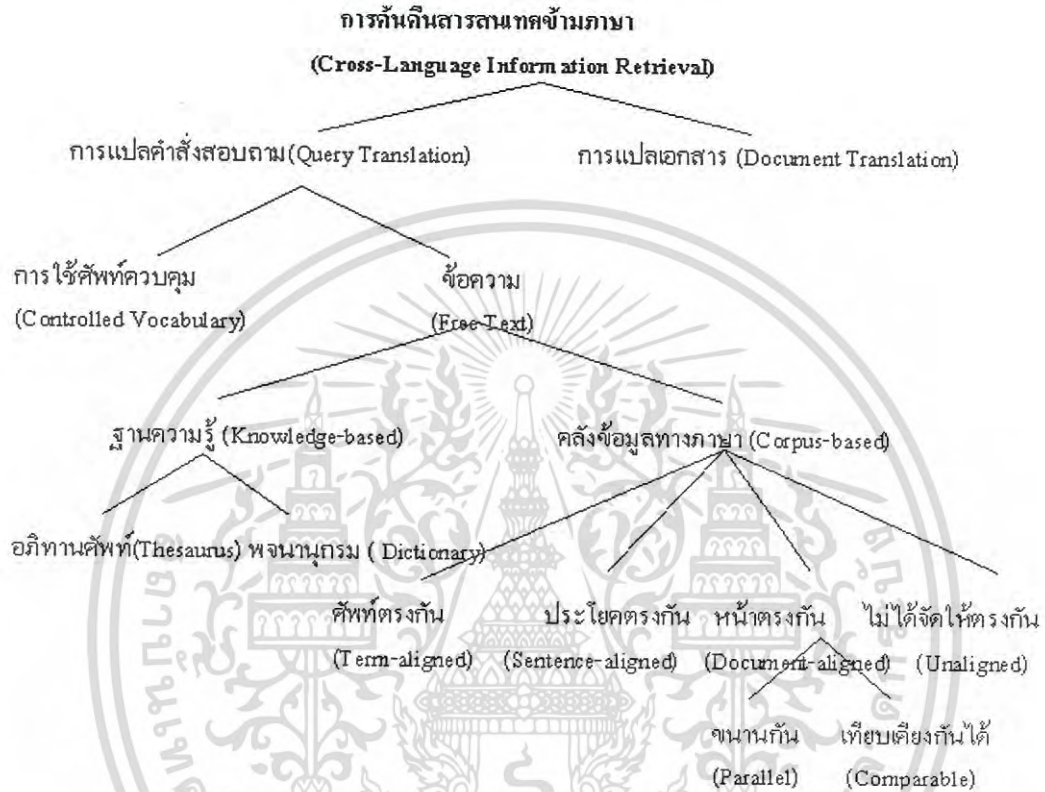
เป้าหมายของการค้นคืนข้ามภาษา คือการที่ให้ผู้ใช้อป้อนคำสั่งสอบถามในภาษาหนึ่งแต่ให้ผลการค้นคืนจากเอกสารในภาษาอื่นๆ วิธีการที่ได้นำมาใช้ ได้แก่

- 1.5.1 การแปลเอกสาร (Document Translation) สามารถทำได้โดยการใช้ระบบแปลภาษา ซึ่งมีขีดจำกัดที่ต้องใช้การประมวลผลสูง ซึ่งสามารถทำได้กับชุดเอกสาร (Collection)ขนาดเล็ก
- 1.5.2 การแปลคำสั่งสอบถาม (Query Translation) สามารถทำได้โดยการใช้ทรัพยากรทางภาษาศาสตร์ ทั้งที่เป็นฐานความรู้และที่เป็นคลังข้อมูลทางภาษา
- 1.5.3 การแปลทั้งคำสั่งสอบถามและเอกสาร ตัวอย่างเช่นการแปลให้เป็นภาษากลาง (Intermediary หรือ Interlingual Representation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการศึกษาของ Ari Pirkola [4] พบว่าการแปลเฉพาะคำสั่งสอบถาม มีประสิทธิภาพของการค้นคืน ได้เท่าเทียมกับการแปลทั้งเอกสาร

นอกจากแบ่งตามวิธีการแปลแล้วยังอาจแบ่งตามทรัพยากรทางภาษาที่ใช้ในการแปล ดังแสดงในรูปที่ 1.3 [5]



รูปที่ 1.3 แนวทางการศึกษาเรื่องการค้นคืนข้ามภาษา

1.6. ขอบเขตของการศึกษา

- 1.6.1. ทรัพยากรทางภาษาบางอย่างที่จำเป็นต้องใช้ แต่ยังไม่มีการจัดทำขึ้น เราอาจจำเป็นต้องการจัดทำขึ้นมาเองเช่นชุดข้อมูลทดสอบ
- 1.6.2. การจัดทำหรือรวบรวมข้อมูลที่จำเป็นต้องใช้แรงคน อาจได้ปริมาณหรือคุณภาพไม่ดีพอเนื่องจากขีดจำกัดเรื่องแรงคน อันอาจเป็นปัญหาที่มีผลให้มีการเปลี่ยนวิธีการทดลองหรือการจัดเตรียมข้อมูล ให้มีการใช้แรงงานคนน้อยลงแม้ว่าจะกระทบกับคุณภาพของข้อมูลในบางส่วน
- 1.6.3. ทรัพยากรทางภาษาที่มีลิขสิทธิ์คุ้มครอง เช่นพจนานุกรม ไม่สามารถเผยแพร่ข้อมูลได้ แต่จะแสดงวิธีการจัดเตรียมหรือการแปลงรูปแบบ ให้สะดวกแก่การใช้งานด้วยคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7. งานวิจัยเกี่ยวข้อง

การศึกษาเรื่องการค้นคืนสารสนเทศข้ามภาษา ภาษาไทย-อังกฤษ ก่อนหน้านี้มีเพียงสองกลุ่มได้แก่ ชูลีรัตน์และคณะ 2543 [6], ทำการแปลคำสั่งสอบถามด้วยพจนานุกรม ซึ่งรวบรวมจากพจนานุกรมที่มีอยู่แล้วซึ่งให้ประสิทธิภาพการค้นคืนที่ไม่ดีนักเนื่องจากพจนานุกรมมีศัพท์จำนวนน้อยไม่มีความครอบคลุมเพียงพอ และยังพบปัญหาคำทับศัพท์ และคำพ้องรูป ในพจนานุกรม

จากการศึกษาของ สมชาย และคณะ 2541, 2542, 2543 [7] ทำการค้นคืนข้ามภาษาคำทับศัพท์ (Transliterated Word) โดยเทียบรหัสเสียง จากรายชื่อบุคคล ทั้งจากภาษาไทยเขียนทับศัพท์ภาษาอังกฤษ และภาษาอังกฤษ เขียนทับศัพท์ภาษาไทย ได้ผลเป็นที่น่าพอใจเมื่อทดสอบกับรายชื่อคน แต่ว่าการนำไปใช้งานจริงสำหรับข้อมูลที่เป็นข้อความ (Text) หรือไม่เป็นโครงสร้าง อาจทำได้ยากเนื่องจากคำทับศัพท์ที่แทรกอยู่ในข้อความ มักจะถูกกระบวนการตัดคำ (Word segmentation) ทำให้กลายเป็นคำขนาดเล็ก

1.8. ขั้นตอนการศึกษา

ขั้นตอนการค้นคืนข้ามภาษาด้วยการแปลคำสั่งสอบถามดังแสดงในรูปที่ 1.4 คำสั่งสอบถามในภาษาต้นทาง (Source Language Query) จะถูกแปลเป็นคำสั่งสอบถามในภาษาเป้าหมาย โดยการใช้ทรัพยากรทางภาษาร่วมกับวิธีการแปลที่เหมาะสมสำหรับทรัพยากรทางภาษาชนิดนั้น ก่อนจะนำเข้าสู่ระบบค้นคืนสารสนเทศของภาษาเป้าหมาย



รูปที่ 1.4 การค้นคืนสารสนเทศข้ามภาษาโดยการแปลคำสั่งสอบถาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการศึกษาจะทำพร้อมกันทั้งการศึกษาการค้นคืนข้ามภาษาจากภาษาไทยเป็นภาษาอังกฤษ และจากภาษาอังกฤษเป็นภาษาไทย และเนื่องจากข้อจำกัดของทรัพยากรทางภาษา เราจึงศึกษาวิธีการแปลคำสั่งสอบถาม ภายหลังจากที่รวบรวมและจัดเตรียมทรัพยากรทางภาษา เพื่อเลือกใช้วิธีการแปลคำสั่งสอบถามที่เหมาะสมกับทรัพยากรทางภาษาที่มี ลำดับขั้นตอนการศึกษามีดังนี้

- 1.8.1. เตรียมทรัพยากรทางภาษาต่างๆ ที่ใช้ในการแปลคำสั่งสอบถาม ได้แก่ พจนานุกรม (Dictionary) ฐานข้อมูลข้อความ (Corpus) ให้ได้คุณภาพดีที่สุดเท่าที่จะทำได้
- 1.8.2. เตรียมข้อมูลสำหรับทดสอบประสิทธิภาพ (Test Collection) โดยอาจทำขึ้นเอง ถ้ายังไม่มีมาก่อน โดยต้องระวังเรื่องความเป็นกลาง ไม่เอนเอียงของคำตอบ
- 1.8.3. เตรียมระบบค้นคืนสารสนเทศ ที่สามารถใช้งานได้ทั้งภาษาไทยและภาษาอังกฤษ
- 1.8.4. ทดสอบเปรียบเทียบประสิทธิภาพของวิธีการแปลแบบต่างๆ ที่สามารถนำมาใช้ กับเตรียมทรัพยากรทางภาษาที่เตรียมได้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ขั้นตอนการเตรียมข้อมูล

บทนี้จะกล่าวถึงการจัดเตรียมข้อมูลที่ต้องใช้ในการศึกษาการค้นคืนข้ามภาษาโดยการแปลคำสั่งสอบถาม ข้อมูลที่ต้องใช้ได้แก่ทรัพยากรทางภาษาสำหรับการแปลคำสั่งสอบถามและชุดข้อมูลทดสอบ เราได้ทำการสำรวจเสาะหาทรัพยากรทางภาษาที่มีคุณภาพดีจากที่มีอยู่แล้วทั้งที่เป็นภาษาไทยและภาษาอังกฤษ แล้วนำมาแปลงรูปแบบให้เหมาะแก่การใช้งาน

การได้มาของทรัพยากรทางภาษานำไปสู่การศึกษาวิธีการแปลคำสั่งสอบถาม เมื่อมีวิธีการแปลหลายวิธี ชุดข้อมูลทดสอบจึงเป็นเครื่องมือสำคัญในการทดสอบประสิทธิภาพการแปลคำสั่งสอบถามแบบในการทดสอบเชิงเปรียบเทียบ ซึ่งชุดข้อมูลทดสอบการค้นคืนข้ามภาษา (CLIR Test Collection) ภาษาไทย-ภาษาอังกฤษ ยังไม่มีการจัดทำเผยแพร่เป็นมาตรฐานเหมือนกับภาษาอื่น (เช่น ภาษาญี่ปุ่น ภาษาจีน ที่จัดทำโดย NII [8] หรือภาษาทางยุโรป ที่จัดทำโดย CLEF [9]) เราจึงได้นำเสนอวิธีการสร้างชุดข้อมูลทดสอบระบบค้นคืนสารสนเทศข้ามภาษาที่มีคุณภาพดีโดยใช้ทุนการจัดสร้างต่ำ ซึ่งสามารถการทดสอบการค้นคืนสารสนเทศข้ามภาษาทั้งจากภาษาไทยเป็นภาษาอังกฤษและจากภาษาอังกฤษเป็นภาษาไทยได้อีกทั้งยังสามารถใช้เป็นชุดข้อมูลทดสอบระบบค้นคืนสารสนเทศภาษาไทยหรือภาษาอังกฤษภาษาเดียวได้อีกด้วย

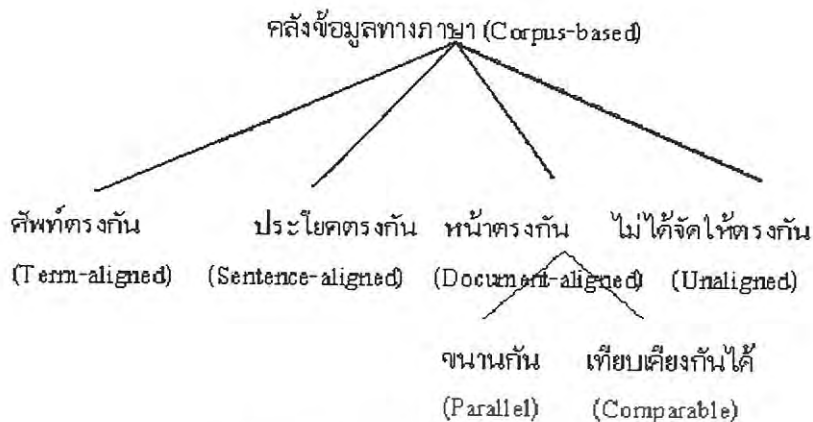
2.1. ทรัพยากรทางภาษาสำหรับการแปลคำสั่งสอบถาม

ทรัพยากรทางภาษาสำหรับการแปลคำสั่งสอบถามหลายชนิด โดยอาจจัดแบ่งทรัพยากรทางภาษาเป็น 3 ประเภทคือ

2.1.1. ฐานความรู้ (Knowledge-based) เช่นพจนานุกรมสองภาษา (Bilingual Dictionary) มีคำแปลจากภาษาหนึ่งให้ความหมายในอีกภาษาหนึ่ง เช่นพจนานุกรม อังกฤษ-ไทย คือพจนานุกรมที่แปลจากศัพท์ภาษาอังกฤษเป็นความหมายภาษาไทย ตัวอย่างพจนานุกรมสองภาษา ได้แก่พจนานุกรมที่พิมพ์จำหน่ายเป็นเล่ม เช่นพจนานุกรม ไทย-อังกฤษของ สอ เสถบุตร

2.1.2. คลังข้อมูลทางภาษา (Corpus-based) คือเอกสารรูปแบบข้อความที่มาจากการรวบรวมที่มีอยู่แล้วหรือแปลจากต้นฉบับ ชนิดของคลังข้อมูลทางภาษาอาจแบ่งได้เป็นหลายชนิด ดังแสดงในรูปที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 ชนิดของคลังข้อมูลทางภาษา

คลังข้อมูลภาษาแบบขนาน (Parallel Corpus) บางชิ้นอาจเป็นคลังข้อมูลภาษาแบบประโยคตรงกัน (Sentence Aligned) อีกด้วยเช่น คลังข้อมูลภาษาแบบขนาน เรื่องจันดารา[10] ในรูปที่ 2.2

<p>พอน้ำ แข็งละลาย ไป ได้สี่ก้อน แล้วผม ก็บอก "ผมขออนุญาตครับ "</p> <p>After four chunks of ice had melted, I told her, "Please allow me."</p> <p>เธอถามโดย ไม่สนใจว่า "ขออนุญาตอะไร"</p> <p>She asked idly: "Allow what?"</p> <p>แทน คำตอบ - ผมวางมือลง ยึดขอบเก้าอี้หวาย ทั้งสองข้าง แล้วแนบหน้าลงกับ เนื้อหลัง อันอวบ และเย็น ยิ่งกะหินอ่อนของเธอ</p> <p>Instead of answering, I grasped the edge of the rattan couch on both sides and buried my face against her plump back, which was as cold as marble.</p>
--

รูปที่ 2.2 ตัวอย่างคลังข้อมูลภาษาแบบประโยคตรงกัน

ในคลังข้อมูลภาษาบางชิ้นมีการกำกับ (Mark-up) ไว้ด้วยเพื่อประโยชน์ในการระบุชนิดของข้อมูล ดังตัวอย่าง คลังข้อมูลภาษาแบบขนาน ภาษาไทย-ภาษาอังกฤษ ดัดแปลงจากนิตยสารไอเดี่ย [11] ในรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

- <DOCUMENT>
  <TITLE>Idea</TITLE>
  <YEAR>1998</YEAR>
  <VOLUME>109</VOLUME>
  <PAGE>17</PAGE>
- <ENGLISH>
  <HEAD>The unkindest cut</HEAD>
  <BODY>More women in developing
  countries are having their babies in
  hospitals, but too many are undergoing
  unnecessary surgery, in most cases
  caesareans or episiotomies.</BODY>

  <KEYWORD>unkindest,undergoing,caesareans
</ENGLISH>
- <THAI>
  <HEAD>การผ่าที่สุดจะไม้อภิรมย์</HEAD>
  <BODY>มีสตรีมีมึนในประเทศกำลังพัฒนาที่คลอด
  บุตรในโรงพยาบาล แต่ก็มีจำนวนมากที่ใ้รับการผ่า
  ตัดโดยไม้อำเป็น กรมีส่วนใหญ่แล้วเป็นการผ่าท้อง
  หรือ ผ่าปากช่องคลอด</BODY>
  <KEYWORD>ไม้อภิรมย์,การผ่าตัด,การผ่าท้อง,ผ่า
  ปากช่องคลอด</KEYWORD>
</THAI>
</DOCUMENT>

```

รูปที่ 2.3. ตัวอย่างคลังข้อมูลภาษาแบบขนานที่มีการกำกับ

2.1.3. ระบบแปลภาษา (Machine Translation หรือ MT) การใช้ระบบแปลภาษาเป็นวิธีการที่สะดวกอย่างหนึ่งในการศึกษาการค้นคว้าข้ามภาษา เนื่องจากสามารถแปลได้ทั้งคำสั่งสอบถามและเอกสาร โดยไม่ต้องคำนึงถึงการทำงานที่ซับซ้อนของระบบ แต่อาจมีจุดด้อยเนื่องด้วยระบบแปลภาษายังไม่มีให้ใช้ทุกภาษา ตัวอย่างของระบบแปลภาษาที่ได้รับความนิยม เช่น SYSTRAN ซึ่งสามารถแปลได้มากกว่า 30 คู่ของภาษา [12]

2.2 การสำรวจทรัพยากรทางภาษา ภาษาไทย-ภาษาอังกฤษ

จากการสำรวจพบว่า มีทรัพยากรทางภาษาทั้งที่เป็นฐานความรู้ คลังข้อมูลทางภาษาและระบบแปลภาษาที่มีศักยภาพในการนำมาใช้แปลคำสั่งสอบถาม รูปแบบของทรัพยากรทางภาษาเหล่านี้มีทั้งที่เป็นแฟ้มข้อความ เป็นสิ่งตีพิมพ์ หรือเป็นซอฟต์แวร์ ดังแสดงในตารางที่ 2.1 และ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 ประเภทฐานความรู้ (Knowledge-based)

ตารางที่ 2.1 ทรัพยากรทางภาษาประเภทฐานความรู้

ชื่อ	แหล่งที่มา	รูปแบบ	ประเภท	จำนวนคำ	ข้อคิดเห็น
CRCL dictionary	CRCL [13]	เพิ่มข้อความ	พจนานุกรม	~3,000 คำ	- แจกจ่ายได้โดยเสรี
LEXiTRON 1.1 [14]	NECTEC	ซอฟต์แวร์	พจนานุกรม	~7,000 คำ	- แจกจ่ายได้โดยเสรี
พจนานุกรม ไทย- อังกฤษ และ อังกฤษ- ไทย [15]	สอ เสถบุตร	ซอฟต์แวร์	พจนานุกรม	~ 47,000 คำ ~ 75,000 ความหมาย	- อธิบายศัพท์และวลีได้ละเอียด - มีลิขสิทธิ์คุ้มครอง
WordNet Thesaurus [16]	Princeton University	ซอฟต์แวร์ และ ฐานข้อมูล	พจนานุกรมคำพ้อง	~ 60,000 คำ	- เป็นอภิธานศัพท์(Thesaurus)ภาษา อังกฤษ ภาษาเดียวที่ครอบคลุมความ หมายดี
พจนานุกรมคำพ้องไทย [17]	ยี่น ภู่วร- วรรณ	สิ่งพิมพ์	พจนานุกรมคำพ้อง	~ 20,000 คำ	-เป็นอภิธานศัพท์ภาษาไทย ภาษาเดียว - มีลิขสิทธิ์คุ้มครอง
พจนานุกรม ไทย-อังกฤษ และ อังกฤษ-ไทย [18]	เชียรชัย เอี่ยมวรรณ	ซอฟต์แวร์	พจนานุกรม	~ 21,500 คำ	- มีคำศัพท์ใหม่มากกว่าของ สอ เสถบุตร - มีลิขสิทธิ์คุ้มครอง

2.2.2. ประเภทคลังข้อมูลทางภาษา

ตารางที่ 2.2 ทรัพยากรทางภาษาประเภทคลังข้อมูลทางภาษา

ชื่อ	แหล่งที่มา	รูปแบบ	ประเภท	จำนวนเอกสาร	ข้อคิดเห็น
นิตยสารไอดี ฉบับปี ค.ศ.1998-1999 [11]	ข่าวแปลจาก Times Magazine	สิ่งพิมพ์	ฐานข้อมูลข้อความ	~500 หัวข้อ	ข่าวหรือบทความแปลจากสิ่งพิมพ์ต่างประเทศ
Thai-English bitexts [10]	CRCL	แฟ้ม HTML	ฐานข้อมูลข้อความ	6 ฉบับ ~3000 ประโยค	เป็นนิยายแปล จัดนานประโยคต่อประโยค
ORCHID corpus	NECTEC	แฟ้มข้อความ	ฐานข้อมูลข้อความ		ฐานข้อมูลข้อความภาษาไทยภาษาเดียว มี Mark-Up ตำแหน่ง POS
ฐานข้อมูลวิทยานิพนธ์ไทย [19]	ศูนย์บริการสารสนเทศทาง เทคโนโลยี (ศสท.)	แฟ้ม HTML	ฐานข้อมูลข้อความ	41,566 บทความ เมื่อปี 2546	ฐานข้อมูลวิทยานิพนธ์ไทย ตั้งแต่ปี พ.ศ. 2528 ถึงปัจจุบัน มีบทความภาษาไทยและภาษาอังกฤษ มีการปรับปรุงเพิ่มเติมอย่างสม่ำเสมอ

2.3. การจัดเตรียมทรัพยากรทางภาษาประเภทฐานความรู้

2.3.1. พจนานุกรมคำพ้องหรืออภิธานศัพท์ (Thesaurus) จากข้อมูลในตารางที่ 2.1 มีพจนานุกรมคำพ้อง 2 ฉบับได้แก่ WordNet Thesaurus และ พจนานุกรมคำพ้องไทย ซึ่งอาจเป็นประโยชน์ในการแปลถ้าพจนานุกรมคำพ้อง 2 ฉบับมีความเชื่อมโยงกัน การสร้างความเชื่อมโยงอาจทำได้โดยการใช้พจนานุกรมสองภาษาซึ่งยังต้องมีการศึกษาต่อไป ดังนั้นเราจึงยังไม่ใช้พจนานุกรมคำพ้องนี้

2.3.2. พจนานุกรมสองภาษา (Bi-Lingual Dictionary) จากข้อมูลในตารางที่ 2.1 มีพจนานุกรมสองภาษา ภาษาไทย-ภาษาอังกฤษ 4 ฉบับได้แก่พจนานุกรมของ CRCL, Lexitron, สอ-เสถียรบุตร และของ เอียร์ชัย เคี่ยมวรเมธ ซึ่งพจนานุกรมของ CRCL มีขนาดข้อมูลน้อยที่สุด เราจึงไม่เลือกใช้

ข้อมูลจากพจนานุกรมทุกฉบับได้ถูกแปลงให้อยู่ในรูปแบบตารางในฐานข้อมูลดังแสดงในรูปที่ 2.4 เพื่อสะดวกต่อการใช้งาน

```
mysql> select * from all_eng_thai where length(src)>6 limit 20;
```

en	th	hit	en_hit	th_hit	src
abate	บรรเทา	2	9	21	pop-lex
abdomen	ท้อง	3	2	8	pop-ss-lex
able	สามารถ	3	5	12	pop-ss-lex
abnormal	ผิดปกติ	2	3	9	pop-lex
abnormal	วิปริต	2	3	3	pop-lex
abolish	เลิกแล้ว	3	3	23	pop-ss-lex
abort	แท้ง	2	10	3	pop-lex
about	ประมาณ	2	17	15	pop-lex
about	เกือบจะ	2	17	8	pop-lex
about	เกี่ยวกับ	3	17	25	pop-ss-lex
above	ข้างบน	2	12	2	pop-lex
above	ข้างบน	2	12	7	pop-lex
abridge	คัดลอก	3	3	12	pop-ss-lex
abroad	ต่างประเทศ	3	7	4	pop-ss-lex
absorb	ดูด	3	9	13	pop-ss-lex
abstract	นามธรรม	2	9	1	pop-lex
abundant	อุดมสมบูรณ์	2	7	31	pop-lex
accent	สำเนียง	2	9	1	pop-lex
accept	ยอมรับ	2	7	14	pop-lex
accept	รับ	3	7	19	pop-ss-lex

รูปที่ 2.4 พจนานุกรมที่แปลงให้อยู่ในรูปตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4. การจัดเตรียมทรัพยากรทางภาษาประเภทคลังข้อมูลทางภาษา

จากข้อมูลในตารางที่ 2.2 คลังข้อมูลทางภาษาจากนิตยสารไอเดีย ซึ่งตีพิมพ์เนื้อหาสองภาษา เราจึงทำการ OCR ข้อความภาษาอังกฤษ และพิมพ์ข้อความภาษาไทยในเนื้อหาส่วนเดียวกันลงในแฟ้มข้อความเดียวกัน โดยให้รูปแบบ XML ใช้เวลาประมาณ 1 เดือนเพื่อแปลงนิตยสารไอเดีย ฉบับปี 1998-1999 ได้บทความจำนวนประมาณ 500 เรื่อง รวมเป็นขนาดได้ประมาณ 3 เมกกะไบต์ ซึ่งจัดได้ว่าไม่มากนัก

ต่อมาผู้วิจัยได้พบว่า ฐานข้อมูลวิทยานิพนธ์ไทย ซึ่งให้บริการที่เว็บไซต์ของศูนย์บริการสารสนเทศทางเทคโนโลยี (ศสท.) <http://www.tiac.or.th> มีบทความจำนวน 41,566 บทความ จากจำนวนประมาณ 52,000 บทความ (ขนาดประมาณ 700 เมกกะไบต์) เฉพาะที่มีทั้งภาษาไทยและภาษาอังกฤษ จัดว่าเป็นฐานข้อมูลข้อความแบบขนานที่มีจำนวนมากและมีการพิมพ์ผิดต่ำคุณภาพดี นำมาใช้เป็นคลังข้อมูลทางภาษาแบบขนาน (Parallel Corpus) ดังตัวอย่างในรูป

2.5

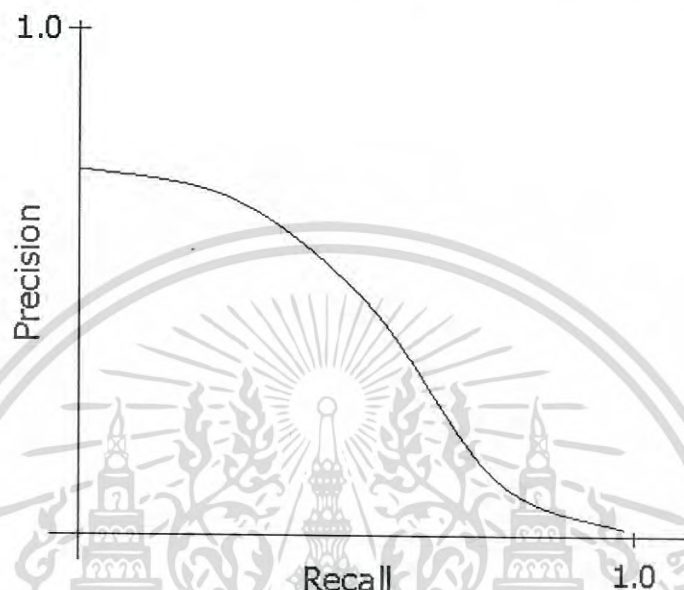


Title	System of Analyzing English Sentences into Interlingua for Machine Translation	ชื่อวิทยานิพนธ์	ระบบการวิเคราะห์ประโยคภาษาอังกฤษไปสู่ภาษากลางเพื่อการ แปลภาษา ด้วยคอมพิวเตอร์
Abstract	In the field of machine translation (MT) system, number of methods had been introduced so far. Most of methods are based on natural language processing and computer software techniques. Alternatively, the interlingua approach is generally used in most MT systems. The interlingua is used to be an intermediate expression in order to transfer the syntactic and semantic information of source language into target language generation. This methodology has been successfully recognized in both accuracy and efficiency.	บทคัดย่อ	ในการแปลภาษาด้วยระบบคอมพิวเตอร์นี้มีหลายวิธี โดยใช้หลักการทางด้าน การประมวลผลภาษาธรรมชาติและวิทยาการ ทางซอฟต์แวร์ ทั้งนี้วิธีการแปล โดยใช้ภาษากลางเป็นสื่อ ถ่ายทอดโครงสร้างทางไวยากรณ์ และความหมาย ของภาษาต้นแบบ ไปสู่การสร้างภาษาเป้าหมายนั้น เป็นวิธีการที่ให้ความถูกต้อง และมีประสิทธิภาพในการแปลสูง นอกจากนั้นยังเหมาะ สำหรับการแปล หลาย ๆ ภาษาในระบบเดียวกัน เพราะจะลดจำนวน ซอฟต์แวร์ จำนวน พจนานุกรม และขั้นตอนการถ่ายทอดให้น้อยลง วิทยานิพนธ์นี้ได้เสนอแนวทาง ในการวิเคราะห์โครงสร้าง ประโยคภาษาอังกฤษ เพื่อถ่ายทอดข้อมูลทาง ไวยากรณ์และความ หมายไปสู่สื่อภาษากลางที่กำหนด
Author	Aranya Preechawai	ชื่อนิสิต	อรุณญา ปรีชาไว
Major	Master. Engineering (Electrical Engineering)	ระดับปริญญา	วิทยานิพนธ์มหาบัณฑิต. วิศวกรรมศาสตร์ (วิศวกรรมไฟฟ้า)
Advisor	Asst.Prof. Dr. Ruttikorn Varakulsiripunth	ชื่ออาจารย์ที่ ปรึกษา	
Institute	King Mongkut's Institute of Technology Ladkrabang. Bangkok (Thailand). Graduate School.	ชื่อสถาบัน	
Year	1993	ปีที่จบการศึกษา	
Descriptor		คำสำคัญ	

รูปที่ 2.5 ตัวอย่างบทคัดย่อจากฐานข้อมูลวิทยานิพนธ์ไทย

2.5. การจัดเตรียมชุดข้อมูลทดสอบ (Test Collection)

การทดสอบประสิทธิภาพของระบบค้นคืนสารสนเทศเป็นสิ่งที่สำคัญมาก เปรียบเสมือนกับเป็นเครื่องมือวัดประสิทธิภาพของระบบ โดยจะวัดค่าความแม่นยำ (Precision) ที่ระดับการเรียกคืน (Recall) ต่างๆกัน แล้วนำเสนอผลการทดสอบในรูปของกราฟ ดังรูปที่ 2.6



รูปที่ 2.6 การแสดงค่าความแม่นยำที่การเรียกคืนระดับต่างๆ

จากรูป 2.6 เป็นลักษณะของกราฟโดยทั่วไป ซึ่งหมายถึงระดับความแม่นยำจะลดลงเมื่อเราต้องการเอกสารที่เกี่ยวข้องจำนวนมากขึ้น เช่นเมื่อต้องการเอกสารที่เกี่ยวข้องทั้งหมดที่มีอยู่ในชุดข้อมูล เราอาจพบว่าความแม่นยำลดลงจนเกือบเป็นศูนย์

2.6.1. ชุดข้อมูลทดสอบ (Test Collection) สำหรับการศึกษาการค้นคืนสารสนเทศ ประกอบด้วย

- 2.6.1.1. ชุดเอกสาร (Document Collection) อาจรวบรวมมาจากสิ่งพิมพ์
- 2.6.1.2. คำสั่งสอบถาม (Query) มาจากการเก็บข้อมูลที่เคยป้อนผู้ใช้ หรือสร้างขึ้นใหม่
- 2.6.1.3. เอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม (Relevance Set) ได้มาจากการเลือกโดยบุคคลที่มีความรู้ในกรณีที่เอกสารในชุดเอกสาร มีจำนวนไม่มากนัก แต่ถ้าชุดเอกสารมีจำนวนมาก เช่นชุดข้อมูลทดสอบของ TREC หรือ CLEF มีเอกสารจำนวนมาก ไม่สามารถใช้คนตรวจสอบได้ จึงมีการใช้วิธีการ Pooling จากระบบค้นคืนสารสนเทศหลายระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2. ชุดข้อมูลทดสอบสำหรับการค้นคืนข้ามภาษา (CLIR Test Collection)

ชุดข้อมูลทดสอบสำหรับการค้นคืนข้ามภาษาจะมีคำสั่งสอบถามกับเอกสารที่เกี่ยวข้องกับคำสั่งสอบถามต่างภาษากัน จากการสำรวจของเราพบว่า ยังไม่มีชุดข้อมูลทดสอบการค้นคืนข้ามภาษา ภาษาไทย-ภาษาอังกฤษ ดังนั้นเราจึงจัดเตรียมขึ้นเอง โดยมีเป้าหมายให้ได้ชุดข้อมูลทดสอบที่มีคุณภาพดี โดยไม่จำเป็นต้องใช้ผู้ที่ชำนาญภาษาอังกฤษ โดยเราจะดัดแปลงจากหัวข้อเรื่องของบทคัดย่อวิทยานิพนธ์ของฐานข้อมูลวิทยานิพนธ์ไทย จากนั้นจึงดัดแปลงหัวข้อวิทยานิพนธ์ที่เป็นภาษาอังกฤษโดยวิธีเดียวกัน

วิทยานิพนธ์ที่มีหัวข้อเป็นภาษาไทยและภาษาอังกฤษจากจากทั้งสิ้นจำนวน 41,566 บทคัดย่อ คัดแยกบทคัดย่อตั้งแต่ปี 2528-2538 จำนวน 7,341 บทคัดย่อแยกเป็นชุดข้อมูลทดสอบ (บทคัดย่อตั้งแต่ปี 2539-2545 จำนวน 34,225 บทคัดย่อใช้เป็นชุดข้อมูลฝึก สำหรับวิธีการแปลคำสั่งสอบถามโดยวิธีการ PRF ดังจะกล่าวถึงในบทที่ 4)

2.6.2.1 การสร้างคำสั่งสอบถามภาษาไทย

บทคัดย่อที่แยกเป็นชุดข้อมูลทดสอบจำนวน 7,341 เรื่อง เรานำหัวข้อวิทยานิพนธ์มาสร้างเป็นคำสั่งสอบถามภาษาไทย โดยมีกระบวนการสร้างชุดข้อมูลทดสอบซึ่งใช้คนทำงานร่วมกับระบบค้นคืนสารสนเทศ

การสร้างคำสั่งสอบถาม ทำโดยการดัดแปลงหัวข้อวิทยานิพนธ์ ให้สัมพันธ์ (Relevance) กับบทคัดย่อวิทยานิพนธ์อื่นๆ ซึ่งปกติหัวข้อวิทยานิพนธ์จะไม่ซ้ำกันและจำกัดการศึกษาให้เฉพาะเจาะจง จึงทำให้พบวิทยานิพนธ์ที่คล้ายกันจำนวนไม่มาก เราจึงต้องทำการดัดแปลงหัวข้อวิทยานิพนธ์ เพื่อสร้างคำสั่งสอบถาม ภาษาไทยด้วยกระบวนการตามลำดับดังต่อไปนี้

- 1) เลือกหัวข้อวิทยานิพนธ์ภาษาไทยมาใช้เป็น Query จำนวน 1 เรื่อง
- 2) ตัดศัพท์เฉพาะทางหรือชื่อเฉพาะออกไป เช่นชื่อสถานที่
- 3) ค้นคืน Query ที่ได้ในข้อ 1 จากหัวข้อวิทยานิพนธ์ที่เหลือ
- 4) ให้ผู้ชำนาญตรวจดูเอกสารที่ค้นได้ประมาณ 400 เอกสารแรก ว่ามีเอกสารใดที่เกี่ยวข้อง (Relevance) กับ Query บันทึกหมายเลขเอกสาร ไว้เป็น Relevance Set
- 5) ถ้ามีเอกสารมากกว่า 5 เอกสารที่เกี่ยวข้อง ให้เลือก Query นี้เป็น Query สำหรับทดสอบ
- 6) ถ้ามีเอกสารน้อยกว่า 5 เอกสารที่เกี่ยวข้อง ให้กลับไปทำข้อ 2)
- 7) ถ้า Query ไม่สามารถให้ผลการค้นคืนที่รัดกุม ให้สุ่มเลือก Query อื่นขึ้นมาแทน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2.2. การสร้างคำสั่งสอบถามภาษาอังกฤษ

เมื่อเสร็จสิ้นขบวนการข้างต้น เราจะได้ชุดข้อมูลทดสอบภาษาไทย แต่เนื่องจากหัวข้อวิทยานิพนธ์มีคำแปลภาษาอังกฤษ เราจึงสามารถสร้างคำสั่งสอบถามภาษาอังกฤษที่เกี่ยวข้องกับเอกสารภาษาไทยได้ โดยการตัดชื่อเฉพาะหรือศัพท์เฉพาะ เช่นเดียวกับการสร้างคำสั่งสอบถามภาษาไทย

2.6.2.3. ตัวอย่างกระบวนการสร้างคำสั่งสอบถามภาษาไทย

1) เลือกหัวข้อเอกสารหมายเลข 2303

"ความรู้ ทักษะ และ การปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์ ของทันตแพทย์ จังหวัด เชียงใหม่ "

2) ตัดชื่อจังหวัดเชียงใหม่ออกไป

"ความรู้ ทักษะ และ การปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์ ของทันตแพทย์"

3) ผลการค้นคืน 400 เอกสารแรก ไม่เกี่ยวข้องกับคำสั่งสอบถามจึงตัด คำว่า "ทันตแพทย์" ออกไป เหลือเพียง "ความรู้ ทักษะ และ การปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์"

อันดับ	หมายเลข	ชื่อเรื่อง
1	6702	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับโรคเอดส์ของนักศึกษาวิทยาลัยพลศึกษา
2	8263	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับโรคเอดส์ ของผู้มารับบริการ ณ คลินิก นิรนาม สภา กษาดไทย
3	2302	ความสัมพันธ์ระหว่าง ความรู้ ทักษะเกี่ยวกับโรคเอดส์ กับการปฏิบัติในการป้องกันการติดเชื้โรคเอดส์ ของพยาบาล วิชาชีพในการพยาบาลสูติศาสตร์และนรีเวชวิทยา โรงพยาบาล มหาราชนครเชียงใหม่
4	6663	การเปิดรับสื่อความรู้ ทักษะ และ พฤติกรรมในการป้องกันโรคเอดส์ ของผู้ปกครองนักเรียน ระดับประถมศึกษาปีที่ 6 เขตกรุงเทพมหานคร
5	10372	การเปิดรับสื่อ ความรู้ ทักษะ และ พฤติกรรมการป้องกันโรคเอดส์ของหญิง มีครรภ์ที่มาฝากครรภ์ที่โรงพยาบาลจุฬาลงกรณ์
6	10649	ผลของการสอนสอดแทรกเรื่องโรคเอดส์และการป้องกันโรคเอดส์ในวิชาสังคมศึกษา ที่มีต่อเจตคติต่อการป้องกันโรคเอดส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1
7	8822	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับสิ่งแวดล้อมของนักศึกษาในมหาวิทยาลัย
8	9072	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับสิ่งแวดล้อมของนักศึกษาวิทยาลัยอาชีวศึกษา
13	8793	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับสิ่งแวดล้อม ของนักเรียน ชั้นมัธยมศึกษาปีที่ 3 ในภาคตะวันออกเฉียงเหนือ

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

14	9392	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับสิ่งแวดล้อมของนักเรียนชั้นประถมศึกษา ปีที่ 6 ในภาคตะวันออกเฉียงเหนือ
15	7989	ความรู้ และ ทักษะเกี่ยวกับโรคเอดส์ของนักเรียนชั้นประถมศึกษาปีที่ 6 ใน โรงเรียนรัฐบาล กรุงเทพมหานคร
16	7229	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับสิ่งเสพติดให้โทษ ของนักศึกษา ระดับประกาศนียบัตรวิชาชีพชั้นสูง ในสถาบันเทคโนโลยีราชมงคล
17	2272	พฤติกรรมอนามัยของชาวไทยภูเขาเผ่าม้ง บ้านหนองหอยเก่า ตำบลแม่แรม อำเภอแมริม จังหวัดเชียงใหม่ : ความรู้และ การปฏิบัติตนเกี่ยวกับการป้องกันโรคเอดส์
22	7967	ความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับการเลือกใช้สารปรุงแต่งรสและวัตถุ เจือปนอาหารของผู้ประกอบอาหารในโรงเรียนประถมศึกษาและมัธยมศึกษา อำเภอเมือง จังหวัดขอนแก่น
23	8111	ความรู้ ทักษะเรื่องโรคเอดส์ และ พฤติกรรมเสี่ยงต่อการติดเชื้อโรคเอดส์ ของผู้นำเยาวชนสภาเยาวชนสุพรรณบุรีโรงเรียนและนักเรียนมัธยมศึกษาตอน ปลาย จังหวัดนครราชสีมา ปี 2536
24	6849	ความรู้ ทักษะ และ การปฏิบัติตนเกี่ยวกับโรคเอดส์ของผู้บริหารสถานศึกษา และครูอนามัยโรงเรียนประถมศึกษา สังกัดสำนักงานคณะกรรมการ การประถมศึกษาแห่งชาติ เขตการศึกษาศึกษา 8
25	1524	การศึกษาความรู้ ทักษะ และ การปฏิบัติเกี่ยวกับกิจกรรมสหกรณ์ในโรงเรียนประถมศึกษา ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานการประถมศึกษาจังหวัดเลย

รูปที่ 2.7 ตัวอย่างแสดงการสร้างชุดข้อมูลทดสอบ

- 4) ทำการค้นคืนอีกครั้ง พบว่าเอกสารที่เกี่ยวข้องกับคำสั่งสอบถามมีจำนวนมากกว่า 5 เอกสาร ดังแสดงตัวอย่างในรูปที่ 2.7 ได้แก่ 6272, 8263, 2302, 6663, 8111, 6849, 8103, 6423, 8090, 7622, 3330, 3263, 7888
- 5) รวบรวมหมายเลขเอกสารที่ Relevance ให้เป็น Relevance Set

2.6.2.4. ตัวอย่างกระบวนการสร้างคำสั่งสอบถามภาษาอังกฤษ

เอกสารหมายเลข 2303 มีชื่อเรื่องภาษาอังกฤษคือ

"Knowledge, Attitude and Practice in AIDS Prevention by Dentists in Chiang Mai Province"

เราตัดคำว่า "เชียงใหม่" และ "ทันตแพทย์" ออกจาก ชื่อเรื่องในภาษาอังกฤษ ดังนั้นคำสั่งสอบถามภาษาอังกฤษคือ "Knowledge, Attitude and Practice in AIDS"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เราได้สร้างคำสั่งสอบถามที่เป็นภาษาไทยและภาษาอังกฤษจำนวน 10 หัวข้อดังแสดงในตารางที่ 2.3 โดยมีเอกสารที่เกี่ยวข้องกับคำสั่งสอบถามดังแสดงในตารางที่ 2.4

ตารางที่ 2.3 ชื่อเรื่องวิทยานิพนธ์ที่นำมาใช้เป็น Query จำนวน 10 เรื่อง

Label	Language	Query
Query1	Thai	ความรู้ ทศนคติ และการปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์
	English	KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION
Query2	Thai	การใช้ประโยชน์และความพึงพอใจจากการเปิดรับรายการข่าวทางโทรทัศน์ ของประชาชน ในกรุงเทพมหานคร
	English	THE USE AN GRATIFICATION OF TELEVISION NEWS AMONG PEOPLE IN BANGKOK METROPOLITAN AREA
Query3	Thai	ความสัมพันธ์ระหว่างความรู้ความเข้าใจ ทักษะการคิดคำนวณ และทักษะการแก้โจทย์ ปัญหาในวิชาคณิตศาสตร์ ในระดับชั้น ประถมศึกษา
	English	RELATIONSHIP AMONG UNDERSTAND COMPUTATIONAL AND WORD-PROBLEM SOLVE SKILL IN PRIMARY SCHOOL MATHEMATICS
Query4	Thai	ทักษะการบริหารของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษา
	English	ADMINISTRATIVE SKILL OF PRIMARY SCHOOL ADMINISTRATOR UNDER THE OFFICE OF PRIMARY EDUCATION
Query5	Thai	การเพาะเลี้ยงสาหร่ายเกลียวทองเพื่อใช้เป็นอาหาร
	English	CULTIVATION OF SPIRULINA AS FEED
Query6	Thai	อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง
	English	THE INFLUENCE OF SUPPLEMENTARY NITROGEN FERTILIZE APPLY AT EAR POD FORMATION ON THE QUALITY AN YIELD OF SOYBEAN
Query7	Thai	การดูแลสุขภาพของผู้สูงอายุ
	English	HEALTH CARE OF THE AGE
Query8	Thai	การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อวางแผนการใช้ที่ดิน
	English	THE APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM FOR LAND USE PLAN
Query9	Thai	การย่อยสลายแป้งมันสำปะหลัง โดยใช้เอนไซม์
	English	HYDROLYSIS OF CASSAVA STARCH BY ENZYME
Query10	Thai	พฤติกรรมของแผ่นพื้นคอนกรีตท้องเรียบชนิดอัดแรงบางส่วน
	English	BEHAVIOUR OF PARTIAL PRESTRESS CONCRETE FLAT PLATE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.4 หมายเลขเอกสารที่ Relevance กับ Query

Lable	หมายเลขเอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม
Query1	6272, 8263, 2302, 6663, 8111, 6849, 8103, 6423, 8090, 7622, 3330, 3263, 7888
Query2	9429, 9809, 6725, 10524, 6636, 19153, 8041, 8122, 7336, 19175
Query3	2102, 1386, 10627, 6528, 6270, 5952, 1389, 2518, 8363, 1602, 2196, 1278, 7477, 9914, 10317, 6941, 7522, 1546
Query4	6101, 1375, 2199, 1409, 9442, 1479, 1512, 1518, 9566, 10528, 8388, 10358, 7743, 6024, 1498, 1380, 9385, 5980, 1345, 2198, 7928, 1410, 3370, 2471, 8156, 3484, 9022, 1599, 2460, 1290, 1472, 2833, 8149, 10795, 2073, 2450, 10736, 10117, 3215, 21354, 1340
Query5	545, 0468, 0326, 8282, 2154, 2156
Query6	2404, 2668, 2227, 3000, 2128, 2137, 2136, 2402
Query7	2261, 3445, 2296, 2714, 3004, 7539, 6474, 2910, 6400, 9120, 2257, 6044, 2295
Query8	9947, 8972, 9366, 9405, 2576, 9043, 7922, 8987, 2753, 9435
Query9	8632, 21896, 21457, 10607, 8123
Query10	7905, 9999, 2789, 6453, 10624, 10520, 6121, 0346

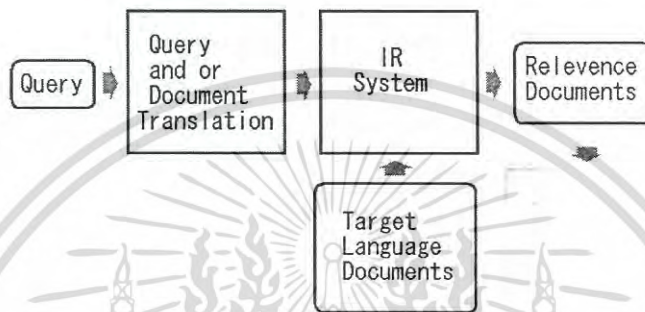
การทดสอบการค้นคืนข้ามภาษา ภาษาไทยเป็นภาษาอังกฤษ จะใช้ชุดข้อมูลภาษาอังกฤษร่วมกับคำสั่งสอบถามภาษาไทย ส่วนการทดสอบการค้นคืนข้ามภาษา ภาษาอังกฤษเป็นภาษาไทยจะใช้ชุดข้อมูลภาษาอังกฤษร่วมกับคำสั่งสอบถามภาษาไทย นอกจากนี้เรายังสามารถใช้ชุดข้อมูลทดสอบนี้ในการค้นคืนสารสนเทศภาษาเดียวได้อีกด้วย เราจะเรียกชุดข้อมูลทดสอบชุดนี้ว่า "ชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทยปี 2528-2538"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

ระบบค้นคืนสารสนเทศสำหรับการค้นคืนข้ามภาษา

ในการศึกษาการค้นคืนข้ามภาษาโดยวิธีการแปลคำสั่งสอบถาม เมื่อคำสั่งสอบถามถูกแปลแล้ว ระบบค้นคืนสารสนเทศจะให้ขั้นตอนสุดท้ายเพื่อค้นเอกสารที่เกี่ยวข้องกันกับคำสั่งสอบถามที่แปลแล้ว ดังรูปที่ 3.1



รูปที่ 3.1 CLIR Block Diagram

เนื้อหาในบทนี้จะกล่าวถึงระบบค้นคืนสารสนเทศที่เราใช้ในการทดลอง ซึ่งทำงานตามแบบจำลองเวกเตอร์สเปซ (Vector Space Model) เราจะอธิบายแนวคิดของแบบจำลองเวกเตอร์สเปซในเบื้องต้น และกล่าวถึงการตัดแบ่ง (Parsing) ข้อความภาษาไทยและภาษาอังกฤษเพื่อสร้างดัชนีแบบย้อนกลับ (Inverted Index) และการคำนวณค่าถ่วงน้ำหนักของคำ (Term Weighting)

เรายังทำการทดสอบระบบค้นคืนสารสนเทศที่เราใช้ในการศึกษานี้ เพื่อให้เกิดความเชื่อมั่นในประสิทธิภาพของระบบในการค้นคืนภาษาเดียว โดยใช้ชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทยปี 2528-2538 ที่ได้สร้างขึ้น

3.1. ระบบค้นคืนสารสนเทศ

การศึกษาเรื่องการค้นคืนสารสนเทศ ส่วนประกอบที่สำคัญในการทดลองคือระบบค้นคืนสารสนเทศซึ่งระบบ SMART [20] เป็นระบบที่ถูกใช้มากที่สุดในงานวิจัย เนื่องจากความยืดหยุ่นในตัวระบบที่เปิดโอกาสให้นักวิจัยสามารถปรับแต่งค่าตัวแปรต่างๆ รวมทั้งเป็นระบบที่แจ่มง่ายโดยไม่คิดมูลค่าพร้อมรหัสต้นฉบับ ระบบ SMART ออกแบบโดยใช้แบบจำลองเวกเตอร์สเปซ (Vector Space Model) ที่ผู้สร้าง SMART เป็นผู้บุกเบิกการใช้งาน ตั้งแต่ปี ค.ศ. 1970

ระบบค้นคืนสารสนเทศสมัยใหม่ ตัวอย่างเช่น Google.com หรือ Altavista.com ยังใช้พื้นฐาน

จากแบบจำลองเวกเตอร์สเปซ แต่มีความสามารถค้นคืนโดยใช้คำสั่งสอบถามได้หลายภาษา เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปโฆษณาการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถแสดงผลได้หลายภาษาโดยใช้การเข้ารหัสอักขระแบบยูนิโคด (Unicode) ดังแสดงในรูปที่ 3.1 ซึ่งมีการแสดงผลเป็นอักขระภาษาอังกฤษ ภาษารัสเซีย ภาษาจีน และ ภาษาญี่ปุ่น ในผลการค้นหาเดียวกัน เรียกว่าเป็นระบบค้นคืนสารสนเทศหลายภาษา (Multilingual IR System) หรือเรียกย่อว่า MLIR

[Строки | Технология .Net | mista.ru](#)

... типами, поддерживаемыми Common Language Runtime, и другими ... String Demo -----
Imports System Imports ... все символы Unicode. Ключ UTF8Encoding ...
www.mista.ru/net/string.htm • Translate

[Japan Lisp User Group: Resources: Lisp Implementations in Japan \(Japanese\)](#)

日本のLisp処理系。解説: プログラマなら一度はLISPを作ると言われており、実験的なものを含めれば無数にあるでしょうが、ここではおおむね...

jp.franz.com/jlug/resources/implementations.htm • Translate

[TCF_TurnSSPHitPoint 00000000](#)

... 据。Length 文本数据的 * 度。Language 文本数据的 * 言字符 ... 文BIG5 TCE_ENGLISH 英 * TCE_UNICODE 000000 HitPointType 命 ... Rec, Hit; CHAR Detail[1024]; /* DEMO000000 */ HTRSSELECT ...

www.trs.com.cn/handbook/TRSClientAPI/Fun...SSPHitPoint.htm • Translate

[More pages from www.trs.com.cn](#)

รูปที่ 3.2 แสดงตัวอย่างผลการค้นของ MLIR จาก Altavista.com

MLIR จะให้ความคล่องตัวมากกว่าระบบค้นคืนสารสนเทศภาษาเดียว (Monolingual IR System) โดยเฉพาะอย่างยิ่งการศึกษาการค้นคืนข้ามภาษา ซึ่งในบางครั้งเอกสารมีอาจหลายภาษาเขียนผสมกัน (Mixed language document) ดังตัวอย่างในรูป 3.2

3.2. แบบจำลองเวกเตอร์สเปซ (Vector Space Model)

แนวคิดเรื่อง แบบจำลองเวกเตอร์สเปซนั้น คำตั้งสอบถาม และ เอกสาร (Document) เป็นเวกเตอร์ในที่ว่าง (Space) ของเมตริกซ์ของคำ (Term) ถ้า ค่ามุมเวกเตอร์ ของคำตั้งสอบถาม กับ เอกสารมีค่าใกล้เคียงกัน แสดงว่า เอกสารกับคำตั้งสอบถาม นั้น ตรงประเด็นกัน (Relevance) หรือเกี่ยวข้องกัน

3.2.1. การทำดรรชนี (Indexing) เพื่อเพิ่มความเร็วในการค้นคืน โดยจะทำดรรชนีไว้ล่วงหน้าก่อนการค้นคืน ดรรชนีนั้นจะเชื่อมโยงระหว่างคำกับเอกสาร เพื่อให้การค้นว่าคำนี้อยู่ในเอกสารใดเป็นไปด้วยความรวดเร็ว ดรรชนีชนิดนี้เรียกว่าดรรชนีย้อนกลับ (Invert Index)

Doc1	cat cat mouse mouse
Doc2	cat dog dog
Doc3	cat mouse
Doc4	cat dog

รูปที่ 3.3 เอกสาร Doc1, Doc2, Doc3, Doc4

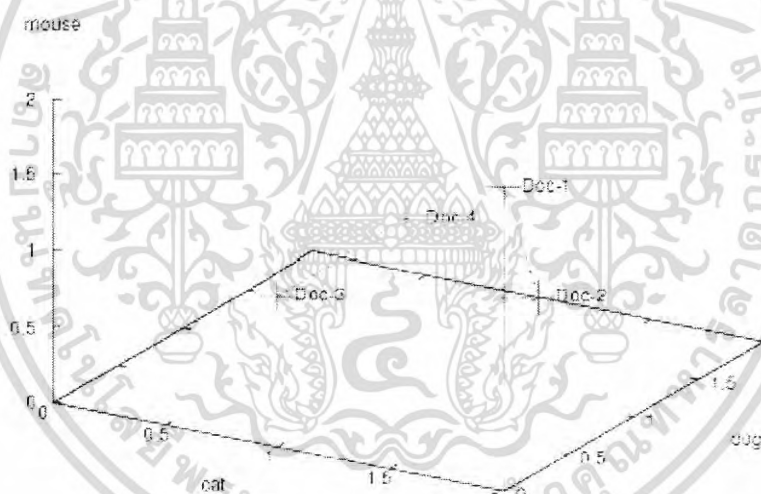
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตัวอย่างในรูปที่ 3.3 เอกสาร Doc1, Doc2, Doc3 และ Doc4 ที่มีคำว่า cat, dog และ mouse เป็นองค์ประกอบ ตัวเลขแสดงจำนวนคำที่ปรากฏในเอกสาร เป็นค่าถ่วงน้ำหนัก TF เขียนเป็นเมตริกซ์ ดังรูป 3.4

Term	Doc 1	Doc 2	Doc 3	Doc 4
cat	2	1	1	1
Dog	0	2	0	1
Mouse	2	0	1	0

รูปที่ 3.4 ตารางนี่ย้อนกลับ (Inverted Index)

3.2.2. การถ่วงน้ำหนัก (Term Weighing) จากรูปที่ 3.4 เราใช้ค่าความถี่ของคำ (Term Frequency หรือ TF) เป็นค่าถ่วงน้ำหนัก เมื่อแสดงเวกเตอร์ของเอกสาร Doc1, Doc2, Doc3 และ Doc4 ในรูป 3.5



รูปที่ 3.5 แสดงเวกเตอร์ของเอกสาร Doc1, Doc2, Doc3 และ Doc4

ในทางปฏิบัติการใช้ TF อย่างเดียวไม่พอเพียงแก่การถ่วงน้ำหนักอย่างยุติธรรม เพราะว่าคำที่พบได้บ่อย เป็นคำที่ไม่มีความสำคัญในการคัดแยกเอกสารที่เกี่ยวข้อง เช่นคำเชื่อม is, am, are, and, เป็น, และ ดังเช่น จะทำให้ผลการค้นคืนได้เอกสารที่ไม่เกี่ยวข้อง ดังนั้นจึงมีการใช้สแกนกลับของความถี่เอกสาร (Invert Document Frequency หรือ IDF) หมายถึง ส่วนกลับของจำนวนเอกสารที่มีคำนั้นปรากฏ เช่นถ้าคำนั้นปรากฏในเอกสารจำนวนมากถึง 100000 เอกสาร ค่า IDF จะมีค่าต่ำเพียง 0.00001 และถ้าคำใดปรากฏในเอกสารจำนวนน้อย เช่น 10 เอกสาร ค่า IDF จะมีค่าสูงกว่า มีค่าเพียง 0.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

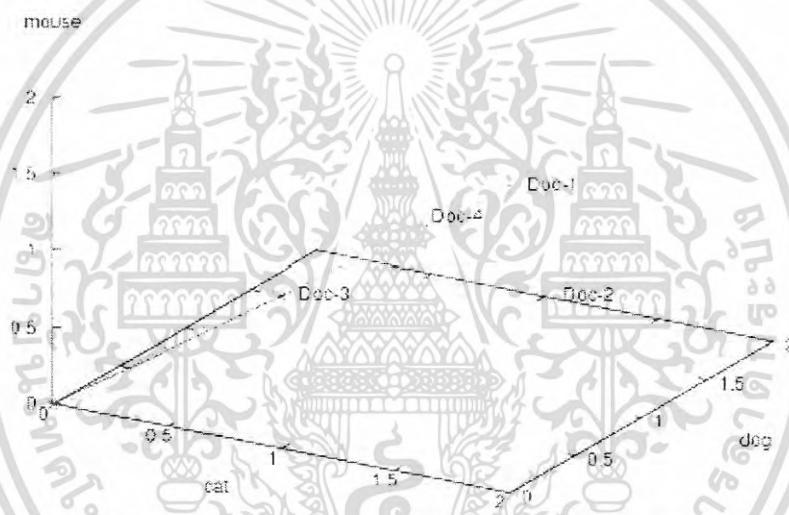
เนื่องจากค่า IDF จะเป็นตัวเลขทศนิยมหลายหลัก และความสัมพันธ์ระหว่าง IDF กับลำดับความสำคัญของคำจะเป็นเส้นโค้งคล้ายกราฟ LOG โดยทั่วไปจึงปรับให้เป็นเชิงเส้นโดยการใช้ค่าถ่วงน้ำหนัก IDF ที่เป็นค่า LOG ดังสมการ

$$\text{IDF} = \log (N / n)$$

N คือจำนวนเอกสารทั้งหมด

n คือจำนวนเอกสารที่มีคำนั้น

3.2.3. การเปรียบเทียบความเหมือน(Similarity Measuring) จากรูป 3.5 เวกเตอร์ Doc1, Doc2, Doc3 และ Doc4 เป็นเวกเตอร์ในที่ว่างของเมตริกคำว่า cat , dog และ mouse ซึ่งถ้าหากจะเปรียบเทียบว่าเอกสารใดมีความคล้ายกัน ด้วยค่ามุม เราจะพบว่า Doc1 และ Doc3 มีค่ามุมเท่ากัน (เส้นทึบในรูป 3.6 ลากตัดผ่าน Doc1 และ Doc3)



รูปที่ 3.6 แสดงเปรียบเทียบความเหมือนด้วย Cosine Similarity

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

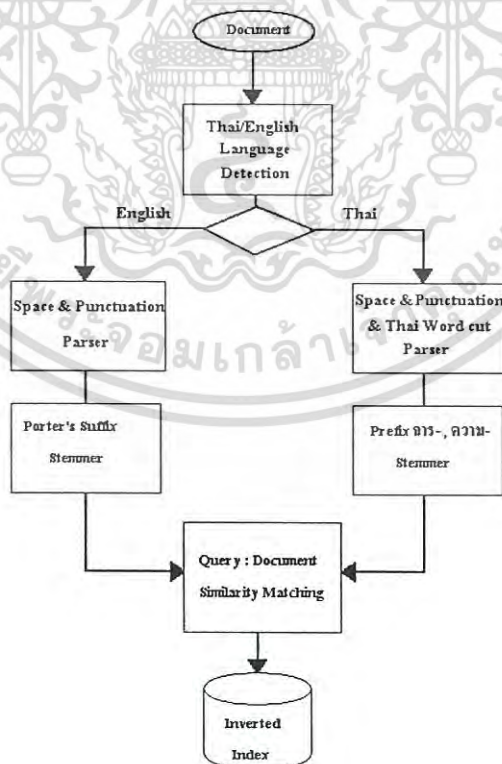
3.3. ระบบค้นคืนสารสนเทศสำหรับการศึกษา CLIR ภาษาไทย-อังกฤษ

ผู้วิจัยได้ดัดแปลงระบบค้นคืนสารสนเทศที่ได้เขียนขึ้นเอง (ThaiSearch) ซึ่งเป็นระบบ MLIR ที่ค้นคืนได้ สองภาษา คือภาษาไทย และภาษาอังกฤษ โดยมีรายละเอียดของระบบดังนี้

	ภาษาไทย	ภาษาไทย
การตัดแยกคำ (Parsing)	แบ่งโดยใช้พจนานุกรมและเครื่องหมายวรรคตอน	แบ่งด้วยเครื่องหมายวรรคตอน
การ Stemming	ใช้กับคำนำหน้า (Prefix) เช่น การ-, ความ-	ใช้กับคำลงท้าย (Suffix) เช่น -ing, -ly, -s, -d, -er
ค่าถ่วงน้ำหนักของคำศัพท์ (Term Weighting)	TF x IDF	TF x IDF
คำหยุด (Stopword)	ใช้	ใช้

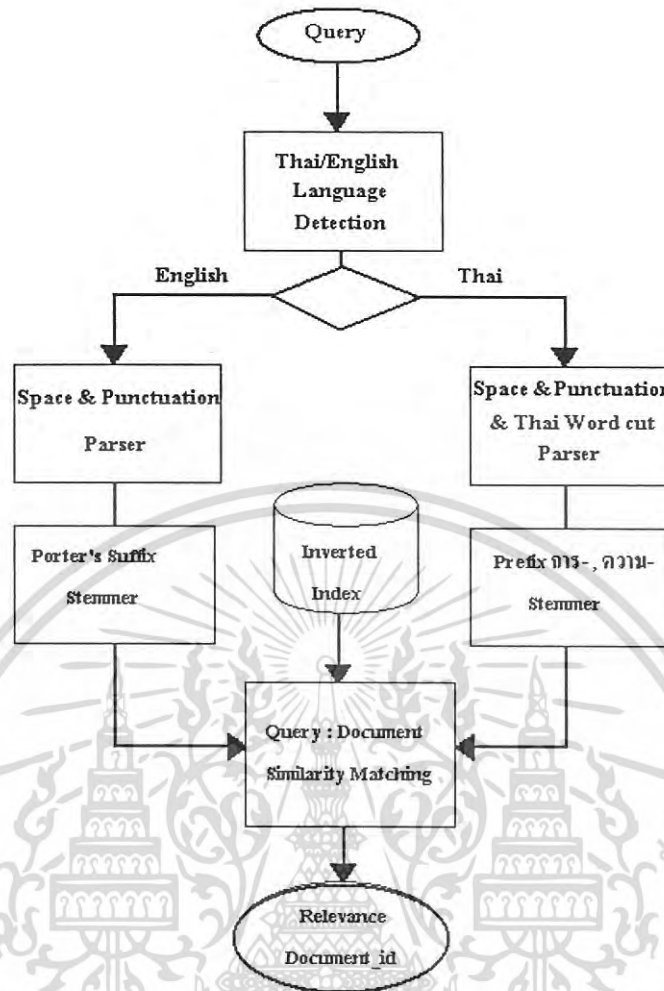
ตาราง 3.1 เปรียบเทียบการสร้างดรรชนีจากเอกสารภาษาไทยและภาษาอังกฤษ

การทำงานของ ThaiSearch ดังแสดงในรูปที่ 3.6 และ 3.7 มีการจัดการภาษาคำสั่งสอบถามภาษาไทยและภาษาอังกฤษแยกจากกัน ในส่วนของ Parser และ Stemmer ดังแสดงในตาราง 3.2



รูปที่ 3.6 การสร้าง Inverted Index

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 การค้นคืนจาก Inverted Index

3.4. ทดสอบประสิทธิภาพระบบค้นคืนสารสนเทศ

ข้อมูลที่ใช้ทดสอบคือ ชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทยปี 2528-2538 ทำการทดสอบประสิทธิภาพการค้นคืน ภาษาไทยภาษาเดียว(Monolingual)

การทดสอบประสิทธิภาพจะทำการวัดค่าความแม่นยำ (Precision) ที่ค่าการเรียกคืน (Recall) ตั้งแต่ 0.1 – 1.0 เป็นช่วงๆ ช่วงละ 0.1 โดยค่าความแม่นยำที่การเรียกคืน 0.0 นั้นได้จาก Interpolate

ผลการทดสอบดังตาราง 3.2 และภาษาอังกฤษภาษาเดียว ดังตาราง 3.3 ผลการทดสอบแต่ละคำสั่งสอบถาม ทั้ง 10 คำสั่งสอบถามโดยละเอียด รวมถึงค่าความแม่นยำที่แท้จริงก่อนการปรับกราฟ (Actual precision) อยู่ในภาคผนวก ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.2 ค่าความแม่นยำของการค้นคืนด้วยคำสั่งสอบถาม ภาษาไทย*

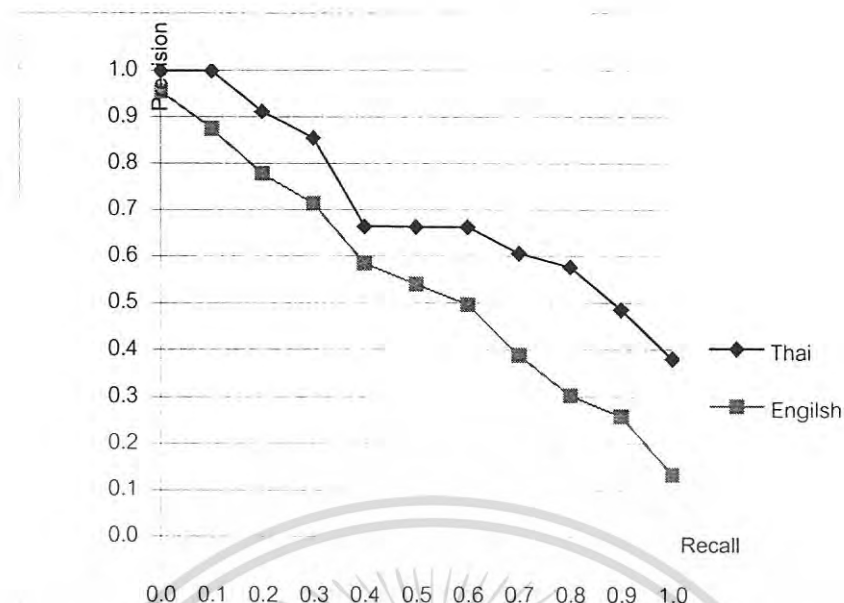
Recall	Query Label										Avg	
	1	2	3	4	5	6	7	8	9	10		
0.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.2	0.83	1.00	0.40	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91
0.3	0.83	0.78	0.40	0.70	1.00	1.00	0.85	1.00	1.00	1.00	1.00	0.86
0.4	0.45	0.78	0.40	0.70	1.00	0.22	0.85	1.00	0.25	1.00	1.00	0.66
0.5	0.45	0.78	0.40	0.68	1.00	0.22	0.85	1.00	0.25	1.00	1.00	0.66
0.6	0.45	0.78	0.40	0.68	1.00	0.22	0.85	1.00	0.25	1.00	1.00	0.66
0.7	0.45	0.78	0.40	0.68	0.57	0.21	0.85	0.88	0.25	1.00	1.00	0.61
0.8	0.45	0.67	0.40	0.68	0.57	0.21	0.85	0.83	0.25	0.86	0.86	0.58
0.9	0.35	0.43	0.36	0.68	0.42	0.21	0.85	0.83	0.22	0.50	0.50	0.49
1.0	0.33	0.43	0.35	0.68	0.08	0.21	0.17	0.83	0.21	0.50	0.50	0.38

ตาราง 3.3 ค่าความแม่นยำของการค้นคืนด้วยคำสั่งสอบถาม ภาษาอังกฤษ*

Recall	Query Label										Avg	
	1	2	3	4	5	6	7	8	9	10		
0.0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	0.25	1.00	1.00	1.00	0.88
0.1	0.50	1.00	1.00	1.00	1.00	1.00	1.00	0.25	1.00	1.00	1.00	0.88
0.2	0.50	0.36	0.75	1.00	1.00	1.00	1.00	0.18	1.00	1.00	1.00	0.78
0.3	0.36	0.36	0.32	0.92	1.00	1.00	1.00	0.18	1.00	1.00	1.00	0.71
0.4	0.27	0.36	0.32	0.67	1.00	0.75	0.08	1.00	0.40	1.00	1.00	0.58
0.5	0.27	0.36	0.32	0.59	1.00	0.67	0.06	1.00	0.33	0.80	0.80	0.54
0.6	0.27	0.22	0.15	0.46	1.00	0.67	0.06	1.00	0.33	0.80	0.80	0.50
0.7	0.27	0.16	0.10	0.43	0.45	0.50	0.05	0.89	0.31	0.71	0.71	0.39
0.8	0.27	0.14	0.10	0.42	0.45	0.24	0.05	0.89	0.31	0.16	0.16	0.30
0.9	0.27	0.05	0.10	0.30	0.45	0.24	0.04	0.75	0.22	0.16	0.16	0.26
1.0	0.09	0.05	0.00	0.05	0.33	0.24	0.01	0.30				0.13

*รายละเอียดของการทดสอบดูได้จากภาคผนวก ค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.8 กราฟความแม่นยำต่อการเรียกคืนของการค้นคืนภาษาเดียว

3.5 สรุปผลการทดสอบระบบค้นคืนสารสนเทศ

จากผลการทดสอบด้วยระบบค้นคืนสารสนเทศ ด้วยชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทย ปี 2528-2538 พบว่าประสิทธิภาพการค้นคืน ในแง่ความแม่นยำ (Precision) ได้ผลดีที่ทุกๆค่า Recall ค่าที่ได้จากค้นคืนเอกสารภาษาไทยดีกว่าค้นคืนเอกสารภาษาอังกฤษ อาจจะเนื่องจากการชุดข้อมูลทดสอบ ที่สร้างขึ้นคัดเลือกเอกสารที่เกี่ยวข้องกันจากเอกสารภาษาไทยโดยอาสาสมัครคนไทย

ระบบค้นคืนสารสนเทศที่จัดทำขึ้นนี้สามารถทำงานได้ประสิทธิภาพดีสำหรับภาษาไทย และภาษาอังกฤษ และจะนำไปใช้เป็นส่วนประกอบสำคัญในการศึกษาการค้นคืนข้ามภาษาในบทต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การแปลคำสั่งสอบถาม

บทนี้จะกล่าวถึงกระบวนการแปลคำสั่งสอบถามโดยใช้ทรัพยากรการแปล (Translation Resource) ชนิดต่างๆ ได้แก่ การแปลด้วยคลังข้อมูลภาษาแบบขนาน (Parallel Corpus) การแปลด้วยพจนานุกรม และการแปลด้วยระบบแปลภาษา (MT) โดยเราจะทำการทดลองเปรียบเทียบประสิทธิภาพการแปลคำสั่งสอบถาม โดยใช้ชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทยปี 2528-2538 จากการเตรียมการในบทที่ 2

4.1. การแปลคำสั่งสอบถามโดยใช้คอร์ปัสแบบขนาน

คลังข้อมูลภาษาแบบขนานที่ได้จากฐานข้อมูลวิทยานิพนธ์ไทยเราแบ่งออกเป็นสองส่วน ส่วนแรกคือบทคัดย่อของปี พ.ศ. 2528-2538 จำนวน 7,341 บทคัดย่อ นำไปสร้างชุดข้อมูลทดสอบ ส่วนที่สองคือบทคัดย่อของปี พ.ศ. 2528-2543 จำนวน 34,225 บทคัดย่อ นำมาใช้ในการแปลคำสั่งสอบถาม

จากการนำเสนอในงานประชุมวิชาการ TREC และ CLEF พบว่าสามารถใช้วิธี Pseudo-Relevance Feedback หรือ PRF ในการแปลคำสั่งสอบถามได้ผลเป็นที่น่าพอใจ [21] เราจะดัดแปลงวิธีการ PRF มาใช้ในการแปลคำสั่งสอบถามในการค้นคืนข้ามภาษา ภาษาไทย-ภาษาอังกฤษ

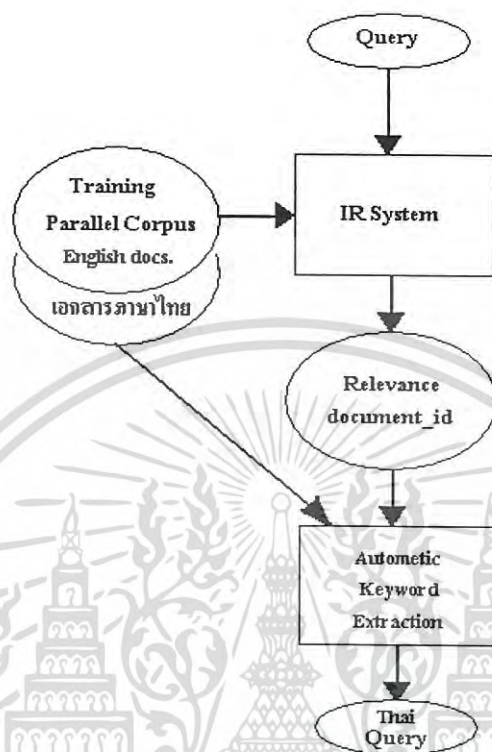
4.1.1. การแปลคำสั่งสอบถามโดยวิธีการ PRF

วิธีการป้อนกลับ (Feedback) ของการค้นคืนสารสนเทศภาษาเดียว จะเป็นการเลือกคำในเอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม (Relevance Document) นำมารวมกับคำสั่งสอบถามเดิมหรือสร้างเป็นคำสั่งสอบถามใหม่ แล้วทำการหาเอกสารที่เกี่ยวข้องจากการค้นคืนครั้งใหม่ ถ้าหากการป้อนกลับเอกสารที่คัดเลือกโดยผู้ใช้ จะเรียกว่า Relevance Feedback (RF) และถ้าเอกสารที่ป้อนกลับมาจากการเลือกโดยอัตโนมัติ จะเรียกว่า Pseudo-Relevance Feedback (PRF)

การใช้วิธีการ PRF เพื่อแปลคำสั่งสอบถาม จะต้องใช้เอกสารที่เป็นคลังข้อมูลภาษาแบบขนานเพื่อค้นคืนเอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม จากนั้นป้อนกลับเอกสารเดียวกันในภาษาที่ต้องการคำแปล เข้าสู่กระบวนการการเลือกคำในเอกสาร กลับมาสร้างเป็น คำสั่งสอบถามใหม่ เราจะเรียกคลังข้อมูลภาษาแบบขนานที่ใช้แปลคำสั่งสอบถามนี้ว่าเป็น Training Corpus

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างที่ 1. การแปลคำสั่งสอบถาม "cat and mouse" ด้วยวิธีการ PRF



รูปที่ 4.3 ขั้นตอนการแปลคำสั่งสอบถาม ด้วย PRF

จากรูปที่ 4.3

1. ป้อนคำสั่งสอบถาม "cat and mouse" เข้าสู่ระบบค้นคืนสารสนเทศให้ค้นคืนเอกสารจาก Training Corpus
2. เลือกเอกสารที่เกี่ยวข้องกับคำสั่งสอบถามได้แก่ Doc1 และ Doc3
3. ทำการแยกคำ (Term Extraction) ในเอกสารภาษาไทย มาสร้างเป็นคำสั่งสอบถามในภาษาไทย

Doc1	cat cat mouse mouse	แมว แมว หนู หนู
Doc3	cat mouse	แมว หนู

คำสั่งสอบถามภาษาไทยคือ "แมว หนู"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างที่ 2. การแปลคำสั่งสอบถาม ด้วยวิธีการ PRF โดยกำหนดให้การป้อนกลับจากเอกสาร 10 หัวข้อ ที่เกี่ยวข้องกับคำสั่งสอบถามมากที่สุด

“อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง เป็นอย่างไร”

1. ป้อนคำสั่งสอบถามเข้าสู่ระบบค้นคืนสารสนเทศให้ค้นคืนเอกสารจาก Training Corpus
2. เลือกเอกสารที่เกี่ยวข้องกันกับคำสั่งสอบถามได้แก่เอกสารหมายเลข 25492, 49422, 20699, 23007, 21897, 26945, 21935, 22610, 25763 และ 4055
3. ทำการแยกคำในเอกสารภาษาอังกฤษ นำมาสร้างเป็นคำสั่งสอบถามภาษาอังกฤษ

(STEP_1)

Query IR system for query : อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง เป็นอย่างไร... Four

RANK#	doc_id#	TITLE	[score#]
1	25492	ผลของไนโตรเจนเสริมที่เพิ่มหลังระยะติดฝักต่อผลผลิต และลักษณะทางสรีรวิทยาบางประการของถั่วเหลือง (37-18)	
2	49422	EFFECT OF SUPPLEMENTAL NITROGEN AFTER POD-SETTING ON YIELDS AND SOME PHYSIOLOGICAL	
3	20699	อิทธิพลของปุ๋ยไนโตรเจนและอัตราปุ๋ยไนโตรเจนต่อการเจริญเติบโต และผลผลิตของถั่วเหลืองที่ปลูกในดินจิมตีด้วยน้ำ (23-43)	
4	23007	Effects of Nitrogen fertilizer and Transplanting Spacings on Yield of Rice (Oryza Sativa L.)	
5	21897	อิทธิพลของพันธุ์และอัตราปุ๋ยไนโตรเจนต่อการเจริญเติบโต และผลผลิตของถั่วเหลืองที่ปลูกในดินจิมตีด้วยน้ำ (23-43)	
6	26942	Effects of Genotypes and Rates of Nitrogen Fertilizer on Growth and Yield of Soybean Grown under Saturat	
7	21935	อิทธิพลของพันธุ์และอัตราปุ๋ยไนโตรเจนต่อการเจริญเติบโตและผลผลิตของถั่วเหลืองที่ปลูกในดินจิมตีด้วยน้ำ (23-43)	
8	22610	Effects of Genotypes and Rates of Nitrogen Fertilizer on Growth and Yield of Soybean Grown under Saturat	
9	25763	อิทธิพลของระยะปลูกต่อผลผลิตและคุณภาพของถั่วเหลืองที่ผลิตพันธุ์ต่าง ๆ (Glycine max [L.] Merr.) (23-43)	
10	4055	Effect of Planting Dates on Fresh Yield and Quality of Vegetable Soybeans. (Glycine max [L.] Merr.)	

รูปที่ 4.4 เอกสารจำนวน 10 เอกสารที่เกี่ยวข้องกับคำสั่งสอบถาม

ค่าที่แยกได้หาค่าถ่วงน้ำหนักจาก

ค่าถ่วงน้ำหนัก = N * IDF

N คือจำนวนเอกสารที่มีคำนี้ปรากฏ (จาก 10 เอกสารที่เลือกมา)

IDF นำมาจากระบบค้นคืนสารสนเทศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อันดับที่	คำ (Term)	IDF	N	IDF * N
1	YIELD	7.74	9	696.69
2	FERTILIZER	9.16	6	549.36
3	SOYBEAN	8.06	6	483.78
4	NITROGEN	8.64	5	431.8
5	GROWN	10.54	4	421.56
6	EFFECT	3.68	10	367.9
7	GROWTH	6.76	5	337.95
8	SOIL	7.18	4	287.2
9	POD	12.48	2	249.56
10	SATURATED	12.26	2	245.1
11	SEED	7.83	3	235.02
12	PLANTING	10.02	2	200.36
13	PHYSIOLOGICAL	9.89	2	197.86
14	VEGETABLE	9.09	2	181.7
15	QUALITY	5.3	3	159.09
16	TOPPING	15.06	1	150.63
17	QUALITY	15.06	1	150.63
18	INTRAPLANT	15.06	1	150.63
19	KPS85	15.06	1	150.63
20	GENOLYPE	15.06	1	150.63
21	RATE	7.36	2	147.1
22	COMMUNIS	13.48	1	134.78
23	RICINUS	13.48	1	134.78
24	TRANSPLANTING	13.48	1	134.78

รูปที่ 4.5 คำที่มีค่าน้ำหนักสูงสุด จากตัวอย่างการแปลคำสังสอบถาม

เลือกคำจำนวน 10 คำที่มีน้ำหนักมากที่สุด มารวมเป็นคำสังสอบถาม ใหม่

“YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2. การแปลคำสั่งสอบถาม ด้วยพจนานุกรม

จากผลการทดสอบระบบค้นคืนข้ามภาษาในงานประชุมวิชาการ CLEF พบว่า[21] การแปลคำสั่งสอบถามด้วยพจนานุกรมให้ผลการแปลที่ดีโดยการใช้พจนานุกรมที่มีความครอบคลุม ผู้วิจัยจึงได้รวบรวมคำแปล จากพจนานุกรม อังกฤษไทย ฉบับต่างๆ ที่ผลิตเป็นซอฟต์แวร์ ได้แก่ LEXITRON จาก NETEC POPular Diction ฉบับของ อ.เรียรชัย เขียมวรเมธ และ สอเสถบุตร โดยบริษัทไทยซอฟต์แวร์อินเทอร์เน็ตไพรซ์

4.2.1. ปัญหาของการแปลคำสั่งสอบถาม ด้วยพจนานุกรม

เดิมผู้วิจัยได้ทำการทดลองในเชิงเปรียบเทียบประสิทธิภาพการแปลคำสั่งสอบถาม โดยพจนานุกรมฉบับต่างๆ ซึ่งก็ให้ผลการแปล คุณภาพต่ำ อันเนื่องมาจาก

4.2.1.1. คำศัพท์ไม่ครอบคลุม

- LEXITRON มีประมาณ 7000 คำศัพท์ 12,000 คำแปล
- ThaiSoftware Dictionary ฉบับ สอ เสถบุตร มีประมาณ 46,800 คำศัพท์ 75,000 คำแปล
- POPular Dictionary ฉบับ เรียรชัย เขียมวรเมธ มีประมาณ 21,500 คำศัพท์ 88,000 คำแปล

คำศัพท์ที่เราใช้เป็นคำสำคัญ มักจะเป็นคำทับศัพท์ หรือศัพท์เฉพาะทาง ซึ่งไม่มีใน

พจนานุกรม 4.2.1.2. คำศัพท์มี หลายความหมาย โดยวิธีการแปลแบบคำต่อคำ ไม่มีข้อมูลใดที่จะช่วยในการเลือกใช้ความหมาย

4.2.2. การแก้ปัญหาการแปลคำสั่งสอบถาม ด้วยพจนานุกรม

ผู้วิจัย ได้หาวิธีการแก้ปัญหาข้างต้น โดย

4.2.2.1. รวมพจนานุกรมทั้ง 3 ฉบับเข้าด้วยกัน เพื่อแก้ปัญหาคำศัพท์ไม่ครอบคลุม ทำให้ได้เป็นพจนานุกรมที่มีจำนวน 48,812 คำศัพท์ 150,499 คำแปล แต่ปัญหาที่ตามมาก็คือ ความหมายของคำศัพท์มีมากขึ้น

4.2.2.2. การเลือกความหมายศัพท์ โดยความหมายที่ใช้อย่างกว้างขวาง ใช้หลักคะแนนเสียงข้างมากในการเลือกใช้ความหมาย โดยให้ พจนานุกรมทั้ง 3 ฉบับมีฉบับละ 1 คะแนนเสียง ความหมายใด ได้คะแนนมากที่สุด จะเลือกความหมายนั้น ถ้าหากคะแนนเท่ากัน จะเลือกจากความหมายของคำแปลในการแปลทิศทางการกลับกัน ผู้วิจัยตั้งสมมุติฐานว่า ความหมายศัพท์ที่ใช้อย่างกว้างขวาง มีโอกาสเป็นคำแปลที่ถูกต้องมากกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการแปลคำว่า "Translate"

การอ้างอิงแหล่งที่มาของคำแปลเราจะใช้ตัวย่อต่อไปนี้ในโปรแกรม

SS = สอ เสถบุตร (So Setapura)

POP = ธิยรรชัฒย ธิยรรวมรเมธ (POPular Dictionary)

LEX = เลกซัทรอน (LEXiTRON)

translate				
Word by Word Translation				
Term	Vote	Meaning	Source	
<u>แปล</u>	3	9	pop-ss-lex	translate
<u>เปลี่ยนแปล</u>	1	28	pop	translate
<u>อธิบาย</u>	1	21	pop	translate
<u>ผา</u>	1	9	-ss	translate
<u>แปลง</u>	1	8	-ss	translate
<u>ถอดความ</u>	1	2	pop	translate
<u>เปลี่ยนหลัก</u>	1	2	pop	translate

รูปที่ 4.7 ตัวอย่างการแปลศัพท์คำเดียว

จาก รูปที่ 4.7 เป็นการแปลศัพท์คำเดียว ซึ่งให้ความหมาย 7 ความหมาย โดยที่เราเลือกความหมายที่เป็นคำแปลของทั้ง 3 พจนานุกรม

<u>เสนอ</u>	2	44	pop-ss	present
<u>ปัจจุบัน</u>	2	8	pop-lex	present
<u>ให้</u>	1	39	pop	present
<u>มี</u>	1	36	-ss	present
<u>ยื่น</u>	1	24	pop	present
<u>แนะนำ</u>	1	22	pop	present
<u>ปรากฏ</u>	1	18	pop	present
<u>ทำทาง</u>	1	18	-ss	present
<u>เข้าร่วม</u>	1	14	pop	present
<u>แสดงออก</u>	1	10	pop	present
<u>เลี้ยง</u>	1	8	pop	present
<u>ขณะนี้</u>	1	8	pop	present
<u>ต่อหน้า</u>	1	6	-ss	present
<u>เลี้ยง (ป็น)</u>	1	5	-ss	present
<u>เข้ามา</u>	1	4	pop	present
<u>ของขวัญ</u>	1	4	pop	present

รูปที่ 4.8 ตัวอย่างการแปลศัพท์คำเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.8 ศัพท์คำเดียวมีหลายความหมาย และมีสองความหมายที่ได้คะแนนเสียงเท่ากัน จึงเลือกใช้ทั้งสองความหมาย คำแปลของ Present คือ "เสนอ" และ "ปัจจุบัน"

4.2.3. การทำ Stemming ศัพท์ภาษาอังกฤษ

กรณีที่ไม่สามารถค้นหาความหมายได้ เราจะทดลองค้นหาความหมายของศัพท์ที่ตัดส่วนเติมท้าย (Suffix) ได้แก่ -y -ies -s -es -ly -ing ดังตัวอย่าง

<u>ความสงสัย</u>	1	9	pop	query
<u>ซักถาม</u>	1	8	-ss	query
<u>เครื่องหมายคำถาม</u>	1	4	pop	query
<u>ตั้งกระทู้ถาม</u>	1	3	pop	query
<u>การตั้งกระทู้ถาม</u>	1	3	pop	query
<u>ตั้งข้อสงสัย</u>	1	2	pop	query
<u>เสนอข้อสงสัย</u>	1	1	pop	query
<u>สอบถามข้อสงสัย</u>	1	1	pop	query

รูปที่ 4.9 ตัวอย่างการศัพท์ที่ไม่พบในพจนานุกรม ด้วยการตัดที่ตัดส่วนเติมท้าย (Suffix)

คำที่ไม่มีความสำคัญในการค้นหา หรือที่เรียกว่าคำหยุด (Stopword) ซึ่งได้แก่คำเชื่อม คำนำหน้า จะถูกละทิ้ง (**รายการคำหยุด ดูได้จากภาคผนวก จ.) ตัวอย่างการแปลคำสั่งสอบในรูป 4.10 คำว่า IN จะถูกละทิ้ง

Term	Vote	Meaning	Sources
<u>ลายนิ้วมือที่นิ่มด้วยหมึก</u>	1	1	-ss fingerprint
<u>คน</u>	3	52	pop-ss-lex man
<u>ผู้ชาย</u>	3	5	pop-ss-lex man
<u>ชาย</u>	2	8	pop-lex man
<u>สามี</u>	2	7	pop-ss man
<u>มนุษย์</u>	1	14	-ss man
<u>ทหาร</u>	1	12	pop man
<u>พลทหาร</u>	1	12	-ss man

Original Query : DNA FINGERPRINTS IN MAN

Translated Query : [DNA] ลายนิ้วมือที่นิ่มด้วยหมึก คน ผู้ชาย ชาย สามี

รูปที่ 4.10 ตัวอย่างการแปล คำต่อคำ (Word by word) โดยละทิ้งคำหยุด

ส่วนคำที่ไม่พบในพจนานุกรม เราจะใส่ เครื่องหมาย [] ครอบไว้ ดังเช่นคำว่า DNA จากรูปที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3. การแปลคำสั่งสอบถาม ด้วยระบบแปลภาษา

จากผลการทดสอบประสิทธิภาพการแปล Query ด้วยระบบแปลภาษา ในงานประชุมวิชาการ TREC และ CLEF พบว่า หลายครั้งการใช้ MT ที่ผลิตเป็นการค้ากับ CLIR ให้ประสิทธิภาพการค้นคืนที่ดีกว่า PRF หรือ พจนานุกรม [22]

ระบบแปลภาษาที่เรานำมาใช้ในการทดลองคือ ภาษิต ซึ่งเป็นผลงานวิจัยของ NECTEC ร่วมกับ บริษัท NEC ประเทศญี่ปุ่น [23] ซึ่งใช้วิธีการแปลเป็นประโยค โดยสร้างภาษากลางก่อน แล้วจึงนำไปสร้างเป็นภาษาเป้าหมาย



รูปที่ 4.11 ตัวอย่างการแปลคำสั่งสอบถาม ด้วย "ภาษิต"

เนื่องจาก ภาษิต ยังอยู่ระหว่างการพัฒนา จึงทำการแปลจากภาษาอังกฤษเป็นภาษาไทยได้เท่านั้น การทดสอบจึงทำการทดสอบเฉพาะการค้นคืนข้ามภาษา ภาษาอังกฤษเป็นภาษาไทย

4.4. การทดลองแปลคำสั่งสอบถาม

เราจะทำการทดลองวัดผลการค้นคืนข้ามภาษาภาษาไทยเป็นภาษาอังกฤษและ ภาษาอังกฤษเป็นภาษาไทย โดยใช้ชุดข้อมูลทดสอบฐานข้อมูลวิทยานิพนธ์ไทยปี 2528-2535

4.4.1. การแปลคำสั่งสอบถามภาษาไทยเป็นภาษาอังกฤษ

- การแปลคำสั่งสอบถามด้วยพจนานุกรม
- การแปลคำสั่งสอบถามด้วย Parallel Corpus

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2. การแปลคำสั่งสอบถามภาษาอังกฤษเป็นภาษาไทย

- การแปลคำสั่งสอบถามด้วยพจนานุกรม
- การแปลคำสั่งสอบถามด้วย Parallel Corpus
- การแปลคำสั่งสอบถามด้วยระบบแปลภาษา

เนื่องจากชุดข้อมูลทดสอบที่ใช้สร้างจาก Parallel Corpus จึงมีคำสั่งสอบที่แปลด้วยคนทุก คำสั่งสอบเราจึงอาศัยคำสั่งสอบนี้เป็นค่า Precision อ้างอิง นอกเหนือไปจากการที่เรายังศึกษาการแปลคำสั่งสอบด้วยพจนานุกรมเพื่ออ้างอิงเป็น Base line

4.5. ผลการทดลองแปลคำสั่งสอบถาม

ผลการแปล Query1 - Query10 จะรวมอยู่ในผลการทดลองบทที่ 5 โดยขอยกผลการแปล Query1 ไว้ ณ ที่นี้เพื่อเป็นตัวอย่าง

Query1	[2303]
<i>Source Query</i>	
Thai	ความรู้ ทัศนคติ และการปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์
English	KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION
<i>Thai to English Query Translation</i>	
Corpus	AID KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN BEHAVIOR
Dictionary	KNOWLEDGE INFORMATION ATTITUDE PRACTICE ABOUT OF PERTAIN PROPHYLAXIS [เอดส์]
<i>English to Thai Query Translation</i>	
Corpus	เอดส์ ความรู้ เจตคติ ป้องกันโรค ทัศนคติ โรค
Dictionary	ความรู้ วิชาความรู้ ท่าที่ ทัศนคติ การปฏิบัติ ช่วยเหลือ ชัดขวาง
MT	ความรู้ทัศนคติในการป้องกันเอดส์ก็จวัตร

จากตัวอย่าง อาจไม่อาจบอกคุณภาพการแปลคำสั่งสอบได้ชัดเจน เนื่องจากผลการแปลคำสั่งสอบถาม มีไว้เพื่อระบบค้นคืนสารสนเทศจะสามารถค้นคืนเอกสารที่ต้องการได้อย่างแม่นยำ ในบทต่อไปเราจะวัดประสิทธิภาพการค้นคืนข้ามภาษาด้วยชุดข้อมูลทดสอบ เพื่อเปรียบเทียบประสิทธิภาพของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีการแปลคำสั่งสอบถาม ทั้งการแปลคำต่อคำด้วยพจนานุกรม การแปลด้วยคลังข้อมูลข้อความโดยวิธี PRF และการแปลด้วยระบบแปลภาษา



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การทดลองค้นคืนสารสนเทศข้ามภาษา

ในบทนี้เราจะทำการทดลองเพื่อพิสูจน์สมมุติฐานเรื่องการค้นคืนข้ามภาษา ภาษาไทยเป็นภาษาอังกฤษและภาษาอังกฤษเป็นภาษาไทย ว่าสามารถทำได้โดยใช้ทรัพยากรทางภาษาที่จัดเตรียมขึ้น นอกจากนี้ยังเป็นการทดสอบเปรียบเทียบการแปลคำสั่งสอบถามด้วยวิธีการที่ต่างกัน

เราจะใช้ชุดข้อมูลทดสอบจากวิทยานิพนธ์ไทยปี 2528-2538 ที่มีคำสั่งสอบถาม 10 เป็นภาษาไทยและอีก 10 คำสั่งสอบถามเป็นภาษาอังกฤษ คำสั่งสอบถามที่แปลเสร็จแล้วจากบทที่ 4 จะป้อนเข้าสู่ระบบค้นคืนสารสนเทศ โดยจะมีคำสั่งสอบถามที่แปลโดยอัตโนมัติจำนวน 9 คำสั่ง และคำสั่งสอบถามที่แปลโดยคน (แปลโดยผู้แต่งวิทยานิพนธ์ และดัดแปลงโดยผู้วิจัย) จำนวน 1 คำสั่ง สำหรับเปรียบเทียบ

5.1 การทดลองแปลคำสั่งสอบถาม

คำสั่งสอบถามที่แปลโดยอัตโนมัติจำนวน 9 คำสั่ง มีรายละเอียดดังนี้

5.1.1. คำสั่งสอบถามที่แปลโดยใช้คลังข้อมูลภาษาแบบขนาน จำนวน 7 คำสั่ง ที่มีจำนวนคำเป็น 0.6, 0.8, 1.0, 1.2, 1.4, 1.6 และ 1.8 เท่าของความยาวคำสั่งสอบถามต้นฉบับ เพื่อทดลองหาความยาวของคำแปล ที่ให้ประสิทธิภาพการค้นคืนได้ดีที่สุดไปพร้อมกัน

ตัวอย่างเช่นคำสั่งสอบถามต้นฉบับ "KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION" เมื่อตัดคำหยุด AN และ IN ออกไปจะมีความยาว 5 คำ เราแปลโดยวิธีการ PRF ได้เป็นคำสั่งสอบถามที่มีความยาวต่างกันดังนี้

Corpus:0.6: เอดส์ ความรู้ เจตคติ คือคำแปลที่มีความยาว 0.6 เท่าของของคำสั่งสอบถามต้นฉบับ

5.1.2. การแปลโดยพจนานุกรมโดยพจนานุกรมหลายฉบับ มีคำแปลทั้งจากไทยแปลเป็นอังกฤษและอังกฤษแปลเป็นไทย ตัวอย่างคำสั่งสอบถามที่แปลแล้ว เช่น

Dic :ความรู้ วิชาความรู้ ทำที่ ทศนคติ การปฏิบัติ ช่วยเหลือ ชัดขวาง

5.1.3. การแปลโดยใช้ระบบแปลภาษา ระบบ "ภษาชิต" ในขณะนี้สามารถแปลภาษาอังกฤษเป็นภาษาไทยเท่านั้น ตัวอย่างคำสั่งสอบถามที่แปลแล้ว เช่น

MT :ความรู้ทศนคติในการป้องกันเอดส์ก็จวัตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.4. คำสั่งสอบถามที่แปลด้วยคน เช่น

Thai: ความรู้ ทัศนคติ และการปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์

5.2 ทดสอบการค้นคืนสารสนเทศ (IR Evaluation)

การทดสอบหาค่าความแม่นยำ (Precision) ที่ระดับการเรียกคืน (Recall) 11 ค่าตั้งแต่ 0.0 - 1.0 โดยนิยามความแม่นยำ และการเรียกคืนดังนี้

$$\text{ความแม่นยำ} = \frac{\text{จำนวนเอกสารที่เกี่ยวข้องที่ค้นคืนได้}}{\text{จำนวนเอกสารทั้งหมดที่ค้นคืนได้}}$$

$$\text{การเรียกคืน} = \frac{\text{จำนวนเอกสารที่เกี่ยวข้องที่ค้นคืนได้}}{\text{จำนวนเอกสารทั้งหมดที่เกี่ยวข้อง}}$$

ค่าความแม่นยำจะถูกปรับ (Interpolation) โดยกฎของ Van Rijsbergen [24] ให้ค่าความแม่นยำที่ค่าการเรียกคืนน้อย มีค่ามากกว่าความแม่นยำที่ค่าการเรียกคืนมากกว่าเสมอ ดังตัวอย่างจากรูปที่ 5.1 ข้อมูลกราฟด้านขวาได้มาจากการปรับจุด A, B, C, D กราฟด้านซ้าย



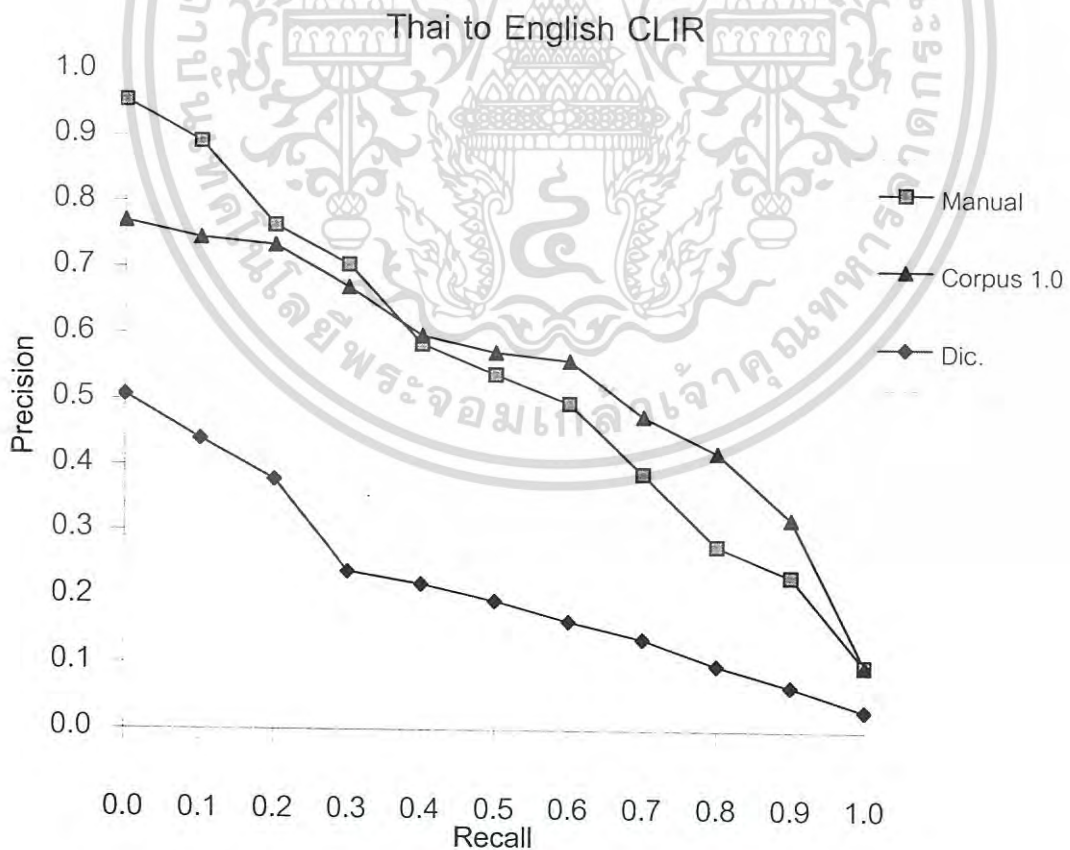
รูปที่ 5.1 ตัวอย่างการปรับกราฟ (Interpolation)

ข้อมูลที่ได้จากการทดลองค้นคืนคำสั่งสอบถามทั้ง 10 คำสั่งเราได้นำเสนอข้อมูลเป็นตารางแสดงค่าความแม่นยำที่ระดับการเรียกคืน 11 ค่าตั้งแต่ 0.0 - 1.0 และแสดงกราฟประสิทธิภาพการค้นคืน ภาษาไทยเป็นภาษาอังกฤษ และภาษาอังกฤษเป็นภาษาไทย อย่างละ 1 ภาพ เพื่อให้เส้นกราฟดูง่ายเราจึงแสดงเส้นกราฟตัวแทนของการแปลด้วยคลังข้อมูลภาษาแบบขนาน เพียง 1 เส้น คือเส้นกราฟคำสั่งสอบถามที่มีความยาวเท่าของความยาวคำสั่งสอบถามต้นฉบับ (Corpus:1.0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ

Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	0.9540	0.5071	0.7307	0.7635	0.7698	0.8178	0.8208	0.8065	0.8423
0.1	0.8917	0.4399	0.7187	0.7450	0.7450	0.7900	0.7886	0.7756	0.7933
0.2	0.7641	0.3780	0.6929	0.7114	0.7345	0.7566	0.7472	0.7365	0.7197
0.3	0.7048	0.2382	0.6826	0.6896	0.6705	0.7065	0.6917	0.6830	0.6463
0.4	0.5848	0.2195	0.5977	0.6076	0.5976	0.5889	0.5822	0.5755	0.5506
0.5	0.5395	0.1947	0.5680	0.5871	0.5729	0.5733	0.5774	0.5588	0.5315
0.6	0.4956	0.1641	0.5465	0.5620	0.5606	0.5643	0.5339	0.5210	0.5244
0.7	0.3882	0.1376	0.5049	0.5215	0.4759	0.4738	0.4601	0.4524	0.4244
0.8	0.2781	0.0984	0.4215	0.4352	0.4207	0.3979	0.4063	0.3987	0.3756
0.9	0.2321	0.0673	0.3259	0.3149	0.3199	0.3308	0.3121	0.3088	0.2907
1.0	0.0986	0.0316	0.1336	0.1199	0.0986	0.1654	0.1755	0.1561	0.1453
Average	0.5392	0.2251	0.5384	0.5507	0.5424	0.5605	0.5542	0.5430	0.5313

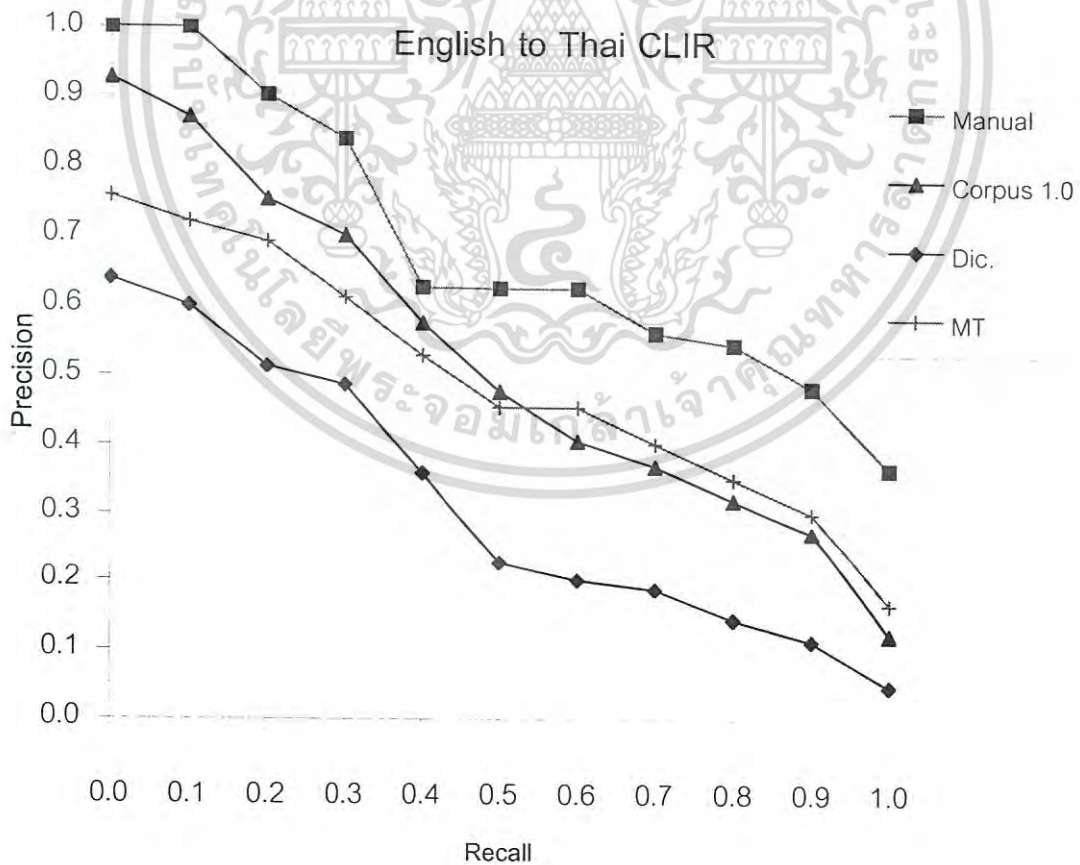


รูปที่ 5.1 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาไทยไปภาษาอังกฤษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.2 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	0.6385	0.8380	0.8538	0.9267	0.8695	0.9399	0.9336	0.8370	0.7564
0.1	1.0000	0.6007	0.8009	0.8148	0.8704	0.8307	0.9048	0.9012	0.8117	0.7204
0.2	0.9025	0.5137	0.7268	0.7176	0.7531	0.7513	0.8365	0.8523	0.7840	0.6927
0.3	0.8392	0.4872	0.6898	0.6979	0.7014	0.7143	0.7995	0.8042	0.7358	0.6126
0.4	0.6274	0.3599	0.5465	0.5785	0.5763	0.6403	0.6773	0.6460	0.6404	0.5296
0.5	0.6260	0.2260	0.4767	0.5235	0.4779	0.5020	0.5122	0.5119	0.5117	0.4559
0.6	0.6260	0.2006	0.4112	0.4127	0.4074	0.4544	0.4694	0.4557	0.4579	0.4558
0.7	0.5626	0.1868	0.3846	0.3639	0.3709	0.4177	0.4189	0.3961	0.3701	0.4035
0.8	0.5456	0.1431	0.3681	0.3078	0.3222	0.3743	0.3827	0.3655	0.3414	0.3527
0.9	0.4838	0.1129	0.2285	0.2473	0.2733	0.2922	0.2899	0.2800	0.2373	0.3032
1.0	0.3678	0.0480	0.1565	0.1420	0.1220	0.1409	0.1359	0.1342	0.1263	0.1648
Average	0.6892	0.3197	0.5116	0.5145	0.5274	0.5443	0.5788	0.5710	0.5321	0.4952



รูปที่ 5.2 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาอังกฤษไปภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์โดยศูนย์วิจัยเทคโนโลยีสารสนเทศและระบบสารสนเทศ การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:1

Source Query

Thai ความรู้ ทักษะ และ การปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์
 English KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION

Thai to English Query Translation

Corpus:0.6 AIDS KNOWLEDGE ATTITUDE PRACTICE
 Corpus:0.8 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION
 Corpus:1.0 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN
 Corpus:1.2 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN BEHAVIOR PREVENTIVE
 Corpus:1.4 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN BEHAVIOR PREVENTIVE
 SAHAWITHAYAKATAKRUNG
 Corpus:1.6 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN BEHAVIOR PREVENTIVE
 SAHAWITHAYAKATAKRUNG RIRST
 Corpus:1.8 AIDS KNOWLEDGE ATTITUDE PRACTICE PREVENTION MEAN BEHAVIOR PREVENTIVE
 SAHAWITHAYAKATAKRUNG RIRST PRATIVE
 Dictionary KNOWLEDGE INFORMATION ATTITUDE PRACTICE ABOUT OF PERTAIN PROPHYLAXIS
 [เอดส์]

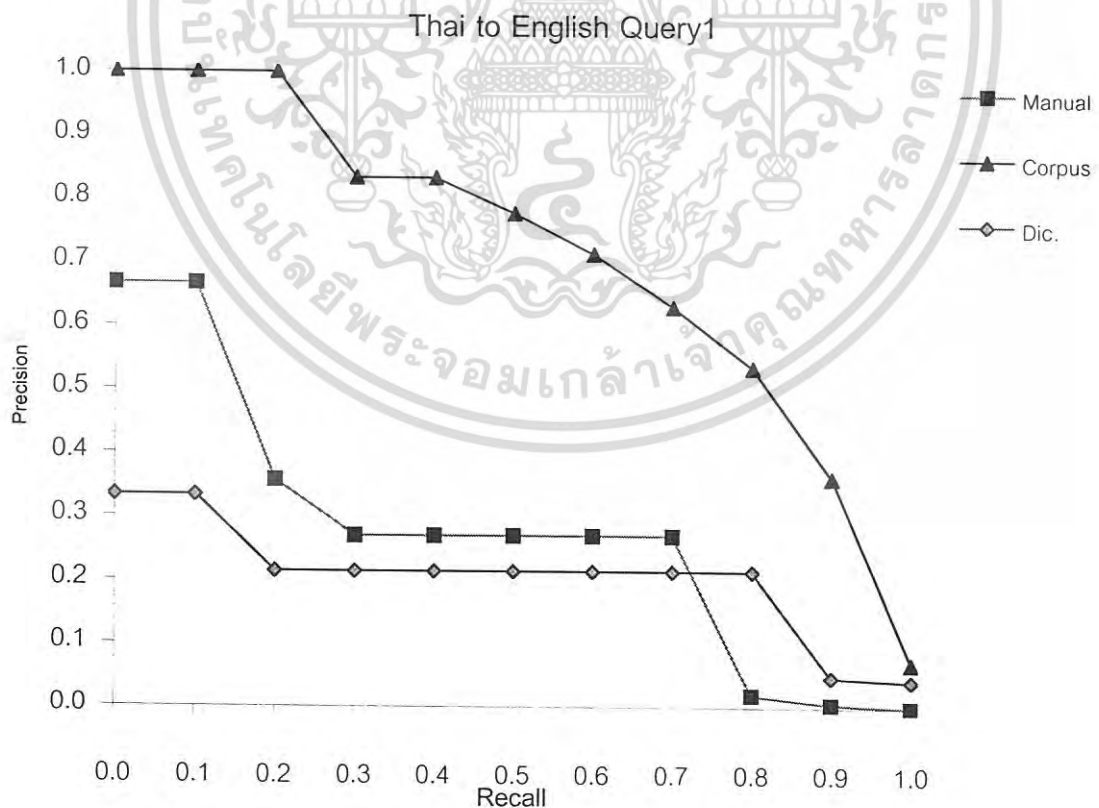
English to Thai Query Translation

Corpus:0.6 เอดส์ ความรู้ เจตคติ
 Corpus:0.8 เอดส์ ความรู้ เจตคติ ป้องกันโรค
 Corpus:1.0 เอดส์ ความรู้ เจตคติ ป้องกันโรค ทักษะ
 Corpus:1.2 เอดส์ ความรู้ เจตคติ ป้องกันโรค ทักษะ โรค
 Corpus:1.4 เอดส์ ความรู้ เจตคติ ป้องกันโรค ทักษะ โรค ปฏิบัติ
 Corpus:1.6 เอดส์ ความรู้ เจตคติ ป้องกันโรค ทักษะ โรค ปฏิบัติ นักศึกษา
 Corpus:1.8 เอดส์ ความรู้ เจตคติ ป้องกันโรค ทักษะ โรค ปฏิบัติ นักศึกษา ความสัมพันธ์
 Dictionary ความรู้ วิชาความรู้ ทักษะ ทักษะ การปฏิบัติ ช่วยเหลือ ชัดขวาง
 MT ความรู้ ทักษะ ในการป้องกันเอดส์ กิจวัตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.3 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query1

Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	0.6667	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	0.6667	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.3571	0.2143	0.8333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	0.2698	0.2143	0.8333	0.8750	0.8333	0.9091	0.9091	0.9091	0.9091
0.4	0.2698	0.2143	0.8333	0.8750	0.8333	0.9091	0.9091	0.9091	0.9091
0.5	0.2698	0.2143	0.8000	0.8750	0.7778	0.9091	0.9091	0.9091	0.9091
0.6	0.2698	0.2143	0.8000	0.8000	0.7143	0.9091	0.9091	0.9091	0.9091
0.7	0.2698	0.2143	0.8000	0.8000	0.6316	0.7500	0.7500	0.8000	0.8000
0.8	0.0196	0.2143	0.4286	0.5172	0.5357	0.5357	0.5357	0.5357	0.5357
0.9	0.0051	0.0484	0.2656	0.3148	0.3617	0.3542	0.3469	0.3617	0.3617
1.0	0.0000	0.0424	0.0720	0.0714	0.0706	0.0657	0.0617	0.0585	0.0549

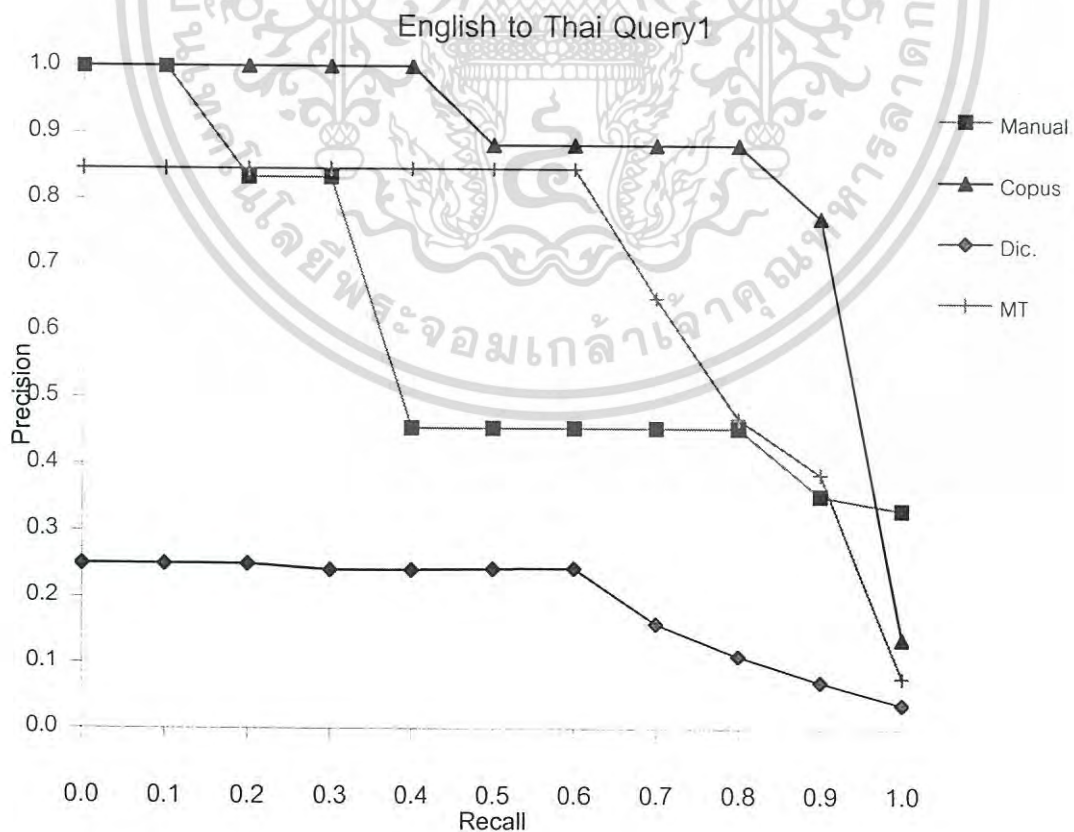


รูปที่ 5.3 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้ใดที่นำเอกสารนี้ไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารจะถือว่าผิดกฎหมาย และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.4 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query1

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	0.2500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8462
0.1	1.0000	0.2500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8462
0.2	0.8333	0.2500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8462
0.3	0.8333	0.2414	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8462
0.4	0.4545	0.2414	0.8571	1.0000	1.0000	1.0000	1.0000	0.8750	0.8750	0.8462
0.5	0.4545	0.2432	0.8571	0.8750	0.8824	0.8571	0.8750	0.8750	0.8750	0.8462
0.6	0.4545	0.2444	0.8571	0.8571	0.8824	0.8571	0.8571	0.8571	0.8571	0.8462
0.7	0.4545	0.1605	0.8571	0.8571	0.8824	0.8571	0.8571	0.8571	0.8571	0.6500
0.8	0.4545	0.1119	0.7895	0.7500	0.8824	0.8333	0.8500	0.8500	0.8333	0.4688
0.9	0.3542	0.0723	0.3333	0.5862	0.7727	0.8095	0.8500	0.8500	0.6538	0.3864
1.0	0.3333	0.0383	0.1570	0.1545	0.1387	0.1338	0.1727	0.1979	0.2021	0.0792



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 รูปที่ 5.4 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาอังกฤษไปภาษาไทย Query1
 ไม่สามารถได้ทั้งสั้น อีกทั้งห้ามเผยแพร่ลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:2

Thai	การใช้ประโยชน์และความพึงพอใจจากการเปิดรับรายการข่าวทางโทรทัศน์ ของประชาชนในกรุงเทพมหานคร
English	THE USE AN GRATIFICATION OF TELEVISION NEWS AMONG PEOPLE IN BANGKOK METROPOLITAN AREA

Thai to English Query Translation

Corpus:0.6	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING
Corpus:0.8	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE
Corpus:1.0	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE PROGRAM CHANNEL
Corpus:1.2	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE PROGRAM CHANNEL USE METROPOLIS
Corpus:1.4	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE PROGRAM CHANNEL USE METROPOLIS BUDDISHISM TODRAHAD
Corpus:1.6	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE PROGRAM CHANNEL USE METROPOLIS BUDDISHISM TODRAHAD POLITICAL HUMOUR
Corpus:1.8	GRATIFICATION EXPOSURE TELEVISION BANGKOK PEOPLE EVENING NEW AUDIENCE PROGRAM CHANNEL USE METROPOLIS BUDDISHISM TODRAHAD POLITICAL HUMOUR SHOW SATISFACTION
Dictionary	USE PLEASURE OPENING ACCEPT RECEIVE TAKE HOLD PROGRAM ITEM INFORMATION NEWS TIDINGS WAY TRAIL PATH TELEVISION TELLY POPULAR BANGKOK

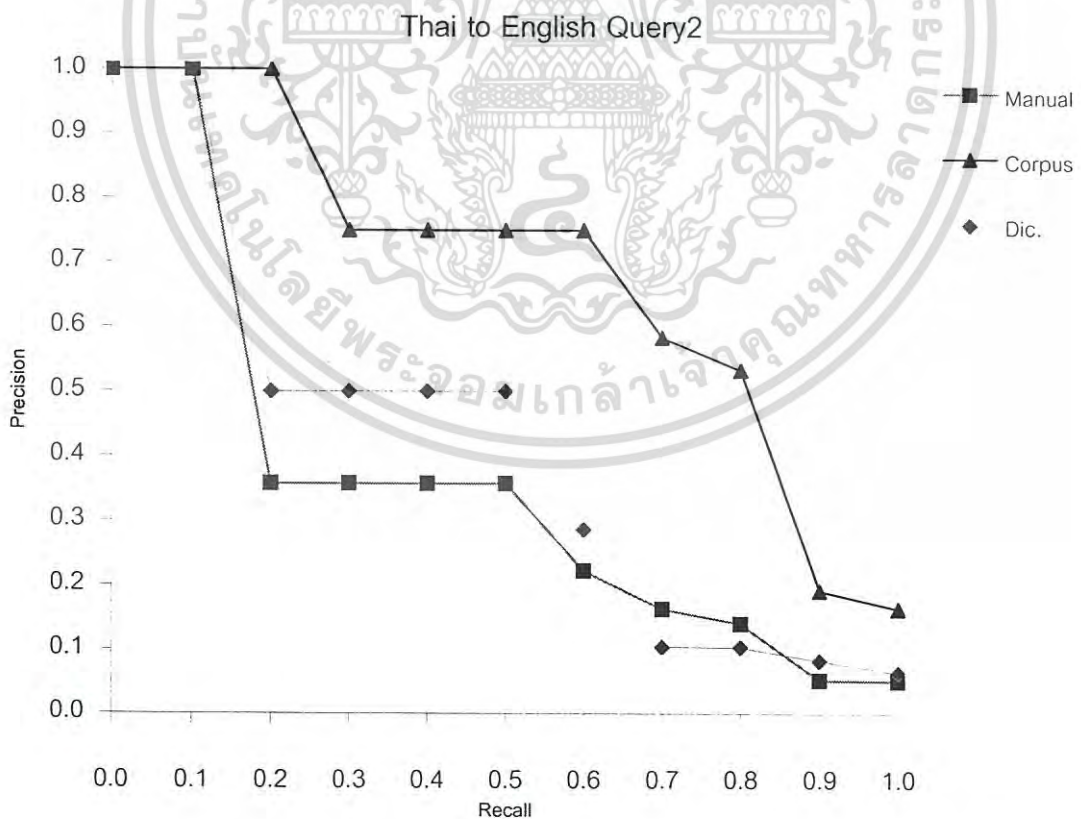
English to Thai Query Translation

Corpus:0.6	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว
Corpus:0.8	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด
Corpus:1.0	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด กองทัพบก
Corpus:1.2	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด กองทัพบก ทาง เกมโชว์
Corpus:1.4	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด กองทัพบก ทาง เกมโชว์ มีต่อ พฤติกรรม
Corpus:1.6	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด กองทัพบก ทาง เกมโชว์ มีต่อ พฤติกรรม ทศนคติ อารมณ์ขัน
Corpus:1.8	รายการ โทรทัศน์ เขตกรุงเทพ มหานคร พึงพอใจ ข่าว รับ เปิด กองทัพบก ทาง เกมโชว์ มีต่อ พฤติกรรม ทศนคติ อารมณ์ขัน ประชาชน ชั่วโมง
Dictionary	ใช้ ก่อให้เกิดความอึดใจ โทรทัศน์ ข่าว ข่าวคราว อยู่ในระหว่าง ประชาชน ชาว กรุงเทพฯ เกี่ยวกับนครหลวง เขต พื้นที่

MT ใช้ของข่าวโทรทัศน์ระหว่างผู้คนที่เขตมหานครกรุงเทพมหานครความพอใจ
 เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การเผยแพร่โดยไม่ได้รับอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.5 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query2

Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	1.0000	0.7143	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.3571	0.5000	0.7143	0.7500	1.0000	0.7143	0.7143	0.7143	0.6667
0.3	0.3571	0.5000	0.7143	0.7500	0.7500	0.7143	0.7143	0.7143	0.6667
0.4	0.3571	0.5000	0.7143	0.7500	0.7500	0.7143	0.7143	0.7143	0.6667
0.5	0.3571	0.5000	0.7143	0.7500	0.7500	0.7143	0.7143	0.7143	0.6250
0.6	0.2222	0.2857	0.6667	0.7500	0.7500	0.6667	0.5455	0.5000	0.6000
0.7	0.1628	0.1045	0.6364	0.7000	0.5833	0.5385	0.4375	0.3684	0.3684
0.8	0.1404	0.1039	0.4706	0.5714	0.5333	0.4211	0.3810	0.3478	0.3333
0.9	0.0526	0.0826	0.3103	0.2727	0.1915	0.1579	0.1538	0.1515	0.1500
1.0	0.051	0.0641	0.1852	0.1515	0.1639	0.1538	0.1538	0.1515	0.1471

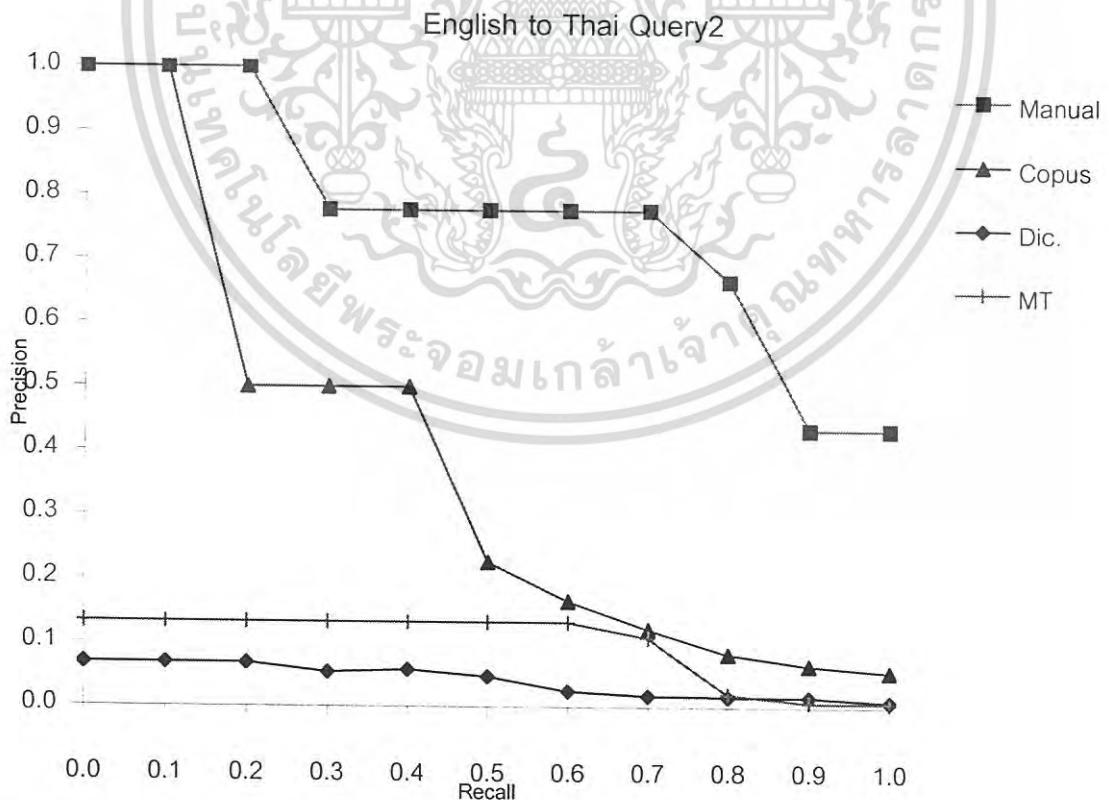


รูปที่ 5.5 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาไทยไปภาษาอังกฤษ Query2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.6 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query2

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	0.0690	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.1333
0.1	1.0000	0.0690	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.1333
0.2	1.0000	0.0690	0.3333	0.5000	0.4444	0.5000	0.6000	0.7500	0.7500	0.1333
0.3	0.7778	0.0545	0.3333	0.5000	0.4444	0.5000	0.6000	0.7500	0.7500	0.1333
0.4	0.7778	0.0588	0.2500	0.5000	0.4444	0.5000	0.5000	0.5000	0.5000	0.1333
0.5	0.7778	0.0481	0.2500	0.5000	0.3571	0.2273	0.1944	0.2188	0.2727	0.1333
0.6	0.7778	0.0249	0.0321	0.0549	0.0923	0.1667	0.1944	0.2188	0.2727	0.1333
0.7	0.7778	0.0179	0.0285	0.0549	0.0619	0.1228	0.1944	0.2188	0.2121	0.1094
0.8	0.6667	0.0166	0.0285	0.0549	0.0600	0.0842	0.1067	0.1194	0.1231	0.0204
0.9	0.4348	0.0164	0.0285	0.0549	0.0600	0.0667	0.0865	0.1000	0.1139	0.0074
1.0	0.4348	0.0090	0.0000	0.0348	0.0444	0.0565	0.0602	0.0613	0.0752	0.0070



เอกสารนี้เป็นรูปที่ 5.6 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query2 ระโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:3

Thai	ความสัมพันธ์ระหว่างความรู้ความเข้าใจ ทักษะการคิดคำนวณ และทักษะการแก้โจทย์ปัญหาในวิชาคณิตศาสตร์ ในระดับชั้น ประถมศึกษา
English	RELATIONSHIP AMONG UNDERSTAND COMPUTATIONAL AND WORD-PROBLEM SOLVE SKILL IN PRIMARY SCHOOL MATHEMATICS

Thai to English Query Translation

Corpus:0.6	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM
Corpus:0.8	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA
Corpus:1.0	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA DIFFICUTTY RELATIONSHIP
Corpus:1.2	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA DIFFICUTTY RELATIONSHIP SIX MATHAYOMSUKSA CALCULATING
Corpus:1.4	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA DIFFICUTTY RELATIONSHIP SIX MATHAYOMSUKSA CALCULATING ANALYTIC INTERGRATED ELEMENTARY
Corpus:1.6	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA DIFFICUTTY RELATIONSHIP SIX MATHAYOMSUKSA CALCULATING ANALYTIC INTERGRATED ELEMENTARY CURRICULUM ABILITY
Corpus:1.8	MATHEMATIC SOLVING SKILL PROBLEM PRATHOMSUKSA SUKSA STUDENT PRATHOM BETWEEN KATHUVITAYA SURANAREEVITHAYA DIFFICUTTY RELATIONSHIP SIX MATHAYOMSUKSA CALCULATING ANALYTIC INTERGRATED ELEMENTARY CURRICULUM ABILITY SCHOOL INSTRUCTIONAL HEURISTIC
Dictionary	RELATION BETWEEN KNOWLEDGE INFORMATION UNDERSTANDING SKILL ESTIMATE SKILL RELEASE PROBLEM PROBLEM QUESTION SUBJECT MATHEMATICS DEGREE ELEMENTARY EDUCATION

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

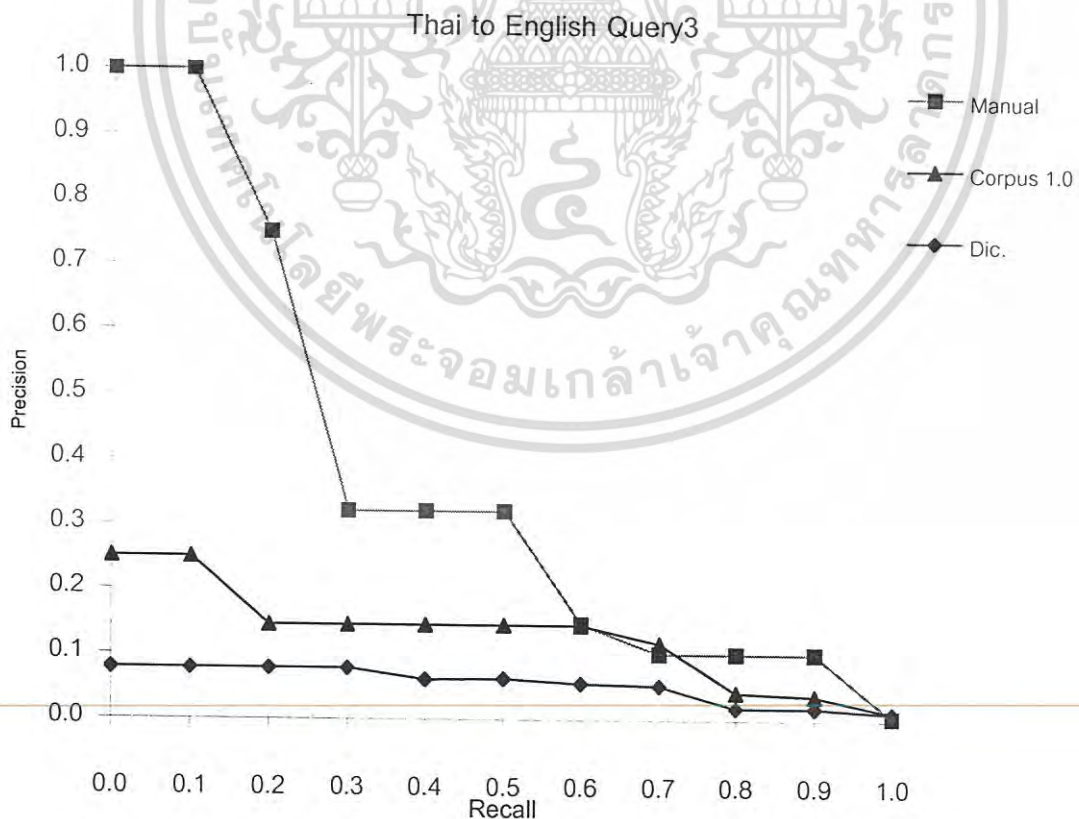
English to Thai Query Translation

- Corpus:0.6 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา
- Corpus:0.8 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์
- Corpus:1.0 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์ ชั้นประถม อ่าน
- Corpus:1.2 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์ ชั้นประถม อ่าน มัธยมศึกษา ชั้น
- Corpus:1.4 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์ ชั้นประถม อ่าน มัธยมศึกษา ชั้น ความสามารถ นูทุม
- Corpus:1.6 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์ ชั้นประถม อ่าน มัธยมศึกษา ชั้น ความสามารถ นูทุม ผลสัมฤทธิ์ ความสัมพันธ์
- Corpus:1.8 แก่โจทย์ นักเรียน เรียน ปีที่ คณิตศาสตร์ ปัญหา ทักษะ โจทย์ ชั้นประถม อ่าน มัธยมศึกษา ชั้น ความสามารถ นูทุม ผลสัมฤทธิ์ ความสัมพันธ์ ตอนปลาย หาเหตุผล
- Dictionary ความสัมพันธ์ อยู่ในระหว่าง เข้าใจ [COMPUTATIONAL] [WORD-PROBLEM] แก้ไข ความสามารถ ทักษะ ดั้งเดิม เบื้องต้น โรงเรียน คณิตศาสตร์
- MT ความสัมพันธ์ระหว่างเชิงคอมพิวเตอร์เข้าใจและทักษะแก้ไขปัญหาคำเกี่ยวกับคณิตศาสตร์โรงเรียนประถมศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.7 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query3

Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	1.0000	0.0781	0.2500	0.2500	0.2500	0.2000	0.2500	0.2000	0.5000
0.1	1.0000	0.0781	0.2500	0.2500	0.2500	0.2000	0.2500	0.2000	0.5000
0.2	0.7500	0.0781	0.1587	0.1636	0.1449	0.1515	0.1220	0.0943	0.0972
0.3	0.3214	0.0781	0.1587	0.1636	0.1449	0.1515	0.1220	0.0943	0.0972
0.4	0.3214	0.0614	0.1587	0.1636	0.1449	0.0645	0.0599	0.0533	0.0972
0.5	0.3214	0.0629	0.1587	0.1636	0.1449	0.0645	0.0599	0.0533	0.0481
0.6	0.1471	0.0559	0.1587	0.1370	0.1449	0.0645	0.0599	0.0506	0.0481
0.7	0.1019	0.0531	0.0656	0.0930	0.1176	0.0642	0.0594	0.0506	0.0418
0.8	0.1019	0.0191	0.0433	0.0369	0.0431	0.0507	0.0468	0.0411	0.0417
0.9	0.1019	0.0190	0.0189	0.0287	0.0381	0.0471	0.0437	0.0386	0.0353
1.0	0.0047	0.0113	0.0172	0.0114	0.0114	0.0154	0.0136	0.0125	0.0110

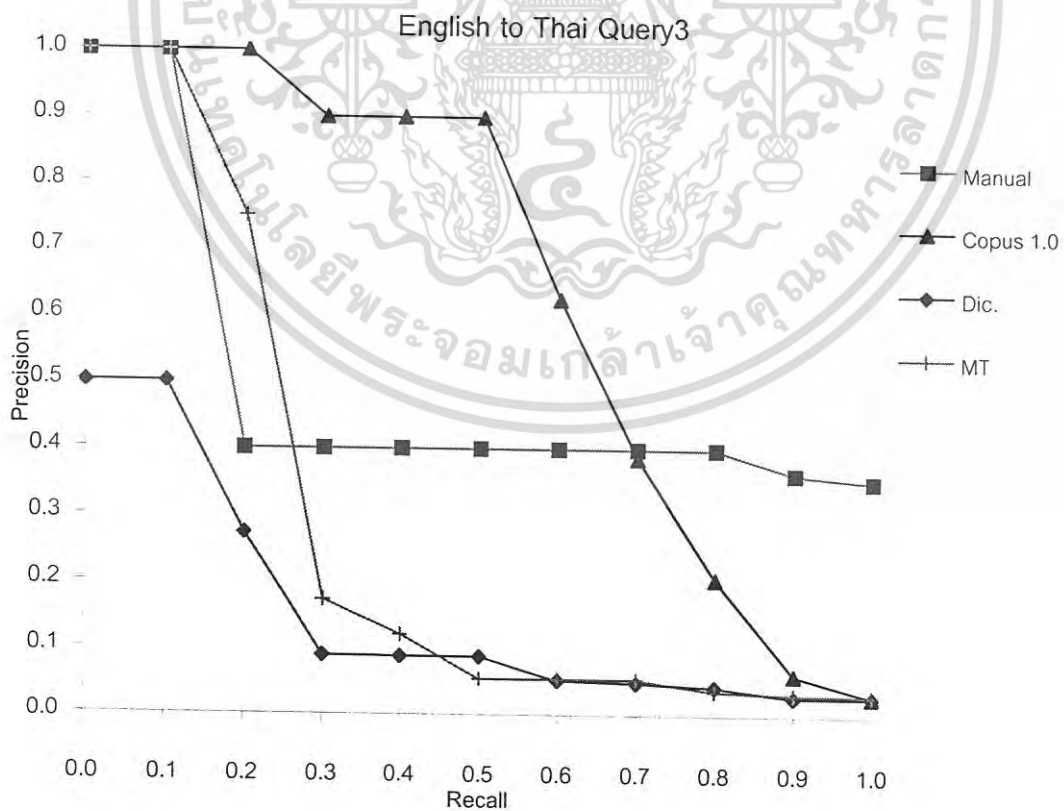


รูปที่ 5.7 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query3

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการค้นคืนข้ามภาษาจากภาษาไทยไปภาษาอังกฤษ Query3
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.8 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query3

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.4000	0.2727	0.8750	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7500
0.3	0.4000	0.0893	0.8750	0.9000	0.9000	1.0000	1.0000	0.9000	0.9000	0.1724
0.4	0.4000	0.0882	0.8750	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.1207
0.5	0.4000	0.0882	0.5625	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.0545
0.6	0.4000	0.0529	0.3333	0.5882	0.6250	0.5882	0.5882	0.5556	0.5263	0.0545
0.7	0.4000	0.0492	0.1101	0.3077	0.3871	0.3243	0.2791	0.2308	0.2069	0.0545
0.8	0.4000	0.0442	0.0838	0.1458	0.2059	0.1918	0.1818	0.1346	0.1228	0.0372
0.9	0.3636	0.0282	0.0678	0.0721	0.0627	0.0580	0.0532	0.0530	0.0537	0.0329
1.0	0.3529	0.0282	0.0387	0.0217	0.0291	0.0298	0.0299	0.0296	0.0293	0.0311



เอกสารนี้เป็นรูปที่ 5.8 ประสิทธิภาพการค้นคืนจากภาษาอังกฤษไปภาษาไทย Query3 ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:4

Thai	ทักษะการบริหารของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษา
English	ADMINISTRATIVE SKILL OF PRIMARY SCHOOL ADMINISTRATOR UNDER THE OFFICE OF PRIMARY EDUCATION

Thai to English Query Translation

Corpus:0.6	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER
Corpus:0.8	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL
Corpus:1.0	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL ADMINISTRATIVE
Corpus:1.2	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL ADMINISTRATIVE ADMINISTRATIONAL PROVINCIAL
Corpus:1.4	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL ADMINISTRATIVE ADMINISTRATIONAL PROVINCIAL CHACHEONGSAO
Corpus:1.6	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL ADMINISTRATIVE ADMINISTRATIONAL PROVINCIAL CHACHEONGSAO EDUCATION JURISDICTION
Corpus:1.8	ADMINISTRATOR PRIMARY OFFICE SKILL UNDER SCHOOL ADMINISTRATIVE ADMINISTRATIONAL PROVINCIAL CHACHEONGSAO EDUCATION JURISDICTION RATCHABURI
Dictionary	SKILL MANAGEMENT ADMINISTRATOR SCHOOL ON OFFICE ELEMENTARY EDUCATION

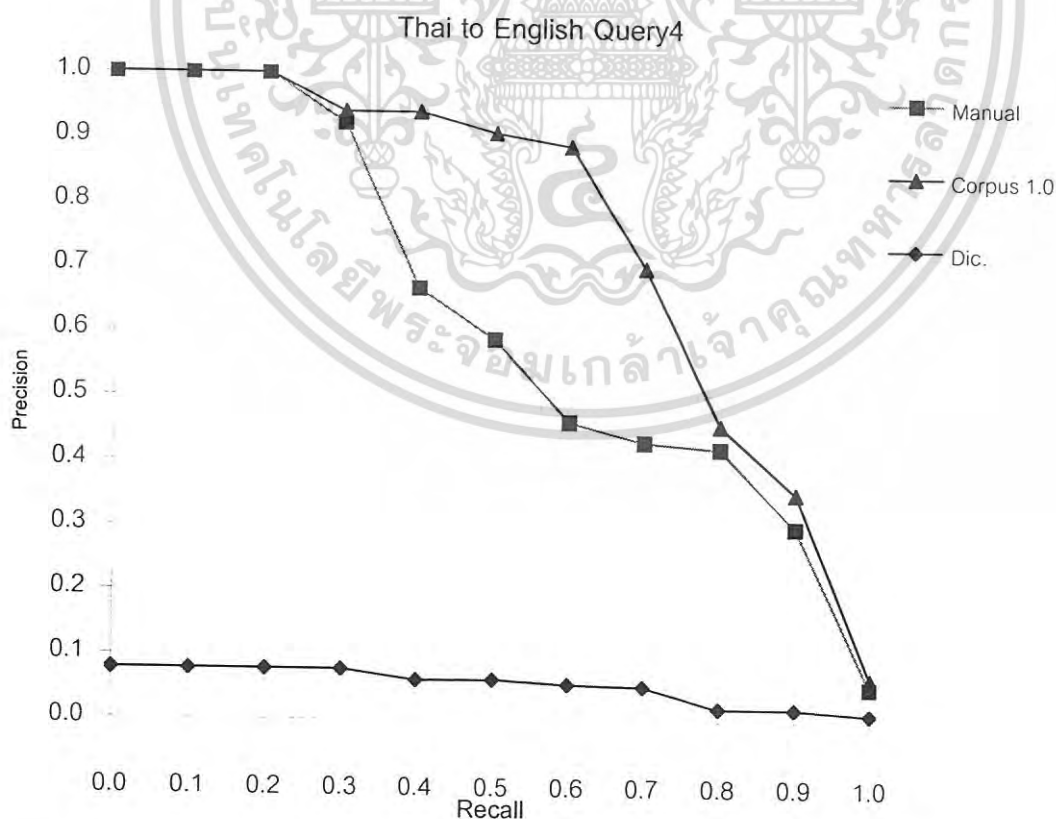
English to Thai Query Translation

Corpus:0.6	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด
Corpus:0.8	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน
Corpus:1.0	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน บริหาร
Corpus:1.2	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน บริหาร ราชบุรี นิเทศ
Corpus:1.4	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน บริหาร ราชบุรี นิเทศ ผลต่อ มั่นใจใน
Corpus:1.6	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน บริหาร ราชบุรี นิเทศ ผลต่อ มั่นใจใน ส่ง พฤติกรรม
Corpus:1.8	สำนักงาน ประถมศึกษา ผู้บริหาร สังกัด โรงเรียน จังหวัด วิชาการ บริหารงาน บริหาร ราชบุรี นิเทศ ผลต่อ มั่นใจใน ส่ง พฤติกรรม ทักษะ กระบวนการ
Dictionary	ควบคุม ความสามารถ ทักษะ ดั้งเดิม เบื้องต้น โรงเรียน ผู้บริหาร ได้ สำนักงาน หน้าที่ ห้องทำงาน ที่ทำงาน ดั้งเดิม เบื้องต้น การศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 MT ทักษะของผู้บริหารโรงเรียนประถมศึกษาข้างใต้สำนักงานของปฐมศึกษาที่ดูการ
 ไม่วารณใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.9 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query4

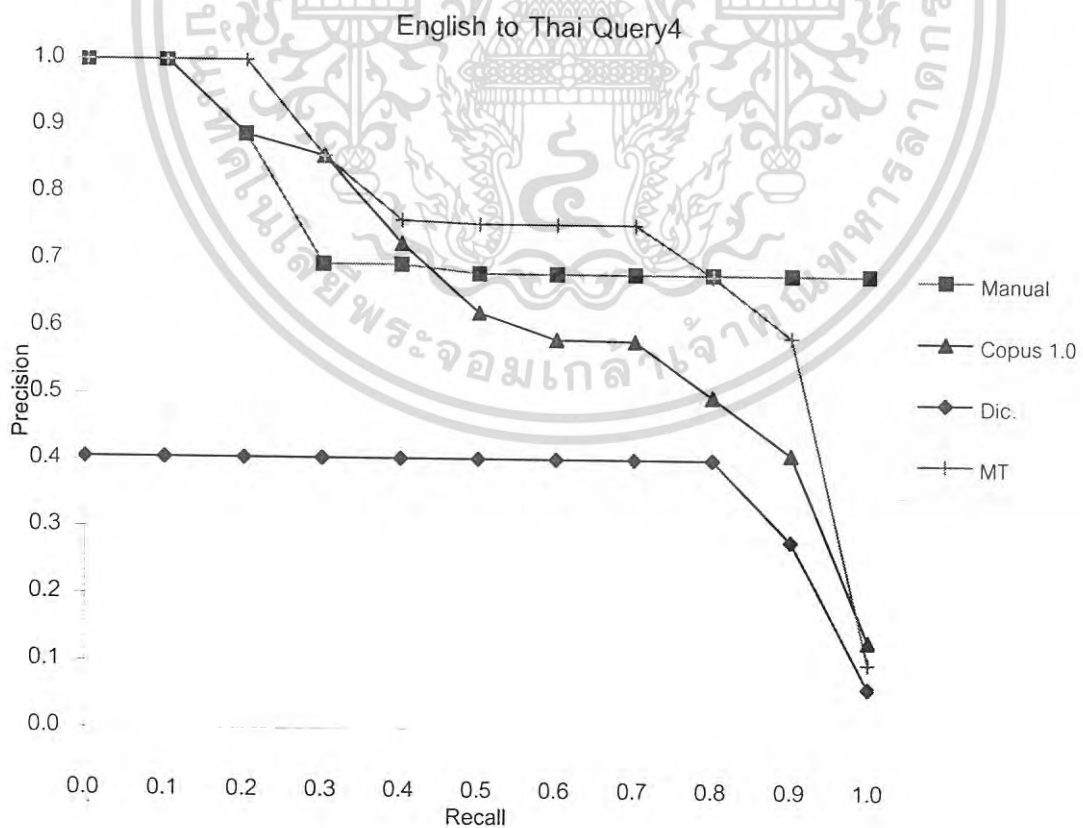
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	1.0000	0.0781	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	0.0781	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	1.0000	0.0781	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	0.9231	0.0781	1.0000	1.0000	0.9412	0.9231	0.8696	0.8889	0.8571
0.4	0.6667	0.0614	0.8889	0.9091	0.9412	0.8889	0.8696	0.8889	0.8000
0.5	0.5882	0.0629	0.8333	0.9091	0.9091	0.8696	0.8696	0.7692	0.7143
0.6	0.4615	0.0559	0.7059	0.8000	0.8889	0.8276	0.7059	0.6667	0.6857
0.7	0.4308	0.0531	0.7000	0.8000	0.7000	0.7000	0.6222	0.6087	0.5490
0.8	0.4211	0.0191	0.5079	0.4923	0.4571	0.4706	0.5000	0.4776	0.4706
0.9	0.3000	0.0190	0.1809	0.1915	0.3529	0.2647	0.2169	0.1915	0.1967
1.0	0.0524	0.0113	0.0555	0.0667	0.0665	0.0653	0.0650	0.0542	0.0535



เอกสารนี้เป็นรูปที่ 5.9 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query4
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.10 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query4

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	0.4051	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	0.4051	0.8750	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7500
0.2	0.8889	0.4051	0.8750	1.0000	0.8889	0.8571	0.8571	0.8571	0.8571	0.7500
0.3	0.6957	0.4051	0.8750	0.9231	0.8571	0.8571	0.8571	0.8571	0.8571	0.7500
0.4	0.6957	0.4051	0.8750	0.6667	0.7273	0.6957	0.6957	0.7273	0.7273	0.7619
0.5	0.6833	0.4051	0.8750	0.6667	0.6250	0.5833	0.5714	0.6061	0.5405	0.7568
0.6	0.6833	0.4051	0.8750	0.6667	0.5854	0.5833	0.5714	0.5455	0.5333	0.7568
0.7	0.6833	0.4051	0.8750	0.6087	0.5833	0.5833	0.5091	0.4746	0.2857	0.7568
0.8	0.6833	0.4051	0.8205	0.5614	0.5000	0.4444	0.4324	0.4444	0.2667	0.6809
0.9	0.6833	0.2835	0.2628	0.3243	0.4138	0.4186	0.4186	0.4186	0.2308	0.5902
1.0	0.6833	0.0657	0.1362	0.1314	0.1349	0.1273	0.1228	0.1336	0.0446	0.1012



เอกสารนี้เป็นรูปที่ 5.10 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query4
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:5

Source Query

Thai	การเพาะเลี้ยงสาหร่ายเกลียวทองเพื่อใช้เป็นอาหาร
English	CULTIVATION OF SPIRULINA AS FEED

Thai to English Query Translation

Corpus:0.6	เพาะเลี้ยง Spirulina ยูเรีย
Corpus:0.8	เพาะเลี้ยง Spirulina ยูเรีย อาหาร
Corpus:1.0	เพาะเลี้ยง Spirulina ยูเรีย อาหาร
Corpus:1.2	เพาะเลี้ยง Spirulina ยูเรีย อาหาร สาหร่าย
Corpus:1.4	เพาะเลี้ยง Spirulina ยูเรีย อาหาร สาหร่าย เป็น
Corpus:1.6	เพาะเลี้ยง Spirulina ยูเรีย อาหาร สาหร่าย เป็น กบมูลฟ
Corpus:1.8	เพาะเลี้ยง Spirulina ยูเรีย อาหาร สาหร่าย เป็น กบมูลฟ กากน้ำตาล
Dictionary	CULTURE ALGA PLY GOLD HAVE FOOD PROVISION GRUB

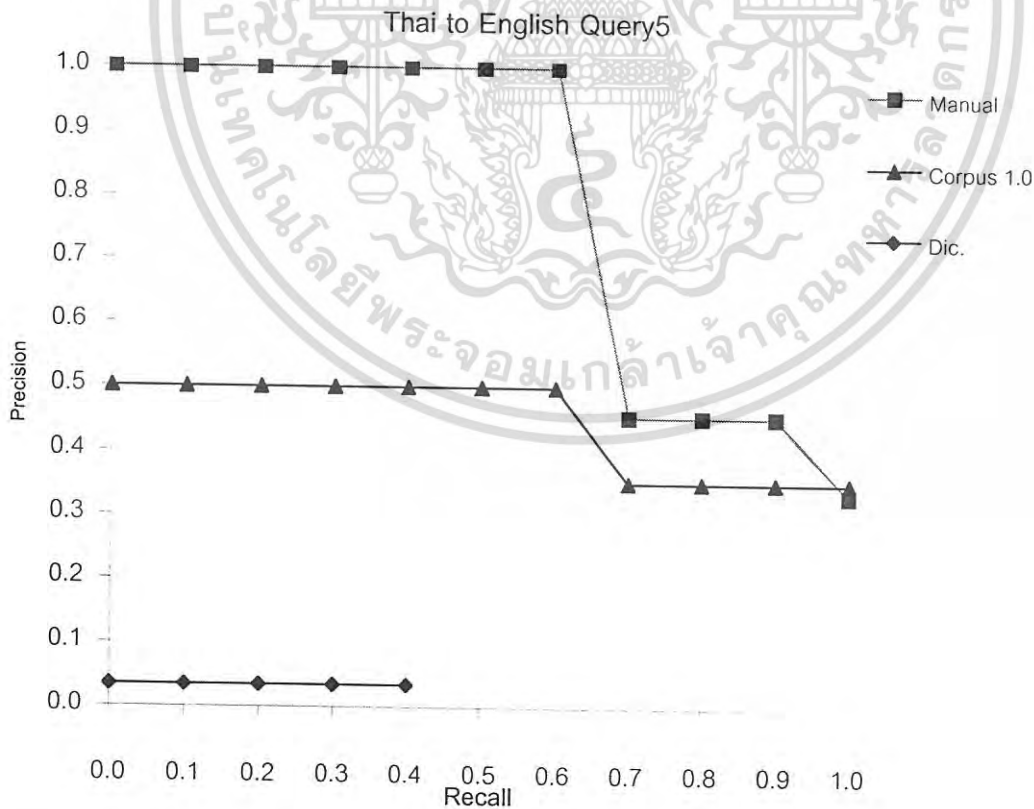
English to Thai Query Translation

Corpus:0.6	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH
Corpus:0.8	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION
Corpus:1.0	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION CHEMICAL
Corpus:1.2	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION CHEMICAL WATER SALTED
Corpus:1.4	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION CHEMICAL WATER SALTED CONDITION
Corpus:1.6	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION CHEMICAL WATER SALTED CONDITION NOODLE FERMENTED
Corpus:1.8	SPIRULINA PLATENSIS CULTIVATED OUTDOOR GROWTH COMPOSITION CHEMICAL WATER SALTED CONDITION NOODLE FERMENTED WASTE
Dictionary	การเพาะปลูก [SPIRULINA] ดั่งที่ เลี้ยง ขุน อาหารสัตว์
MT	การปลูกของ SPIRULINA ในฐานะจัดหาให้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.11 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query5

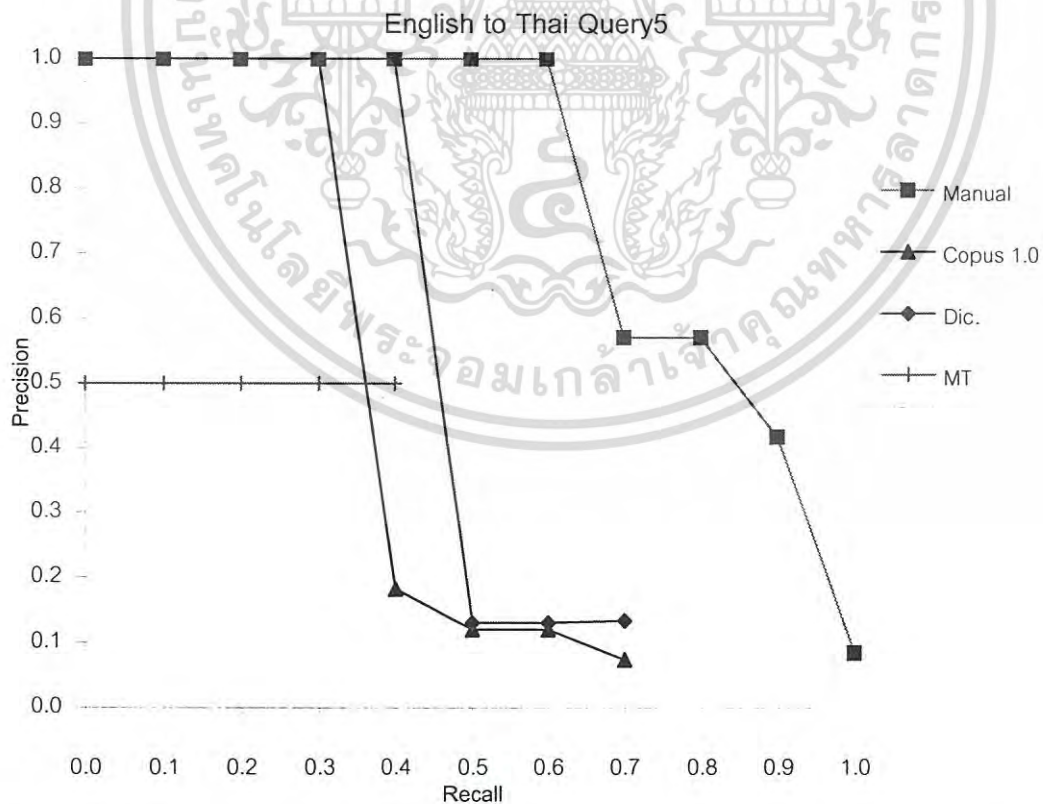
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	1.0000	0.0351	0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.1	1.0000	0.0351	0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.2	1.0000	0.0351	0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.3	1.0000	0.0351	0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.4	1.0000	0.0351	0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.5	1.0000		0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.6	1.0000		0.5000	0.5000	0.5000	0.5000	0.4286	0.4286	0.3750
0.7	0.4545		0.4286	0.4286	0.3529	0.2400	0.2222	0.2222	0.1667
0.8	0.4545		0.4286	0.4286	0.3529	0.2400	0.2222	0.2222	0.1667
0.9	0.4545		0.4286	0.4286	0.3529	0.2400	0.2222	0.2222	0.1667
1.0	0.3333		0.4286	0.4286	0.3529	0.2400	0.2222	0.2222	0.1667



เอกสารนี้เป็นรูปที่ 5.11 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query5 ระเบียบขั้นตอนการคำนวณว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.12 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query5

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5000
0.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5000
0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5000
0.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5000
0.4	1.0000	1.0000	0.2500	0.1818	0.1818	0.6667	1.0000	1.0000	1.0000	0.5000
0.5	1.0000	0.1304	0.1429	0.1200	0.1200	0.6667	0.7500	0.7500	0.7500	
0.6	1.0000	0.1304		0.1200	0.1200	0.6667	0.7500	0.7500	0.7500	
0.7	0.5714	0.1333		0.0741	0.0741	0.6667	0.6667	0.6667	0.6667	
0.8	0.5714					0.6667	0.6667	0.6667	0.6667	
0.9	0.4167					0.2222	0.1463	0.1429	0.1429	
1.0	0.0845					0.2222	0.1463	0.1429	0.1429	



รูปที่ 5.12 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาอังกฤษไปภาษาไทย Query5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:6 [02799]

Thai	อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง
English	THE INFLUENCE OF SUPPLEMENTARY NITROGEN FERTILIZE APPLY AT EAR POD FORMATION ON THE QUALITY AN YIELD OF SOYBEAN

Thai to English Query Translation

Corpus:0.6	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL
Corpus:0.8	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED
Corpus:1.0	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED SEED PLANTING
Corpus:1.2	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED SEED PLANTING PHYSIOLOGICAL VEGETABLE QUALITY
Corpus:1.4	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED SEED PLANTING PHYSIOLOGICAL VEGETABLE QUALITY QUALITY GENOLYPE
Corpus:1.6	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED SEED PLANTING PHYSIOLOGICAL VEGETABLE QUALITY QUALITY GENOLYPE KPS85 TOPPING INTRAPLANT
Corpus:1.8	YIELD FERTILIZER SOYBEAN NITROGEN GROWN EFFECT GROWTH SOIL POD SATURATED SEED PLANTING PHYSIOLOGICAL VEGETABLE QUALITY QUALITY GENOLYPE KPS85 TOPPING INTRAPLANT RATE TRANSPLANTING
Dictionary	INFLUENCE SUPPLEMENT FERTILIZER NITROGEN LENGTH BEGIN ATTACH STICK SHEATH POD TO WITI PER QUALITY PRODUCT SOYBEAN SOYABEAN

English to Thai Query Translation

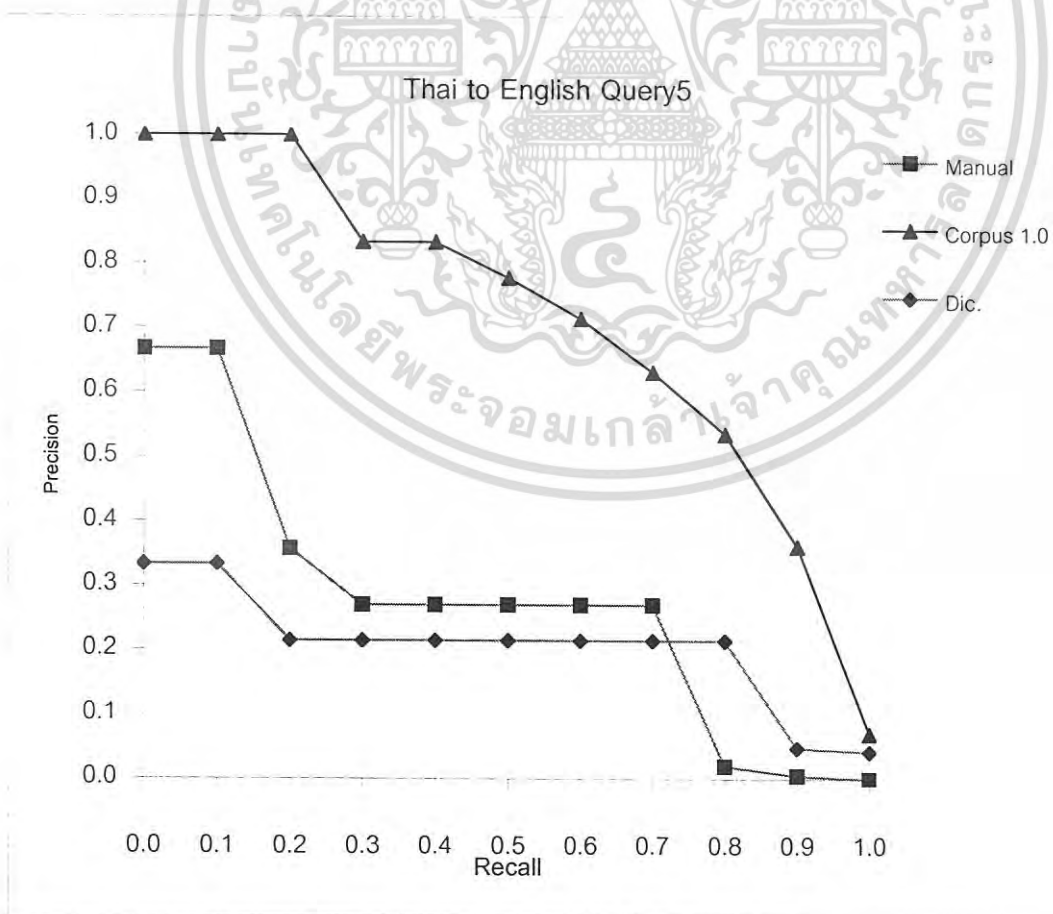
Corpus:0.6	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก
Corpus:0.8	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต
Corpus:1.0	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต คุณภาพ พันธุ์
Corpus:1.2	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต คุณภาพ พันธุ์ ดิน กิน ปุ๋ย
Corpus:1.4	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต คุณภาพ พันธุ์ ดิน กิน ปุ๋ย หญ้า ติด
Corpus:1.6	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต คุณภาพ พันธุ์ ดิน กิน ปุ๋ย หญ้า ติด เมล็ด บางชนิด
Corpus:1.8	ผลผลิต ถั่วเหลือง อิทธิพล ไนโตรเจน ฝัก ต่อ ปุ๋ยหมัก อิมตัว เจริญเติบโต คุณภาพ พันธุ์ ดิน กิน ปุ๋ย หญ้า ติด เมล็ด บางชนิด ภาคกลาง maximum
Dictionary	อิทธิพล เสริม ไนโตรเจน อุดมสมบูรณ์ ทา สมัคร ประยุกต์ใช้ หู รวงข้าว ฝัก ขบวนการ คุณภาพ ก่อให้เกิด ถั่วเหลือง

MT เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และสงวนสิทธิในเนื้อหา การนำเอกสารนี้ไปใช้ในการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.13 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query6

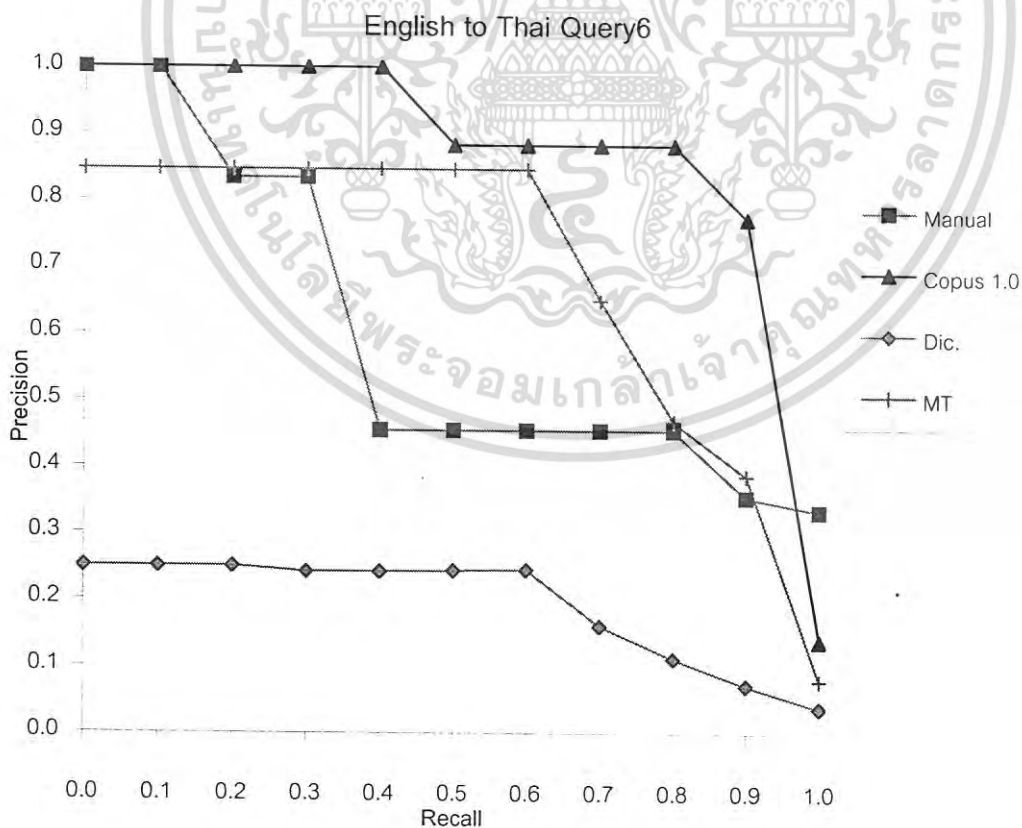
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0	1.0000	0.6000	0.5000	0.5000	0.5000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	0.6000	0.5000	0.5000	0.5000	1.0000	1.0000	1.0000	1.0000
0.2	1.0000	0.6000	0.5000	0.5000	0.5000	1.0000	1.0000	1.0000	1.0000
0.3	1.0000	0.6000	0.5000	0.5000	0.4286	0.6667	0.6667	0.6667	0.5000
0.4	0.7500	0.6000	0.5000	0.5000	0.4286	0.3333	0.3333	0.3333	0.3000
0.5	0.6667	0.3846	0.3333	0.3333	0.3077	0.2857	0.2857	0.2857	0.2857
0.6	0.6667	0.3846	0.3333	0.3333	0.3077	0.2857	0.2857	0.2857	0.2857
0.7	0.5000	0.3846	0.2941	0.2941	0.2941	0.2381	0.2381	0.2381	0.2273
0.8	0.2353	0.3158	0.2121	0.2059	0.2059	0.1628	0.1628	0.1628	0.1522
0.9	0.2353	0.3043	0.2121	0.2059	0.2059	0.1628	0.1628	0.1628	0.1522
1.0	0.2353	0.1039	0.0777	0.0750	0.0727	0.0584	0.0584	0.0584	0.0482



เอกสารนี้เป็นรูปที่ 5.13 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query6 โยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.14 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query6

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0	1.0000	1.0000	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	1.0000	1.0000	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	1.0000	1.0000	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	1.0000	1.0000	0.6667	0.5000	0.6667	0.6667	0.6667	0.6667	0.6667	1.0000
0.4	0.2222	0.6000	0.3529	0.5000	0.5000	0.6000	0.6000	0.5000	0.5000	1.0000
0.5	0.2222	0.5000	0.3529	0.4000	0.3333	0.2000	0.2000	0.1739	0.1739	1.0000
0.6	0.2222	0.5000	0.3529	0.4000	0.3333	0.2000	0.2000	0.1739	0.1739	1.0000
0.7	0.2051	0.4545	0.3529	0.3571	0.3333	0.2000	0.2000	0.1667	0.1667	0.7500
0.8	0.2051	0.3333	0.3529	0.3529	0.2857	0.1935	0.1935	0.1579	0.1579	0.7500
0.9	0.2051	0.3043	0.3077	0.3500	0.2692	0.1842	0.1842	0.1556	0.1556	0.6364
1.0	0.2051	0.2581	0.3077	0.3478	0.1951	0.1429	0.1356	0.1159	0.1159	0.3333



รูปที่ 5.14 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query6

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่สามารถนำออกจำหน่ายหรือทำซ้ำโดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:7

Source Query

Thai การดูแลสุขภาพของผู้สูงอายุ
 English HEALTH CARE OF THE AGE

Thai to English Query Translation

Corpus:0.6 CARE SELF
 Corpus:0.8 CARE SELF ELDERLY
 Corpus:1.0 CARE SELF ELDERLY
 Corpus:1.2 CARE SELF ELDERLY HEALTH
 Corpus:1.4 CARE SELF ELDERLY HEALTH CLUB
 Corpus:1.6 CARE SELF ELDERLY HEALTH CLUB
 Corpus:1.8 CARE SELF ELDERLY HEALTH CLUB HOSPITAL
 Dictionary OFFICE HEALTH PERSON AGED

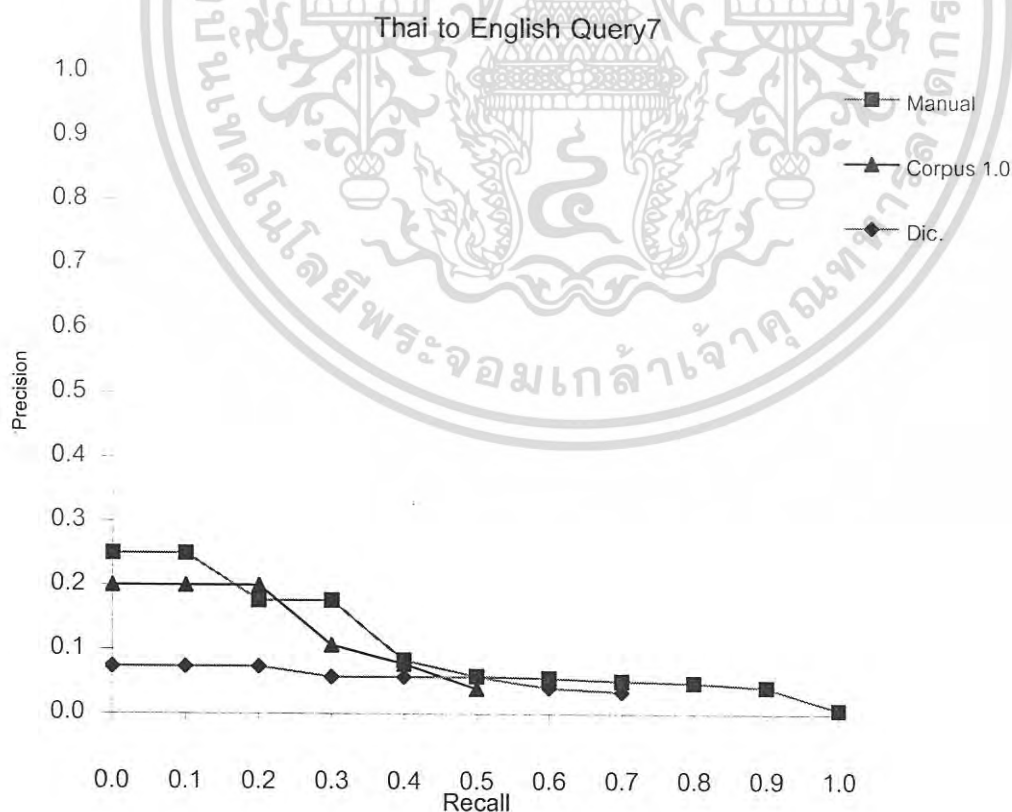
English to Thai Query Translation

Corpus:0.6 ดูแล วัยเรียน
 Corpus:0.8 ดูแล วัยเรียน ตนเอง
 Corpus:1.0 ดูแล วัยเรียน ตนเอง
 Corpus:1.2 ดูแล วัยเรียน ตนเอง มูลฐาน
 Corpus:1.4 ดูแล วัยเรียน ตนเอง มูลฐาน เด็ก
 Corpus:1.6 ดูแล วัยเรียน ตนเอง มูลฐาน เด็ก
 Corpus:1.8 ดูแล วัยเรียน ตนเอง มูลฐาน เด็ก อาสาสมัคร
 Dictionary แข็งแรง ความดูแล อายุ ยุคสมัย
 MT การดูแลรักษาสุขภาพร่างกายและจิตใจของอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.15 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query7

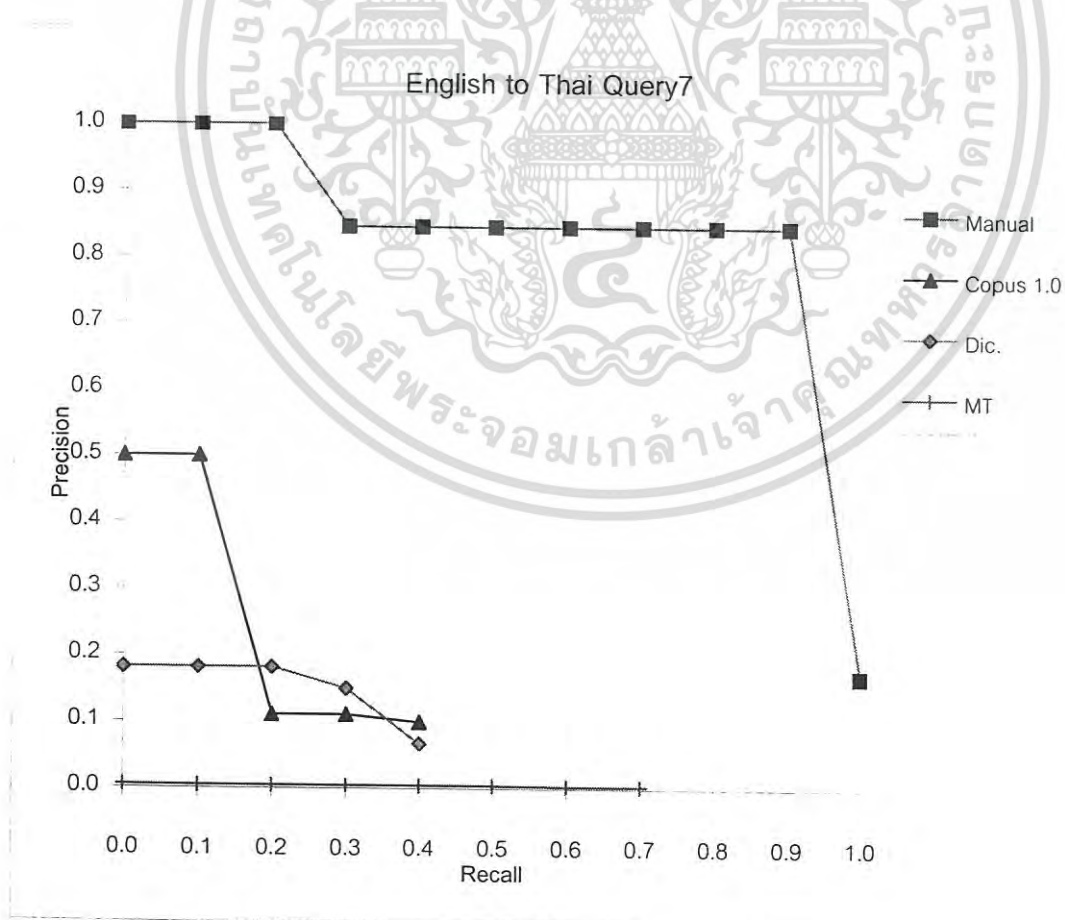
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0000	0.2500	0.0741	0.2222	0.2000	0.2000	0.2000	0.2069	0.1277	0.0577
0.1000	0.2500	0.0741	0.2222	0.2000	0.2000	0.2000	0.2069	0.1277	0.0577
0.2000	0.1765	0.0741	0.2222	0.2000	0.2000	0.2000	0.2069	0.1277	0.0577
0.3000	0.1765	0.0577	0.1200	0.1071	0.1071	0.2000	0.2069	0.1277	0.0577
0.4000	0.0833	0.0577	0.0820	0.0781	0.0781	0.1786	0.2069	0.1277	0.0577
0.5000	0.0588	0.0577	0.0403	0.0395	0.0395	0.0897	0.2069	0.1277	0.0577
0.6000	0.0556	0.0405				0.0897	0.1047	0.0692	0.0405
0.7000	0.0511	0.0349				0.0409	0.1047	0.0692	0.0349
0.8000	0.0485					0.0310	0.0482	0.0332	
0.9000	0.0415					0.0293	0.0482	0.0332	
1.0000	0.0065						0.0363	0.0264	



เอกสารนี้เป็นรูปที่ 5.15 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query7 โดยขั้นตอนการคำนวณค่า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.16 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query7

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0000	1.0000	0.1818	0.5000	0.5000	0.5000	0.1429	0.1429	0.1111	0.0556	0.0045
0.1000	1.0000	0.1818	0.5000	0.5000	0.5000	0.1429	0.1429	0.1111	0.0556	0.0045
0.2000	1.0000	0.1818	0.1250	0.1250	0.1111	0.0714	0.0714	0.0638	0.0556	0.0045
0.3000	0.8462	0.1500	0.1250	0.1250	0.1111	0.0714	0.0714	0.0638	0.0556	0.0045
0.4000	0.8462	0.0678	0.1250	0.1250	0.1000	0.0667	0.0667	0.0617	0.0258	0.0045
0.5000	0.8462								0.0243	0.0043
0.6000	0.8462								0.0225	0.0034
0.7000	0.8462									0.0033
0.8000	0.8462									
0.9000	0.8462									
1.0000	0.1688									



เอกสารนี้เป็นรูปที่ 5.16 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query7 ระโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:8

Source Query

Thai	การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อวางแผนการใช้ที่ดิน
English	THE APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM FOR LAND USE PLAN

Thai to English Query Translation

Corpus:0.6	GEOGRAPHIC LAND APPLICATION INFORMATION
Corpus:0.8	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING
Corpus:1.0	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM
Corpus:1.2	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM GIS LANDUSE
Corpus:1.4	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM GIS LANDUSE EVALUATION
Corpus:1.6	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM GIS LANDUSE EVALUATION USE
Corpus:1.8	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM GIS LANDUSE EVALUATION USE POTENTIAL
Dictionary	APPLICATION SYSTEM INFORMATION GEOGRAPHY DESIGN USE SPEND PLY LAND

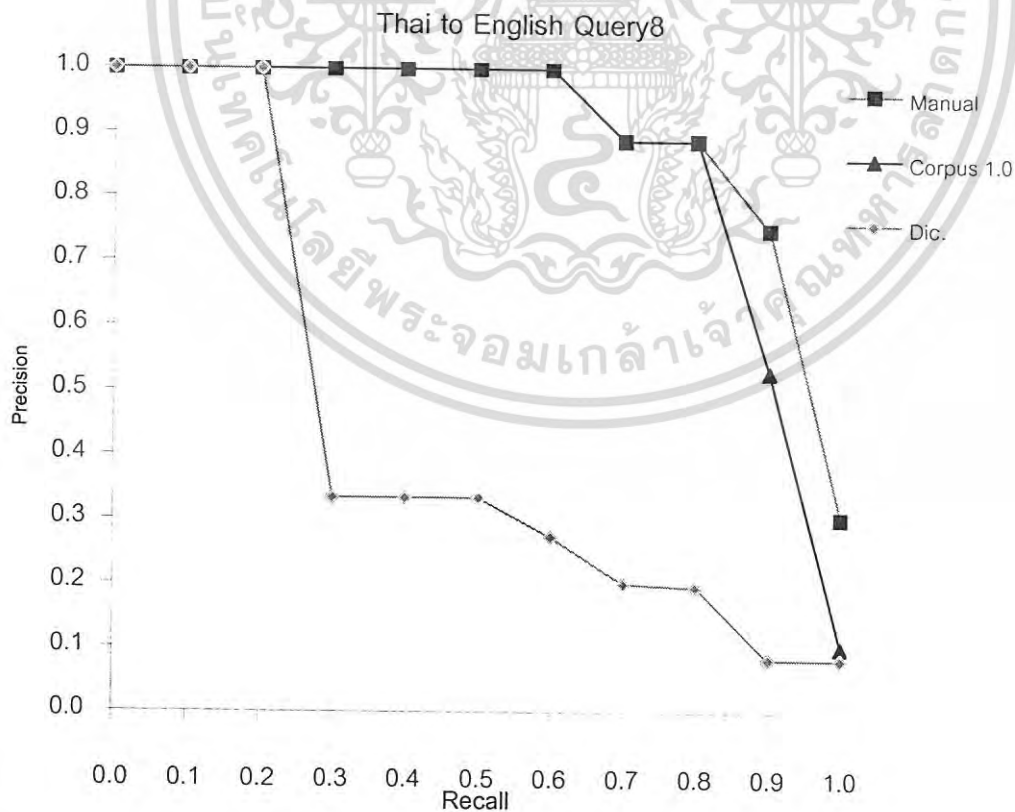
English to Thai Query Translation

Corpus:0.6	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ
Corpus:0.8	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา
Corpus:1.0	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา ใช้
Corpus:1.2	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา ใช้ กุด กุ่ม
Corpus:1.4	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา ใช้ กุด กุ่ม ทруд ผังเมือง
Corpus:1.6	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา ใช้ กุด กุ่ม ทруд ผังเมือง โป่ง
Corpus:1.8	ภูมิศาสตร์ ระบบสารสนเทศ ที่ดิน ประยุกต์ใช้ วางแผนการ พื้นที่ ประเมินราคา ใช้ กุด กุ่ม ทруд ผังเมือง โป่ง สวนสาธารณะ ทะเลสาบ
Dictionary	เอาใจใส่ ในทางภูมิศาสตร์ ชาว ความรู้ ร่างกาย ระบบ ระบอบ ที่ดิน ประเทศ ขึ้นฝั่ง ใช้ วางแผน
MT	ประยุกต์ใช้ของระบบสารสนเทศเชิงภูมิศาสตร์เพื่อแผนการใช้ที่ดิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.17 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query8

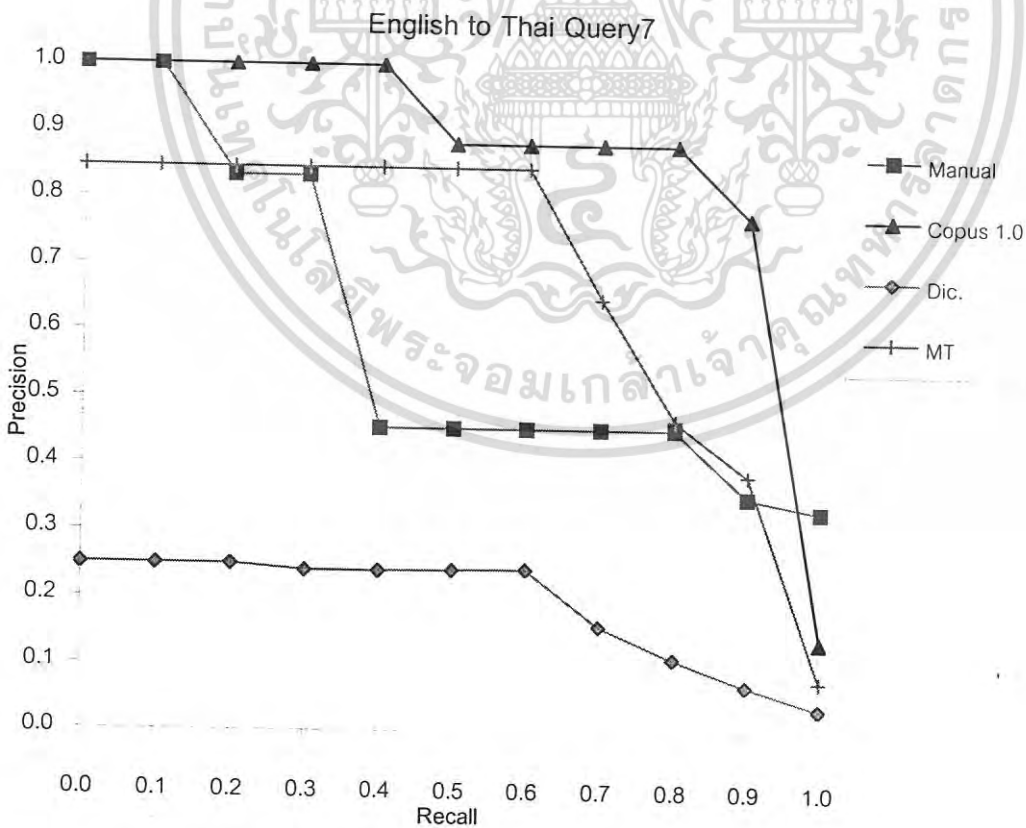
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3000	1.0000	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4000	1.0000	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5000	1.0000	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.6000	1.0000	0.2727	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7000	0.8889	0.2000	0.8889	0.8889	0.8889	1.0000	1.0000	1.0000	0.8889
0.8000	0.8889	0.1951	0.8889	0.8889	0.8889	0.9000	1.0000	1.0000	0.8889
0.9000	0.7500	0.0826	0.7500	0.7500	0.5294	0.9000	0.9000	0.9000	0.8182
1.0000	0.3030	0.0826	0.2500	0.2439	0.1020	0.5556	0.4167	0.2500	0.2439



เอกสารนี้เป็นรูปที่ 5.17 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query8
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.18 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query8

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	1.0000	0.4444	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3000	1.0000	0.4444	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4000	1.0000	0.4444	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5000	1.0000	0.3333	1.0000	1.0000	0.8333	0.8333	0.8333	0.8333	0.8333	1.0000
0.6000	1.0000	0.1622	1.0000	0.7778	0.7778	0.7778	0.7778	0.7500	0.7500	1.0000
0.7000	0.8750	0.1750	1.0000	0.7778	0.7778	0.7778	0.7778	0.7000	0.7000	1.0000
0.8000	0.8333	0.0909	1.0000	0.6667	0.7273	0.7273	0.7273	0.6667	0.6667	0.9091
0.9000	0.8333	0.0732	0.8182	0.6000	0.6429	0.6429	0.6429	0.6000	0.6000	0.9091
1.0000	0.8333	0.0324	0.7692	0.5882	0.5556	0.5556	0.5556	0.5263	0.5263	0.9091



เอกสารนี้เป็นรูปที่ 5.18 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query8
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:9

Source Query

Thai	การย่อยสลายแป้งมันสำปะหลัง โดยใช้เอนไซม์
English	HYDROLYSIS OF CASSAVA STARCH BY ENZYME

Thai to English Query Translation

Corpus:0.6	STARCH TAPIOCA CASSAVA WHEAT FLOUR
Corpus:0.8	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE
Corpus:1.0	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE MIXED
Corpus:1.2	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE MIXED MODIFICATION NAISE
Corpus:1.4	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE MIXED MODIFICATION NAISE SHOREA
Corpus:1.6	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE MIXED MODIFICATION NAISE SHOREA DEGRADABLE MAYON
Corpus:1.8	STARCH TAPIOCA CASSAVA WHEAT FLOUR POLYETHYLENE MIXED MODIFICATION NAISE SHOREA DEGRADABLE MAYON COMBLNATION
Dictionary	DIGESTION COLLAPSE POWDER FLOUR CASSAVA USE SPEND PLY SINEW TENDON [ไซม์]

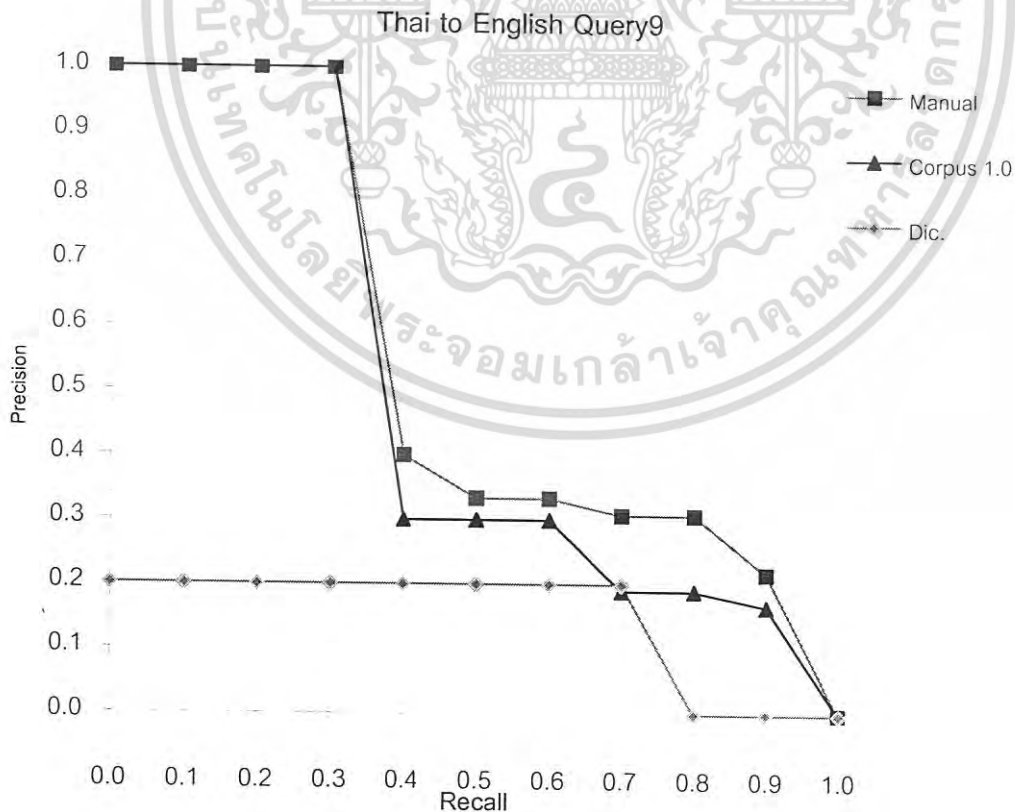
English to Thai Query Translation

Corpus:0.6	มันสำปะหลัง แป้ง Manihot
Corpus:0.8	มันสำปะหลัง แป้ง Manihot Crantz
Corpus:1.0	มันสำปะหลัง แป้ง Manihot Crantz
Corpus:1.2	มันสำปะหลัง แป้ง Manihot Crantz niger
Corpus:1.4	มันสำปะหลัง แป้ง Manihot Crantz niger เอนไซม์
Corpus:1.6	มันสำปะหลัง แป้ง Manihot Crantz niger เอนไซม์ Aspergillus
Corpus:1.8	มันสำปะหลัง แป้ง Manihot Crantz niger เอนไซม์ Aspergillus กรด
Dictionary	[HYDROLYSIS] มันสำปะหลัง แป้ง เอนไซม์
MT	การสลายด้วยน้ำของการลงแป้งมันสำปะหลังโดยเอนไซม์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.19 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query9

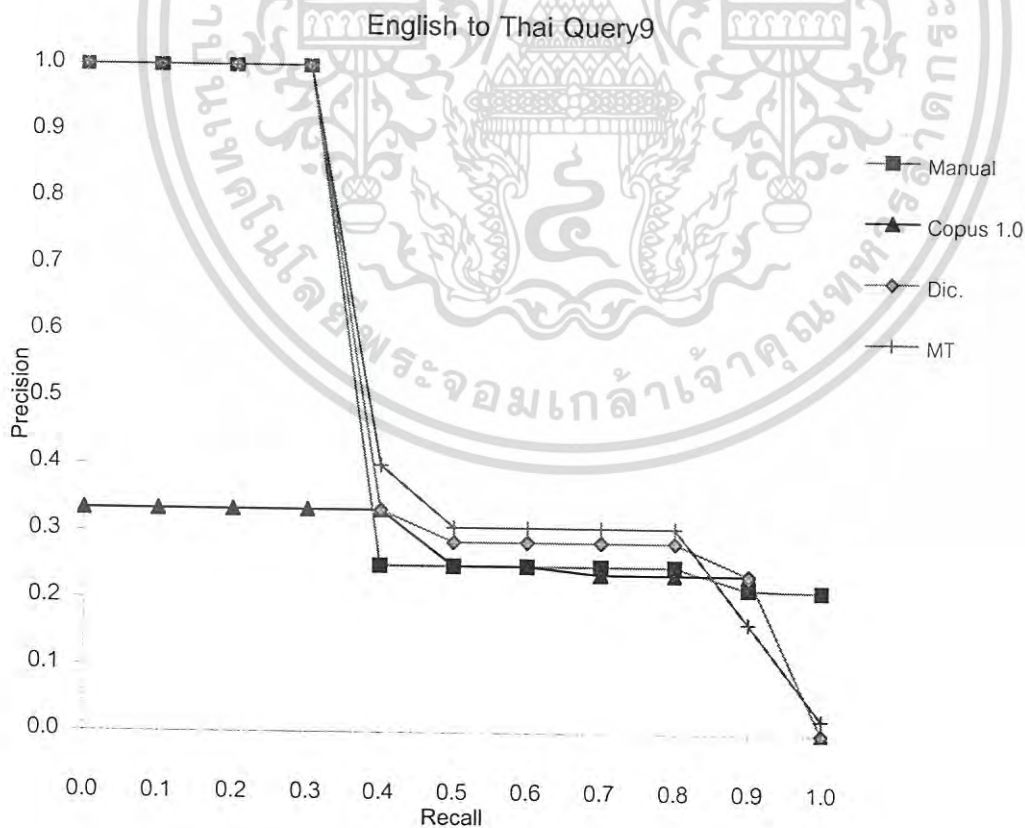
Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0000	1.0000	0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1000	1.0000	0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	1.0000	0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3000	1.0000	0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4000	0.4000	0.2000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000
0.5000	0.3333	0.2000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000
0.6000	0.3333	0.2000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000
0.7000	0.3077	0.2000	0.2353	0.2105	0.1905	0.1667	0.1667	0.1667	0.1667
0.8000	0.3077		0.2353	0.2105	0.1905	0.1667	0.1667	0.1667	0.1667
0.9000	0.2174		0.2174	0.1786	0.1667	0.1515	0.1515	0.1515	0.1515
1.0000									



เอกสารนี้เป็นรูปที่ 5.19 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query8 โดยใช้ประโยชน์ด้านการคำนวณที่มากกว่าวิธีอื่น ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.20 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query9

Recall	Manual	Dic.	Corpus							MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0000	1.0000	1.0000	0.3333	0.3333	0.3333	0.3333	1.0000	1.0000	0.5000	1.0000
0.1000	1.0000	1.0000	0.3333	0.3333	0.3333	0.3333	1.0000	1.0000	0.5000	1.0000
0.2000	1.0000	1.0000	0.3333	0.3333	0.3333	0.3333	1.0000	1.0000	0.5000	1.0000
0.3000	1.0000	1.0000	0.3333	0.3333	0.3333	0.3333	1.0000	1.0000	0.5000	1.0000
0.4000	0.2500	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.2500	0.2353	0.4000
0.5000	0.2500	0.2857	0.2500	0.2500	0.2500	0.2500	0.2857	0.2500	0.2353	0.3077
0.6000	0.2500	0.2857	0.2500	0.2500	0.2500	0.2500	0.2857	0.2500	0.2353	0.3077
0.7000	0.2500	0.2857	0.2381	0.2381	0.2381	0.2273	0.2857	0.2500	0.2353	0.3077
0.8000	0.2500	0.2857	0.2381	0.2381	0.2381	0.2273	0.2857	0.2500	0.2353	0.3077
0.9000	0.2174	0.2381	0.2381	0.2381	0.2381	0.2273	0.2273	0.2000	0.1852	0.1667
1.0000	0.2143									0.0222



รูปที่ 5.20 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:10

Source Query

Thai	พฤติกรรมของแผ่นพื้นคอนกรีตท้องเรียบชนิดอัดแรงบางส่วน
English	BEHAVIOUR OF PARTIAL PRESTRESS CONCRETE FLAT PLATE

Thai to English Query Translation

Corpus:0.6	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT
Corpus:0.8	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST
Corpus:1.0	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST PRECAST SIMPLEX
Corpus:1.2	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST PRECAST SIMPLEX CONTINUITY DESIGN
Corpus:1.4	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST PRECAST SIMPLEX CONTINUITY DESIGN METHOD PLANK
Corpus:1.6	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST PRECAST SIMPLEX CONTINUITY DESIGN METHOD PLANK KOQ LATERALLY
Corpus:1.8	PRESTRESSED CONCRETE TENSIONED PLATE SLAB FLAT OPTIMUM POST PRECAST SIMPLEX CONTINUITY DESIGN METHOD PLANK KOQ LATERALLY LOADED WIDTH
Dictionary	ACT BOARD GROUND FLOOR CONCRETE ABDOMEN BELLY TUMMY EVEN FLAT RUN ROLL STRENGTH FORCE STRONG STRONGLY PORTION

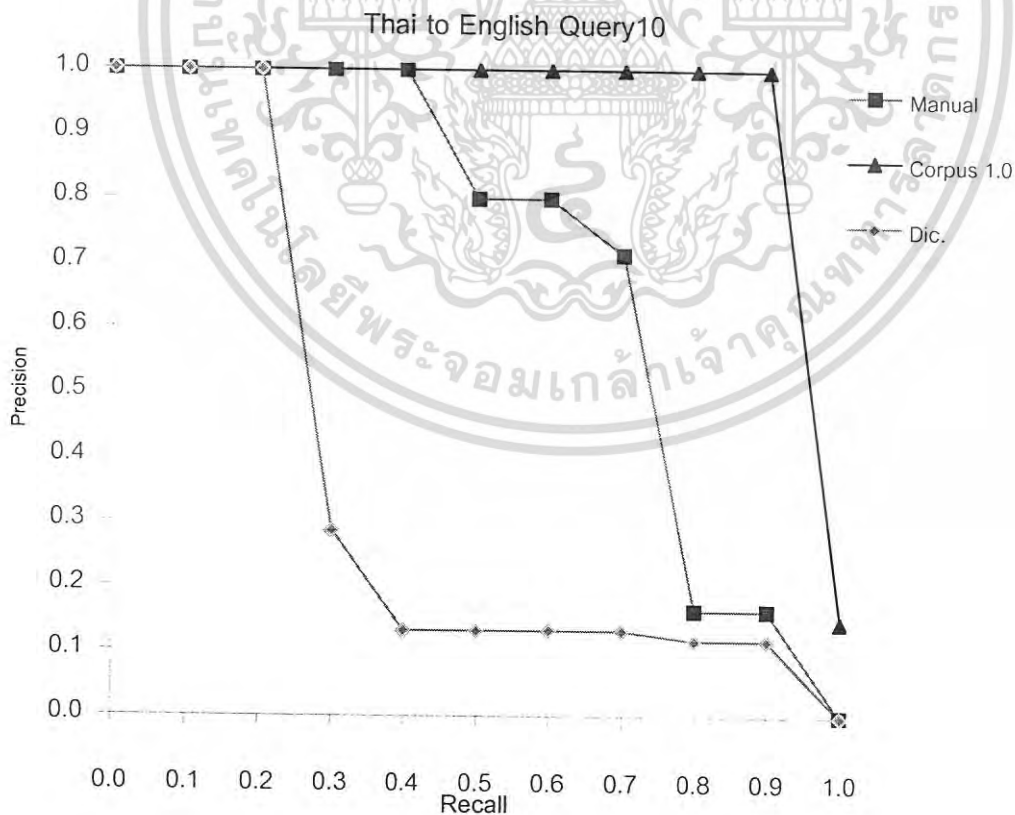
English to Thai Query Translation

Corpus:0.6	คอนกรีต พื้น คาน แรง
Corpus:0.8	คอนกรีต พื้น คาน แรง แผ่น
Corpus:1.0	คอนกรีต พื้น คาน แรง แผ่น อัด
Corpus:1.2	คอนกรีต พื้น คาน แรง แผ่น อัด ไร้ เหล็ก
Corpus:1.4	คอนกรีต พื้น คาน แรง แผ่น อัด ไร้ เหล็ก คำนวณ
Corpus:1.6	คอนกรีต พื้น คาน แรง แผ่น อัด ไร้ เหล็ก คำนวณ ออกแบบ
Corpus:1.8	คอนกรีต พื้น คาน แรง แผ่น อัด ไร้ เหล็ก คำนวณ ออกแบบ ภายหลัง
Dictionary	ความประพฤติ ส่วนหนึ่ง อัดแรงล่วงหน้า รูปธรรม คอนกรีต แบน ราบ เรียบ ห้อยชุด ราบเรียบ จาน
MT	พฤติกรรมของจานราบเรียบคอนกรีตแรงอัดล่วงหน้าบางส่วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.21 ค่าความแม่นยำการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query10

Recall	Manual	Dic.	Corpus						
			0.6	0.8	1.0	1.2	1.4	1.6	1.8
0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3000	1.0000	0.2857	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4000	1.0000	0.1316	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5000	0.8000	0.1316	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.6000	0.8000	0.1316	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7000	0.7143	0.1316	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8000	0.1628	0.1167	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.9000	0.1628	0.1167	0.8750	0.7778	1.0000	1.0000	0.8750	0.8750	0.8750
1.0000			0.2500	0.1509	0.1455	0.5000	0.7273	0.7273	0.7273

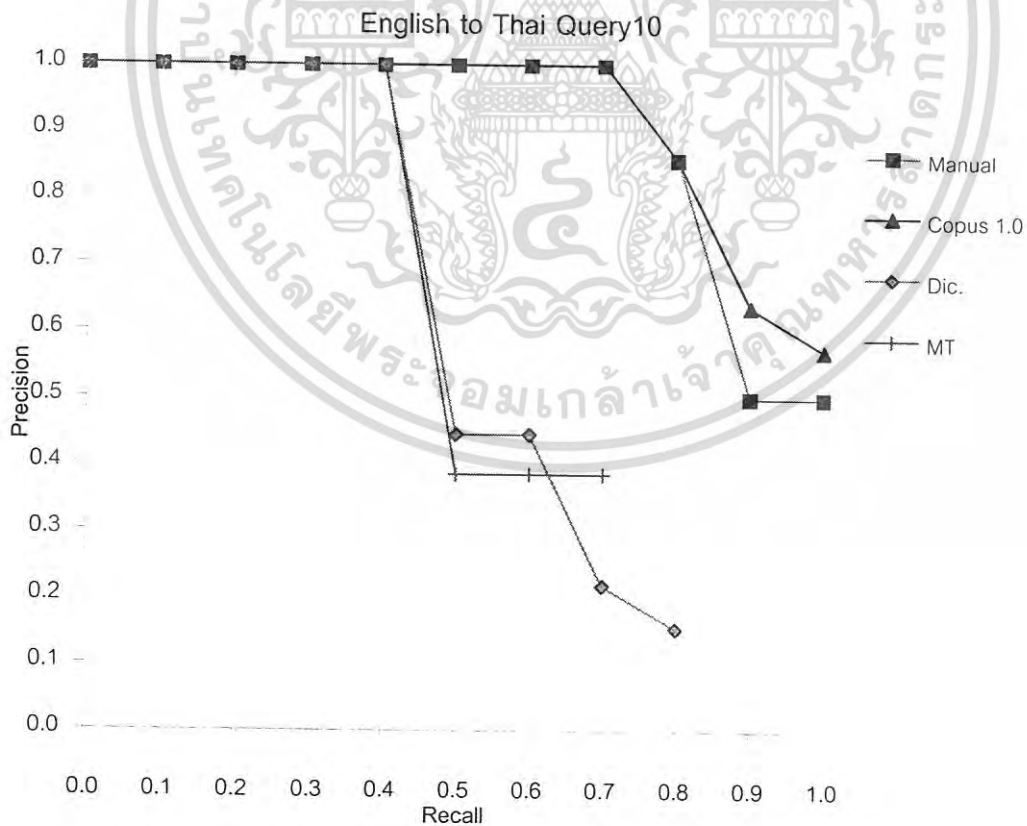


รูปที่ 5.21 ประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ Query10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.22 ค่าความแม่นยำของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย Query10

Recall	Manual	Dic.								MT
			0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5000	1.0000	0.4444	0.8571	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3846
0.6000	1.0000	0.4444	0.8571	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3846
0.7000	1.0000	0.2174	0.8571	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3846
0.8000	0.8571	0.1538	0.8571	0.8750	0.8571	0.8571	0.8571	0.8571	0.8571	
0.9000	0.5000		0.4118	0.8750	0.6364	0.5385	0.5385	0.3889	0.3889	
1.0000	0.5000		0.3333	0.8000	0.5714	0.4706	0.4211	0.3636	0.3636	



รูปที่ 5.22 ประสิทธิภาพการค้นคืนข้ามภาษาจากภาษาอังกฤษไปภาษาไทย Query10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้มีการเผยแพร่ข้อมูลใดๆ
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 การทดลองวัดค่าความแม่นยำเมื่อค่า Recall คงที่

วิธีการวัดประสิทธิภาพของระบบค้นคืนสารสนเทศ อีกแนวทางหนึ่งทำโดย กำหนดค่า Recall ให้คงที่ เช่น 5, 10, 20, 50 หรือ 100 เอกสาร เราได้ทำการแสดงผลความแม่นยำจากการค้นคืนข้ามภาษา โดยใช้คลังข้อมูลภาษาแบบขนาน

ตารางที่ 5.23 ค่าความแม่นยำเมื่อค่า Recall คงที่ของการค้นคืนฯ จากภาษาไทยไปภาษาอังกฤษ

วิธี PRF โดยมีความยาว Query 1.2 เท่า

Recall	Query1	Query2	Query3	Query4	Query5	Query6	Query7	Query8	Query9	Query10	Average
5	0.80	0.60	0.20	1.00	0.50	0.40	0.16	1.00	0.30	0.30	0.53
10	0.60	0.50	0.20	0.80	0.30	0.30	0.15	0.90	0.30	0.30	0.44
20	0.45	0.25	0.10	0.70	0.15	0.20	0.15	0.50	0.15	0.15	0.28
50	0.30	0.16	0.09	0.46	0.15	0.14	0.10	0.20	0.08	0.08	0.18
100	0.17	0.09	0.05	0.33	0.06	0.07	0.07	0.10	0.05	0.05	0.10

ตารางที่ 5.24 ค่าความแม่นยำเมื่อค่า Recall คงที่ของการค้นคืนฯ จากภาษาอังกฤษไปภาษาไทย

วิธี PRF โดยมีความยาว Query 1.4 เท่า

Recall	Query1	Query2	Query3	Query4	Query5	Query6	Query7	Query8	Query9	Query10	Average
5	0.80	0.60	0.60	1.00	0.40	0.60	0.14	0.80	0.33	1.00	0.63
10	0.80	0.38	0.60	0.80	0.30	0.30	0.10	0.70	0.20	0.60	0.48
20	0.65	0.30	0.45	0.65	0.15	0.20	0.05	0.50	0.14	0.40	0.35
50	0.36	0.16	0.20	0.42	0.12	0.14	0.04	0.20	0.10	0.16	0.19
100	0.19	0.10	0.11	0.36	0.06	0.08	0.05	0.10	0.05	0.08	0.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

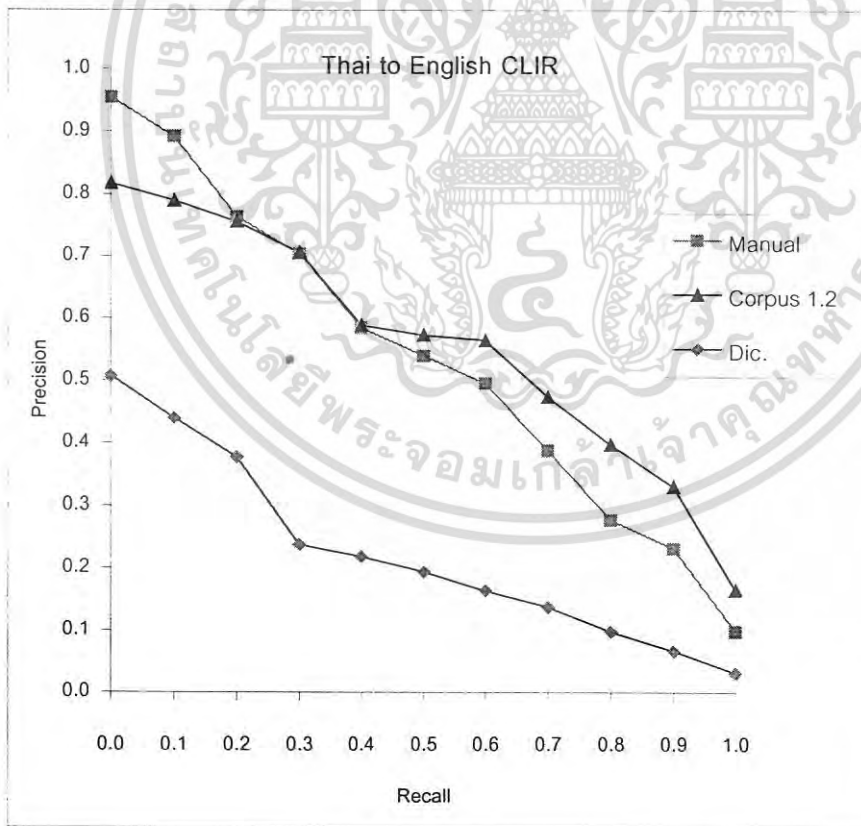
สรุปผลการทดลองและข้อเสนอแนะ

จากการทดลองวัดผลประสิทธิภาพการค้นคืนข้ามภาษา จากภาษาไทยเป็นภาษาอังกฤษ และ จากภาษาอังกฤษเป็นภาษาไทยโดยวิธีการแปลคำสั่งสอบถาม พบว่าเมื่อใช้ฐานข้อมูลข้อความร่วมกับวิธีการ PRF สามารถให้ประสิทธิภาพการค้นคืนที่ดีกว่าวิธีการอื่น และยังสามารถค้นคืนข้ามภาษาได้ทั้งภาษาไทยเป็นภาษาอังกฤษ และ จากภาษาอังกฤษเป็นภาษาไทย

ตารางที่ 6.1 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาไทยไปภาษาอังกฤษ

Manual	Dic.	Corpus							
		0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.5392	0.2251	0.5384	0.5507	0.5424	0.5605	0.5542	0.5430	0.5313	

การใช้ฐานข้อมูลข้อความร่วมกับวิธีการ PRF แปลคำสั่งสอบถามให้มีความยาว 1.2 เท่าของสิ่งสอบถามต้นฉบับความแม่นยำเฉลี่ยสูงสุด เมื่อแปลคำสั่งสอบถามภาษาไทยเป็นภาษาอังกฤษ



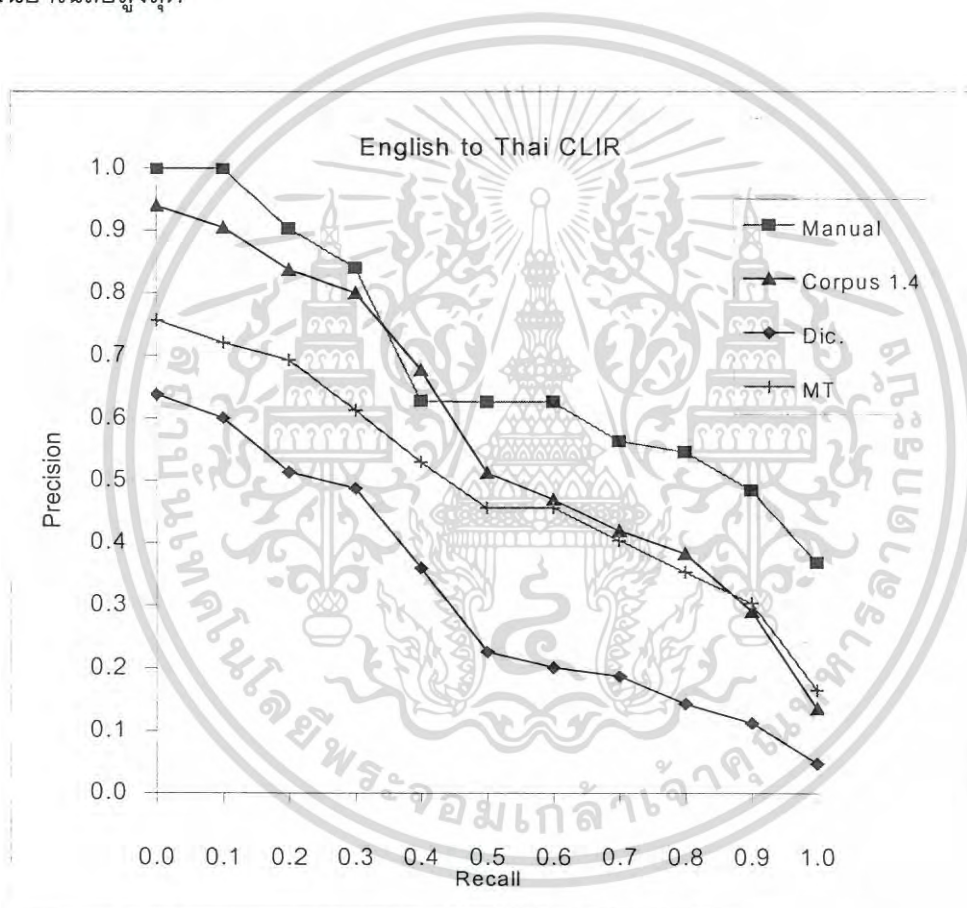
รูปที่ 6.1 กราฟค่าความแม่นยำเฉลี่ย CLIR จากภาษาไทยไปภาษาอังกฤษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 6.2 ค่าความแม่นยำเฉลี่ยของการค้นคืนข้ามภาษา จากภาษาอังกฤษไปภาษาไทย

Manual	Dic.	Corpus							MT
		0.6	0.8	1.0	1.2	1.4	1.6	1.8	
0.6892	0.3197	0.5116	0.5145	0.5274	0.5443	0.5788	0.5710	0.5321	0.4952

ในทำนองเดียวกัน เมื่อแปลคำสั่งสอบถามภาษาไทยเป็นภาษาอังกฤษ การใช้ฐานข้อมูลข้อความร่วมกับวิธีการ PRF แปลคำสั่งสอบถามให้มีความยาว 1.4 เท่าของคำสั่งสอบถามต้นฉบับ จะให้การค้นคืนที่มีความแม่นยำเฉลี่ยสูงสุด



รูปที่ 6.2 กราฟค่าความแม่นยำเฉลี่ย CLIR จากภาษาอังกฤษไปภาษาไทย

6.1. การแปลคำสั่งสอบถามโดยการใช้พจนานุกรม

แม้ว่าผลการวิจัยในต่างประเทศ [23] การแปลคำสั่งสอบถามด้วยพจนานุกรมให้ผลการค้นคืนที่มีประสิทธิภาพไม่แพ้วิธีการอื่น แต่เนื่องจากการจัดเตรียมพจนานุกรมที่เราใช้ในการทดลองนำมาจากพจนานุกรมที่ใช้กันทั่วไป คำศัพท์เฉพาะที่มีอยู่ในชุดข้อมูลทดสอบจึงไม่มีในพจนานุกรม อีกทั้งวิธีการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แปลแบบคำต่อคำสามารถมีโอกาสที่จะเลือกคำแปลที่ไม่ถูกต้องเป็นไปได้มาก แม้ว่าผลการทดลองจะดีกว่าการศึกษาคั้งก่อน [24] แต่ผลที่ได้เมื่อเทียบกับวิธีการอื่น ยังด้อยกว่าอยู่มาก

6.2. การแปลคำสั่งสอบถามโดยใช้ระบบแปลภาษา

เนื่องจากระบบแปลภาษา "ภาษาซี" ได้ถูกออกแบบมาใช้งานทั่วไป โดยเฉพาะการแปลข้อความจากเว็บเพจ คำบางคำที่มีหลายความหมาย ระบบแปลภาษาจะเลือกความหมายที่คิดว่าเหมาะสมที่สุดมาเพียงคำเดียว จึงอาจเป็นสาเหตุให้ความแม่นยำในการแปลคำสั่งสอบถามด้อยกว่าการแปลด้วยฐานข้อมูลข้อความ

ประสิทธิภาพการแปลที่เหนือกว่าการแปลด้วยพจนานุกรม แสดงถึงความสามารถในการแก้ปัญหาความกำกวมของศัพท์ของระบบแปลภาษา รวมทั้งความสะดวกในการใช้งานทำให้ระบบแปลภาษาน่าจะมีบทบาทในการศึกษาการค้นคว้าข้ามภาษาในอนาคต โดยเฉพาะเมื่อมีการพัฒนาให้สามารถแปลภาษาไทยเป็นภาษาอังกฤษได้ จะช่วยให้การเข้าถึงข้อมูลภาษาอังกฤษของผู้ใช้ชาวไทยทำได้ดีขึ้น

6.3. การแปลคำสั่งสอบถามโดยใช้ฐานข้อมูลข้อความร่วมกับวิธีการ PRF

ฐานข้อมูลข้อความร่วมกับวิธีการ PRF ให้ผลการค้นคว้าที่ดีมาก โดยเมื่อแปลความยาวของคำสั่งสอบถามที่แปลได้จาก 0.6-1.8 เท่าของความยาวคำสั่งสอบถามภาษาต้นฉบับ พบว่าผลการแปลที่ดีที่สุดคือ

6.3.1. การแปลคำสั่งสอบถามจากภาษาไทยเป็นภาษาอังกฤษโดยมีความยาว 1.2 เท่าของความยาวคำสั่งสอบถามภาษาต้นฉบับ

6.3.2. การแปลคำสั่งสอบถามจากภาษาอังกฤษเป็นภาษาไทยโดยมีความยาว 1.4 เท่าของความยาวคำสั่งสอบถามภาษาต้นฉบับ

คำสั่งสอบถามที่แปลแล้วบางคำสั่ง ให้ประสิทธิภาพดีกว่าบางคำสั่งสอบถามที่แปลด้วยคน อาจเนื่องมาจากการมีคำแปลของคำพ้องความหมายหลายคำ เช่น

คำสั่งสอบถามต้นฉบับ	การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อวางแผนการใช้ที่ดิน
แปลด้วยคน	THE APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM FOR LAND USE PLAN
แปลด้วย PRF	GEOGRAPHIC LAND APPLICATION INFORMATION PLANNING SYSTEM GIS LANDUSE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแปลด้วย PRF มีทั้งคำว่า GIS ซึ่งเป็นการขยายความคำสั่งสอบถาม (Query Expansion) ได้โดยอัตโนมัติ

6.4. ค่าความแม่นยำเมื่อค่า Recall คงที่ของวิธีการ PRF

จากผลการทดลองในตารางที่ 5.23 และ 5.24 แสดงให้เห็นความแม่นยำ ในระดับประมาณ 0.5 ที่การเรียกคืน (Recall) 5 ถึง 10 เอกสาร นั้นหมายความว่า เอกสารที่ค้นได้ในอันดับต้นๆ จะเป็นเอกสารที่ผู้ใช้ต้องการโดยเฉลี่ยประมาณครึ่งหนึ่ง ซึ่งจัดว่าสามารถใช้งานจริงได้

6.5. ข้อคิดเห็นเพิ่มเติมและการศึกษาในอนาคต

6.5.1. การทดสอบด้วยชุดข้อมูลทดสอบที่มีที่มาเดียวกับชุดข้อมูลฝึก จึงอาจเป็นผลให้วิธีการ PRF สามารถแปลคำสั่งสอบถามได้เป็นอย่างดี หากมีชุดข้อมูลทดสอบมีเนื้อหาแตกต่างกับชุดข้อมูลฝึกอย่างมากผลการแปลด้วยวิธีการ PRF อาจจะมีประสิทธิภาพลดลง

6.5.2. คำสั่งสอบถามที่แปลด้วยคนบางคำสั่ง ให้ประสิทธิภาพในการค้นคืนต่ำกว่าวิธีการ PRF อาจเนื่องมาจากกระบวนการสร้างคำสั่งสอบถามซึ่งใช้วิธีการตัดคำออกเพียงอย่างเดียวไม่มีการเพิ่มคำสำคัญเข้าไป

6.5.3. การสันนิษฐานว่า ข้อมูลในชุดข้อมูลฝึก (Training Corpus) จำนวน 10 ตัวอย่างเป็นเอกสารที่เกี่ยวข้องกันกับคำสั่งสอบถามต้นฉบับ อาจจะไม่ใช้วิธีการเลือกเอกสารที่เกี่ยวข้องกันกับคำสั่งสอบถามที่ดีที่สุด อาจจะมีเพียง 5 เอกสารเท่านั้นที่เกี่ยวข้องกัน ซึ่งทำให้เอกสารที่ไม่เกี่ยวข้องอีก 5 เอกสาร เพิ่มคำที่ไม่เกี่ยวข้องลงในคำสั่งสอบถามที่แปลได้

6.5.4. การประยุกต์ใช้การค้นคืนสารสนเทศข้ามภาษาอย่างเป็นรูปธรรม อาจเริ่มจากการให้บริการค้นคืนจาก "ฐานข้อมูลวิทยานิพนธ์ไทย Online" เพราะกลุ่มผู้ใช้มีความรู้ภาษาอังกฤษดีกว่ากลุ่มอื่น ทำให้ไม่จำเป็นต้องใช้ระบบแปลภาษา ถือเป็นงานอำนวยความสะดวกในการค้นคืนงานวิจัยจากต่างประเทศไปด้วย เพื่อให้งานวิจัยในระดับบัณฑิตศึกษามีความหลากหลายมากขึ้น

6.5.5. การประยุกต์เทคนิคการประมวลผลภาษาธรรมชาติเข้าช่วยในการแปลคำสั่งสอบถาม อาจทำให้การแปลคำสั่งสอบถามที่เป็นรูปประโยคสมบูรณ์ มีความถูกต้องมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] Global Reach. 2003. Global Internet Statistics (by Language) [Online]. Available: <http://www.greach.com/globstats/>, 30 Sept., 2003
- [2] D. Oard and B. Dorr. 1996. A Survey of Multilingual Text Retrieval. Technical Report UMIACS-TR-96-19 CD-TR-3615, University of Maryland, College Park.
- [3] Peter Schauble, Paraic Sheridan 1997. Cross-Language Information Retrieval (CLIR) Track Overview [Online]. Available: <http://trec.nist.gov/pubs/trec6/papers/>, page 31
- [4] Ari Pirkola. 2001. The Effects of Query Structure and Dictionary Setups in Dictionary – Based Cross Language Information Retrieval [Online]. Available: <http://www.cs.mu.oz.au/sigir98/program.html>.
- [5] Peter Schauble. 2000. CLEF 2000 State- of- the Art Multilingual Information Access [Online], Available: <http://www.clef-campaign.org>.
- [6] Jaruskulchai Chuleerat. 2001. Dictionary Based Thai CLIR: Experimental survey of Thai CLIR, CLEF 2001 Workshop, 3-4 September, Darmstadt, Germany.
- [7] Suwanvisat P. and Prasijutrakul S. 2000. Thai-English Cross-Language Transliterated Word Retrieval Soundex Technique, NCSEC2000.
- [8] Kando, N., Koyama, T., Oyama, K., Kageura, K., Yoshioka, M., Nozue, T., Matsumura, A., Kuriyama, K. 1998. NTCIR : NACSIS Test Collection Project." In [Poster] the 20th Annual Colloquium of the British Computer Society Information Retrieval Specialist Group. Autrans, France
- [9] Micheal Kluck. 2003. Test collection report for the CLEF 2003 Campaign[Online]. Available: <http://clef.iei.pi.cnr.it:2002>.
- [10] Doug Cooper. 1999. Thai On-Line Library - Bitext Corpus [Online] , Available: <http://searcs.th.net/bitext>
- [11] Noppong Eoh, 1998-1999. The Idea Magazine. Bangkok, Thailand. Penboon Distribution Co.,Ltd.
- [12] Systan software Inc. 2003. The SYSTRAN Linguistics Platform [Online], Available: <http://www.systransoft.com/Technology/SLP.pdf>

[13] Doug Cooper. 2004. Center for Research in Computational Linguistics[Online].
 เอกสารอ้างอิง
 ไม่ว่ากรรมใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Available: <http://seasrc.th.net>

- [14] S. Sethaputra, 1996. *Thaisoft So Sethaputra Dictionary Version 2.0* [Computer Software], Bangkok : Thaisoft Co.,Ltd
- [15] NECTEC, 2000. *LEXITRON Thai English Dictionary version 1.0* [Computer Software], Bangkok, Thailand
- [16] George A. Miller,Christiane Fellbaum, Randee Teng, Susanne Wolff, Pamela Wakefield Helen Langone, Benjamin Haskell , 2002 *WordNet 2.0* [Computer Software], Cognitive Science Laboratory Princeton University
- [17] ยืน ภู่วรวรรณ. 2540. *พจนานุกรมคำพ้องไทย*
- [18] เรียงชัย เขี่ยมวรเมธ, 2546. *พจนานุกรม ไทย-อังกฤษ และ อังกฤษ-ไทย* [Computer Software]
- [19] ศูนย์บริการสารสนเทศทางเทคโนโลยี. 2546. *ฐานข้อมูลวิทยานิพนธ์ไทย Online* [Online]. Available: <http://thesis.tiac.or.th>
- [20] Gerard Salton. 1991. *The SMART Information Retrieval System after 30 years*, SIGIR 1991: page 356-358
- [21] Monica Rogati. 2001. *CMU PRF using a Comparable Corpus*. CLEF Working Notes, Carnegie Mellon University
- [22] Yiming Yang ,Nianli Ma. 2002. *CMU in Cross-Language Information Retrieval at NTCIR-3*.: Proceedings of the Third NTCIR Workshop
- [23] Lisa Ballesteros. 1997. *Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval* .
- [24] Nathanan Srisa-ard, Chotipat Pornavalai. 2002. *Thai-English Query Translation in Cross-Language Information Retrieval*. SNLP-O-COCOSDA 2002

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

Publication

Thai-English Query Translation in Cross-Language Information Retrieval

From

The Fifth Symposium on Natural Language Processing 2002 + Oriental

COCOSDA Workshop 2002

9-11 May 2002 Hua Hin, Prachuapkirikhan, Thailand



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thai-English Query Translation in Cross-Language Information Retrieval

Nathanan Srisa-ard and Chotipat Pornavalai

Faculty of Information Technology
King Mongkut Institute of Technology Ladkrabang,
Chalongkrung Road, Ladkrabang
Bangkok Thailand 10520
e-mail : nathanan@chaiyo.com, chotipat@it.kmitl.ac.th

Abstract

This document presents Thai to English and English to Thai cross-language information retrieval, which retrieves the document from one language by using query from another language. We use the query translation approach and prepare the translation resource, parallel corpus collected from Thai thesis abstract, and article in the “Idea magazine”. We also prepared the bi-lingual dictionary from desktop published dictionary “So Setapura”. Parallel corpus can help the disambiguation translation by using the proposed modified pseudo-relevance feedback method. Our experiment shows that the proposed method has much better precision than bi-lingual dictionary. Its performance is also very near to the monolingual information retrieval.

1 Introduction

Recently many researchers have interested cross-language information retrieval (CLIR). This is because of widely spread of information across the Internet. Most documents in the WWW today are written in English. Non-native English user who need the information from the internet, need the process of retrieval and then translate using man or machine translation (MT). MT still requires a huge processing time and not available in all languages. Manual translation is also expensive. An efficient CLIR will reduce that cost by retrieving small amount of relevance documents from another language.

The CLIR approach can be divided into two categories:

1. Require translation
 - 1.1. Query only translation
 - 1.2. Document only translation
 - 1.3. Both query and document translation
2. No Translation required
 - 2.1. Sound matching for proper noun.
 - 2.2. Latent semantic indexing(LSI)

For the require translation approach, document or query can be translated using these translation resource:

1. Bi-lingual dictionary – using word by word translation and dis-ambiguation method. [4]
2. Machine translation. (MT) – there is a report on using commercial MT product with a good result for long query translation.[5]
3. Corpus based translation – Parallel corpus can be used as a translation resource. Sentence aligned parallel corpus is very useful for translation [4]. But creating this kind of corpus requires a lot of manual process. Content aligned and comparable corpus is easier to create and had been reported to use as a translation resource.

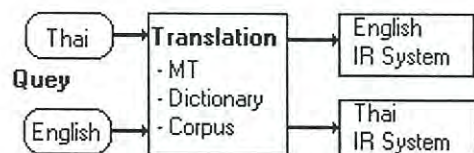


Figure 1. Query translation

Figure 1 show that the query in source language is translated using translation resources and feed to the target language IR system.

In this paper, we use the query only translation approach. The translation resources we used are parallel corpus and bi-lingual dictionary. We compare dictionary mapping method and our proposed modified pseudo-relevance feedback method using parallel corpus.

Previous CLIR for Thai and English language have been proposed by Chuleerat et.al. [2], using the query translation by dictionary approach. But the problem on dictionary coverage, and problem with synonym and transliterate word, made this approach low performance. The transliterate word retrieval by compare the sound coding of proper noun has been attempted by Prasitjutraku Et al. [3]. It learns the word coding from the list of transliterate word. Practically, this is difficult to separate proper noun that embedded into free text because it likely to be small word after it passed to the segmentation process.

2 Pseudo relevance feedback

Relevance feedback has been used to improve performance in monolingual retrieval for a long time. Start by selecting top N relevant documents with query. Then choose the top most relevant terms in document and add to the query.

CLIR, using Pseudo relevance feedback, retrieves relevant documents in source language then choose the top most existing terms in document pair in target language to form a new query in target language. The top most term can be selected using one of these factors.

- 3.1 using number of document that contain term (N)
- 3.2 using number of occurrences of term(F)
- 3.3 using $N * IDF$
- 3.4 using $F * IDF$

3 Thai-English translation resource.

3.1 Bi-lingual dictionary

We prepared bi-lingual dictionary from the popular So-Setapura Thai-English and English-

Thai dictionary [6]. We got 9718 Thai words and 10743 English words, but the problem for So-Setapura dictionary that it is not up to date, new words in computer field does not exist in this dictionary. For example there is no translation of word “database”.

information = ข่าว	ข่าว = advice
ข้อความ	gospel
ข้อมูล	information
สนเทศ	intelligence
ความรู้	item
นิเทศ	news
แจ้งให้ทราบ	story
	tidings
	word

Figure 2. Bi-lingual Dictionary

Figure 2 show the many meanings for one word in Thai-English bi-lingual dictionary prepared from So-Setapura dictionary.

3.2 Parallel text corpus

We found that some magazines are published in two languages; some articles were translated from English to Thai, for example, the Idea Magazine. We had used OCR to make the electronic file from the English part of document. But we manually typed Thai words by hand. In one month one staff can convert about 500 (~ 3MB) document pair from 1998-1999 Idea magazine.

```

- <DOCUMENT>
  <TITLE>Idea</TITLE>
  <YEAR>1998</YEAR>
  <VOLUME>109</VOLUME>
  <PAGE>17</PAGE>
- <ENGLISH>
  <HEAD>The unkindest cut</HEAD>
  <BODY>More women in developing
  countries are having their babies in
  hospitals, but too many are undergoing
  unnecessary surgery, in most cases
  caesareans or episiotomies.</BODY>
  <KEYWORD>unkindest,undergoing,caesareans
  </ENGLISH>
- <THAI>
  <HEAD>การผ่าที่สุดาะไม่อภิรมย์</HEAD>
  <BODY>มีสตรีมีมขินในประเทศกำลังพัฒนาที่คลอด
  บุตรในโรงพยาบาล แต่ก็มีจำนวนมากที่ได้รับการผ่า
  ตัดโดยไม่จำเป็น กรณีส่วนใหญ่แล้วเป็นการผ่าท้อง
  หรือ ผ่าปากของคลอด</BODY>
  <KEYWORD>ไม่อภิรมย์,การผ่าตัด,การผ่าท้อง,ผ่า
  ปากของคลอด</KEYWORD>
  </THAI>
</DOCUMENT>

```

Figure 3. Idea Magazine corpus

Figure 3 the parallel corpus prepared from the Idea Magazine. We had collected the important information such as title, keywords.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

We found the ready to use corpus “Thai Thesis Abstract” in the library CDRom and TIAC web site [http://www.tiac.or.th] that 21,736 of 34,000 document (~ 400MB) contains Thai and English abstract in parallel. Not a sentence by sentence translation as the Idea magazine but it is big enough and useful for our study.

Title	Phytochemical Study of <i>Dysoxylum Grande Hiem</i> Leaves
Abstract	From the leaves extract of <i>Dysoxylum grande Hiem</i> . An alkaloid is isolated and identified by means of chromatographic and crystallization from this species. Their chemical structures and physical properties
Author	Amwawan Srivilai
Major	Master. Science in Pharmacy (Pharmaceutical Chemistry)
Advisor	Asso Prof. Dr. Ekaram Saifah
Institute	Chulalongkorn University, Bangkok (Thailand), Graduate School.
Year	1993
Descriptor	DYSOXYLUM GRANDE / MELIACEAE / PHYTOCHEMICAL
ISBN	974-583-343-6
Page	144 P.

ชื่อวิทยานิพนธ์	การศึกษาทางพฤกษเคมีของไม้ต้นทาสือ
บทคัดย่อ	จากการสกัดแยกสารจากไม้ต้นทาสือ (<i>Dysoxylum grande Hiem</i>) ได้อัลคาลอยด์ ชื่อ rohituzine และ โครโมโน ชื่อ noreugenin สารเคมีทั้ง 2 ศึกษา โครงสร้างทางเคมีและคุณสมบัติทางกายภาพของสารที่สกัดได้
ชื่อนิติกร	อัมพรพรรณ ศรีวิไล
ระดับปริญญา	วิทยานิพนธ์มหาบัณฑิต, เกษตรศาสตร์ (เภสัชเคมี)
ชื่ออาจารย์ที่ปรึกษา	รศ.ดร. เอกรินทร์ สายฟ้า

Figure 4. Thai Thesis Abstract corpus

4 Experimental

We want to see the effectiveness of query translation method between bi-lingual dictionary and PRF method. We measure the precision and recall and compare with monolingual run.

4.1 Test collection

To make a test result in standard precision recall, we need Thai-English Cross-Language test collection but it is not available at that time. We adopt the English test collection from NPL test collection to use in our experiment. [9]

4.1.1 Thai to English query translation, the NPL is a collection of 11,429 document titles. We choose this collection because we know that domain of this collection is related to science and engineering domain as Thai Thesis Abstract corpus. We had translated 10 queries to use in experimental.

4.1.2 English to Thai query translation, we did not measure the performance because test collection is not available but we show the result of some translation compare between translation

methods. The result of the test will be reported later.

4.2 IR System

Thai language has to be segmented before other processing. We already had sub routines for thai word segmentation written in PHP then we decided to use the custom made system instead of popular SMART, but we borrow stopwords and weighting function from SMART. We made live demonstration at <http://test.chaiyo.com>.

In the word segmentation process, we use the maximal matching algorithm [7] with Thai wordlist that made from bi-lingual dictionary combined with Thai wordlist from CTTEX [8].

4.3 Bi-Lingual Dictionary translation

We use word by word translation dictionary mapping without disambiguation, If term has more than one meaning (Figure 2.), We add all meaning to the new query. This translation result is use as the baseline to compare with other method.

4.4 Pseudo Relevance Feedback

We propose the modified pseudo relevance feedback in target language term selection using concept:

“The Number of term in target language is equal or more than in source language.”

We use this relevance feedback rule.

L = number of term in query

D = number of selected document that most relevance to query

N = number of document that contain term in target language

Algorithm:

(1). Retrieve D document

(2). Retrieve terms in target language that exist in document pair of (1)

(3). Rank terms in (2) using $N*IDF$ value

(4). Select at least $L+1$ term to make new query

We use 10% from retrieved document (D) feed to our PRF process. Our experimental use the

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

title from Thai thesis abstract and title from the Idea magazine as a training corpus,

5 Experimental result

We translate the queries to PRF and dictionary lookup. Then we feed the translated queries to the IR system, and measure the precision value at various recalls.

5.1 Thai to English queries translation result

Result acronym description.

Q = Original NPL English query

T = Thai manual translated

D = Dictionary automatic translated

P = The proposed PRF automatic translated

Table 1 Thai to English query translation result

Label	Query
Q1	MATHEMATICAL ANALYSIS AND DESIGN DETAILS OF WAVEGUIDE FED MICROWAVE RADIATIONS
T1	การวิเคราะห์ทางคณิตศาสตร์และการออกแบบรายละเอียดของท่อนำคลื่นสำหรับแพรงส์ไมโครเวฟ
D1	analyse analyze anatomize assay construe criticize delineate diagnose dissect explore riddle scan arithmetic mathematics design plan project style detail specification wave
P1	design analysis wave mathematics mathematical system development model microwave waveguide pipe
Q2	SPHERICAL HARMONIC ANALYSIS OF THE EARTH'S MAGNETIC FIELD
T2	การวิเคราะห์ฮาร์โมนิกทรงกลมของสนามแม่เหล็กโลก
D2	analyse analyze anatomize assay construe criticize delineate diagnose dissect explore riddle scan sphere magnetic field earth globe world
P2	phitsanulok magnetic harmonic province field study world
Q3	SOLUTION OF DIFFERENTIAL EQUATIONS BY COMPUTER
T3	การแก้สมการดิฟเฟอเรนเชียลด้วยคอมพิวเตอร์
D3	counteract cure loose loosen negate

	neutralize ravel release unclasp unloose unravel equation
P3	computer computerized assist instruction lesson system equation stud differential sukxa

Table 1 show 3 translated Thai Queries and output of PRF and dictionary translation.

Table 2. Thai to English efficiency

Precision at	MONO	DICT	PRF
10 docs	0.3250	0.1333	0.2714
20 docs	0.2563	0.0667	0.1625
30 docs	0.2083	0.0666	0.1417
50 docs	0.1600	0.0400	0.1050

Table 2 show the average precision compare between monolingual (MONO), Pseudo relevance feedback (PRF) and Bi-lingual Dictionary

The results from the table 1 show the better translation of PRF over dictionary translation. The number of selected term in dictionary translation has no limit, but our approach in PRF has limited to N+1 term. The precision of PRF is also much better, as shown in the table 2.

Table 3 English to Thai Precision recall

RECALL	MONO	PRF
0.00	1.0000	0.7778
0.10	0.1261	0.0897
0.20	0.0701	0.0497
0.30	0.0504	0.0361
0.40	0.0398	0.0306
0.50	0.0324	0.0268
0.60	0.0282	0.0237
0.70	0.0253	0.0214
0.80	0.0228	0.0187
0.90	0.0214	0.0166
1.00	0.0204	0.0151

As we already know, the CLIR performance is lower than monolingual IR. The results in table 3 show a little drop in precision and recall of our PRF comparing to the monolingual IR.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

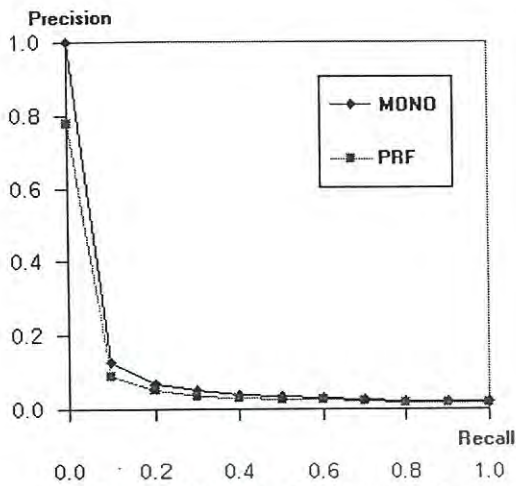


Figure 5. Precision Recall trade off

Figure 5 show that the performance using pseudo relevance feedback is dropped a little, when comparing with monolingual IR system.

5.2 English to Thai queries translation

Results from 5 NPL English Queries

Table 3 English to Thai query translation result

Label	Query
Q4	MATHEMATICAL ANALYSIS AND DESIGN DETAILS OF WAVEGUIDE FED MICROWAVE RADIATIONS
D4	วิภาค รวมทั้ง กำหนด คัด จุดประสงค์ มุ่งหมาย ออกแบบ เจตนา แบบแผน แผนการ โครงการ รายละเอียด ในฐานะ
P4	ออกแบบ คณิตศาสตร์ ไมโครเวฟ แบบจำลอง วิเคราะห์ระบบ คลื่น แบบ
Q5	SPHERICAL HARMONIC ANALYSIS OF THE EARTH'S MAGNETIC FIELD
D5	สอดคล้องกัน วิภาค ในฐานะ พื้นดิน โลก ดึงดูด ขอบเขต ทุ่ง ทุ่งหญ้า สนาม เกษตร แผนก
P5	แม่เหล็ก สนามแม่เหล็ก สนาม ฮาร์ ฮาร์โมนิก ไฟฟ้า นิค โม วิเคราะห์ ตัวนำ
Q6	SOLUTION OF DIFFERENTIAL EQUATIONS BY COMPUTER

D6 คำตอบ ทางออก วิธีแก้ ในฐานะ ซึ่งแตกต่างกัน ตาม ต่อ ทีละ ผ่านไป โดย ใกล้เคียง

P6 สมการ สारละลาย นีเซล ดิฟเฟอ เร

Q7 THE EFFECT OF OXIDATION ON CIRCUIT BREAKER CONTACTS

D7 ทรัพย์สิน ทำให้เกิดผล ประสิทธิภาพ ปรากฏการณ์ ผลกระทบ ยังผล สรรพคุณ อิทธิพล โน้มน้าวใจ ข้าวของ ความ ในฐานะ ตาม ต่อเนื่องต่อไป ประจำ เกี่ยวกับ โดย ในเรื่อง วงจร เครื่องบด ติดต่อ พบปะ สัมผัส

P7 วงจร ออกแบบ มอ ออกซิเดชัน ซี แบบประยุกต์ใช้ กระแส เจริญ

Q8 PRINTED CIRCUIT DESIGN FOR A RANDOM PULSE GENERATOR OF LOW FREQUENCY

D8 จารึก ตัวพิมพ์ ประทับ ผัง พิมพ์ ภาพพิมพ์ ลายพิมพ์ สลัก อัด วงจร กำหนด คัด จุดประสงค์ มุ่งหมาย ออกแบบ เจตนา แบบแผน แผนการ โครงการ สำหรับ เนื่องจาก เพราะว่า เพื่อ แทนที่ จังหวะ ซึ่พจร ถั่ว สะท้อน สั่น เดิน ในฐานะ ค่อยต่ำ พร่อง คู่ม เตี้ย เสว

P8 วงจร ออกแบบ มอ ถี่ ซี แบบ กำเนิด พัลส์ สัญญาณ เจริญต่ำ เครื่อง

At this moment, we do not have Thai test collection to measure the performance in precision and recall. However, our PRF that limits the number of terms in translated query shows good translation in selected topic.

6 Result Discussion

Thai-English Parallel Corpus translation is outperform the dictionary translation and has almost the same performance comparing with monolingual system.

Both translation methods may add unnecessary noise term, see the result P2: "phitsanulok" or in Q3: "cure".

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

ชุดข้อมูลทดสอบ สร้างจากฐานข้อมูลวิทยานิพนธ์ไทย พ.ศ.
2528-2538

Query:1	[2303]
Orig Thai	ความรู้ ทศนคติ และการปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์
Orig Eng	KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION
<i>Relevance Docs</i>	
1	[06702] ความรู้ ทศนคติ และการปฏิบัติเกี่ยวกับโรคเอดส์ของนักศึกษาวิทยาลัยพลศึกษา(18.51)
2	[08263] ความรู้ ทศนคติ และการปฏิบัติเกี่ยวกับโรคเอดส์ ของผู้มารับบริการ ณ คลินิกนิรนาม สภากาชาดไทย(17.62)
3	[06663] การเปิดรับสื่อความรู้ ทศนคติ และพฤติกรรมในการป้องกันโรคเอดส์ ของผู้ปกครองนักเรียนระดับประถมศึกษาปีที่ 6 เขตกรุงเทพมหานคร(15.24)
4	[10372] การเปิดรับสื่อ ความรู้ ทศนคติ และพฤติกรรมการป้องกันโรคเอดส์ของหญิง มีครรภ์ที่มาฝากครรภ์ที่โรงพยาบาลจุฬาลงกรณ์(14.81)
5	[10649] ผลของการสอนสอดแทรกเรื่องโรคเอดส์และการป้องกันโรคเอดส์ในวิชาสังคมศึกษา ที่มีต่อเจตคติต่อการป้องกันโรคเอดส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1(14.61)
6	[02272] พฤติกรรมอนามัยของชาวไทยภูเขาเผ่าม้ง บ้านหนองหอยเก่า ตำบลแม่แรม อำเภอแมริม จังหวัดเชียงใหม่ : ความรู้และ การปฏิบัติตนเกี่ยวกับการป้องกันโรคเอดส์ (13.47)
7	[09912] ความรู้ ทศนคติ และการปฏิบัติเกี่ยวกับการบริโภคเพื่อสุขภาพของนักเรียน ชั้นประถมศึกษาปีที่ 6 ในโรงเรียนสังกัดกรุงเทพมหานคร(13.12)
8	[02742] ทศนคติและพฤติกรรมในการป้องกันโรคเอดส์ของหญิง อาชีพพิเศษ : กรณีศึกษาจังหวัดแม่ฮ่องสอน(12.90)
9	[08111] ความรู้ ทศนคติเรื่องโรคเอดส์ และพฤติกรรมเสี่ยงต่อการติดเชื้อโรคเอดส์ ของผู้นำเยาวชนสาธารณสุขประจำโรงเรียนและนักเรียนมัธยมศึกษาตอน ปลาย จังหวัดนครราชสีมา ปี 2536(12.03)
10	[08103] ความรู้ ทศนคติและพฤติกรรมที่เกี่ยวข้องกับการติดเชื้อโรคเอดส์ในโสเภณีหญิงและชายจังหวัดภูเก็ต(10.91)

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11	[06423] ความรู้และทัศนคติเรื่องโรคเอดส์ ของนักเรียน ระดับมัธยมศึกษาตอนปลาย ของจังหวัดพิษณุโลกปี 2534(10.62)
12	[08090] ความรู้ ทัศนคติและพฤติกรรมเสี่ยงต่อการติดเชื้อโรคเอดส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 เขตการศึกษา 9(10.55)
13	[07778] สภาพและปัญหาการจัดดำเนินงานป้องกันโรคเอดส์ของโรงเรียนเทศบาล (10.11)
14	[07622] ความรู้ ทัศนคติ และพฤติกรรมป้องกันการติดเชื้อโรคเอดส์ของพนักงานและคนงานในโรงงานอุตสาหกรรมในเขต อำเภอปากช่องจังหวัดนครราชสีมา(9.954)
15	[03330] ความรู้ ทัศนคติต่อโรคเอดส์และความตั้งใจรับการทดสอบหา ร่องรอยการติดเชื้อเอชไอวีในหญิงมีครรภ์ที่คลินิกฝากครรภ์ โรงพยาบาลแม่และเด็ก จังหวัดเชียงใหม่ (9.858)
16	[03263] ความรู้และการปฏิบัติในการดูแลทารกของมารดาที่ติดเชื้อ เอชไอวี(7.758)
17	[07888] การวิเคราะห์รูปแบบและเนื้อหาของภาพโฆษณาและแผ่นพับที่ใช้เผยแพร่ความรู้โรคเอดส์สำหรับวัยรุ่น กรุงเทพมหานคร(7.567)
18	[21947] ความรู้ เจตคติ และพฤติกรรมป้องกันการติดเชื้อโรคเอดส์ของนิสิตชาย มหาวิทยาลัยเกษตรศาสตร์(7.470)
19	[02314] ความรู้และการปฏิบัติตนในการป้องกันการติดเชื้อโรคเอดส์ ในผู้ใช้แรงงานชาย โรงงานอุตสาหกรรมธานีนครคอนเดนเซอร์ อำเภอเมือง จังหวัดเชียงใหม่(6.859)
Query:2	[9169]
Orig Thai	การใช้ประโยชน์และความพึงพอใจจากการเปิดรับรายการข่าวทางโทรทัศน์ ของประชาชนในกรุงเทพมหานคร
Orig Eng	THE USE AN GRATIFICATION OF TELEVISION NEWS AMONG PEOPLE IN BANGKOK METROPOLITAN AREA

Relevance Docs

1	[09429] การเปิดรับ การคาดหวังประโยชน์ และความพึงพอใจในรายการธรรมะ ทางสื่อโทรทัศน์ของสมาชิกชมรมทางพระพุทธศาสนาในกรุงเทพมหานคร(23.75)
2	[09809] การใช้ประโยชน์และความพึงพอใจจากการเปิดรับสื่อเกี่ยวกับนักร้อง : กรณีศึกษานักเรียนชั้นมัธยมศึกษาในพื้นที่กรุงเทพมหานคร(21.44)
3	[06725] พฤติกรรมการเปิดรับข่าวสารโฆษณาเกี่ยวกับสิ่งแวดล้อมทางสื่อโทรทัศน์ ของประชาชนในกรุงเทพมหานคร(19.60)

เอกสารนี้เป็นเอกสารที่เผยแพร่ในอินเทอร์เน็ตโดยไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4	[10524] การเปิดรับและความต้องการรายการโทรทัศน์ของผู้สูงอายุในกรุงเทพมหานคร (19.30)
5	[06636] การศึกษาพฤติกรรมการเปิดรับสาร ความรู้ และทัศนคติต่อการเสนอข่าว ต่างประเทศทางสื่อมวลชนของประชาชนในกรุงเทพมหานคร(19.08)
6	[19153] ทัศนคติและพฤติกรรมการเปิดรับชมรายการขายสินค้าตรงทาง โทรทัศน์ของผู้ชมในเขตกรุงเทพมหานคร(17.65)
7	[08041] การเปิดรับข่าวสารสิ่งแวดล้อมทางโทรทัศน์ ความรู้ ความตระหนัก และ การมีส่วนร่วมรักษาคุณภาพสิ่งแวดล้อม ของประชาชนในเขตกรุงเทพมหานคร(17.19)
8	[08122] ความคาดหวังและความพึงพอใจในการเปิดรับข่าวสารจากสื่อ เพื่อการประชาสัมพันธ์ภายในมหาวิทยาลัยรามคำแหงของนักศึกษาและบุคลากร(14.78)
9	[07336] ความคิดเห็นของกลุ่มผู้ผลิตรายการและนักเรียนชั้นประถมศึกษาในโรงเรียนสังกัดกรุงเทพมหานคร เกี่ยวกับรายการโทรทัศน์สำหรับเด็ก(13.24)
10	[19175] การเปิดรับข่าวสาร กับ พฤติกรรมการบริโภคอาหารฟาสต์ฟู้ด แบบตะวันตก ของประชาชนในกรุงเทพมหานคร(13.19)
Query:3	[2091]
Orig Thai	ความสัมพันธ์ระหว่างความรู้ความเข้าใจ ทักษะการคิดคำนวณ และทักษะการแก้โจทย์ปัญหาในวิชาคณิตศาสตร์ ในระดับชั้น ประถมศึกษา
Orig Eng	RELATIONSHIP AMONG UNDERSTAND COMPUTATIONAL AND WORD-PROBLEM SOLVE SKILL IN PRIMARY SCHOOL MATHEMATICS
<i>Relevance Docs</i>	
1	[02102] การออกแบบสร้างชุดการสอนวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษา ปีที่ 2 เรื่อง ความเข้าใจในโจทย์ปัญหาคณิตศาสตร์(22.63)
2	[01386] การวัดผลและการประเมินผลกลุ่มทักษะคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 ของโรงเรียน สังกัดสำนักงานการประถมศึกษาจังหวัด ในโครงการส่งเสริมการประถมศึกษาภาค ตะวันออกที่มีผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์สูงและต่ำ(18.21)
3	[10627] ความสัมพันธ์ระหว่างคุณภาพการสอนของครูและบรรยากาศในห้องเรียนตาม การรับรู้ ของนักเรียนกับแรงจูงใจใฝ่สัมฤทธิ์ต่อวิชาคณิตศาสตร์ของนักเรียนระดับชั้น มัธยมศึกษา(14.10)
เอกสารนี้เป็น 4 กสาร	[06528] ความสัมพันธ์ระหว่าง ความวิตกกังวลในวิชาคณิตศาสตร์ เซาว์น ปัญญา กับภา

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	ผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ของนักเรียนชั้นมัธยม ศึกษาปีที่ 4 ใน กรุงเทพมหานคร(14.04)
5	[06270] ผลของการใช้เครื่องคิดเลขที่มีต่อผลสัมฤทธิ์ในการแก้โจทย์ปัญหาคณิตศาสตร์ ที่มีระดับความยากต่างกัน ของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่มีผลสัมฤทธิ์ ทาง การ เรียนคณิตศาสตร์ต่างกัน(12.41)
6	[05952] การเปรียบเทียบการถ่ายโยงการเรียนรู้ ระหว่างการเรียนรู้ทักษะท่าครอวัล ที่มี ต่อการเรียนรู้ทักษะท่าผีเสื้อ กับการเรียนรู้ ทักษะท่าผีเสื้อ ที่มีต่อการ เรียนรู้ทักษะท่าครอวัลในกีฬา ว่ายน้ำ(12.40)
7	[01389] สภาพและปัญหาการสอนซ่อมเสริมวิชาคณิตศาสตร์ของครูคณิตศาสตร์ชั้น ประถมศึกษาปีที่ 6 ของโรงเรียนในโครงการส่งเสริมการประถมศึกษาภาคตะวันออกเฉียง (12.00)
8	[02518] สมรรถภาพพื้นฐานทางคณิตศาสตร์ที่ใช้ในการแก้โจทย์ปัญหา ของนักเรียนชั้น ประถมศึกษาปีที่ 5 ที่มีผลสัมฤทธิ์ทาง คณิตศาสตร์สูง(11.97)
9	[08363] การเปรียบเทียบความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียน ชั้น ประถมศึกษาปีที่ 4 ที่เรียนโดยใช้ภาพกับไม่ใช้ภาพ(11.85)
10	[01602] การพัฒนาชุดการเรียนการสอนเพื่อซ่อมเสริมการเรียนคณิตศาสตร์เรื่อง การแก้ โจทย์ปัญหา เศษส่วน ของนักเรียนชั้นประถมศึกษาปีที่ 5 ที่มีผลการเรียนคณิตศาสตร์ต่ำ (11.24)
11	[02196] การสร้างแบบทดสอบวินิจฉัยการแก้โจทย์ปัญหาคณิตศาสตร์ สำหรับนักเรียน ชั้นประถมศึกษาปีที่ 5(10.96)
12	[01278] เจตคติต่อวิชาคณิตศาสตร์ของครูสอนคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงาน การประถมศึกษาจังหวัด ในโครงการส่งเสริมการประถมศึกษาภาค ตะวันออก(10.91)
13	[07477] ความสัมพันธ์ระหว่างความรู้เรื่องสิ่งแวดล้อมกับระดับการให้เหตุผลเชิง จริย ธรรมสิ่งแวดล้อม ของนักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียน ประถมศึกษา สังกัด สำนักงานประถมศึกษาจังหวัดเพชรบูรณ์(10.81)
14	[09914] ผลของการร่วมมือและการให้ข้อมูลป้อนกลับเฉลี่ยที่มีผลสัมฤทธิ์ทางการเรียน วิชา คณิตศาสตร์ และความคงทนทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นประถม ศึกษ(10.47)
15	[10317] การพัฒนาวิธีการวินิจฉัยการแก้โจทย์ปัญหาคณิตศาสตร์ที่คำนึงถึงสภาพที่เป็น

เอกสารนี้เป็นเอกสาร
 จริงของ นักเรียนชั้นประถมศึกษาปีที่ 6(10.25)ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรรมใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

16	[06941] ผลของการใช้เทคนิคการสอนของสติฟที่มีต่อผลสัมฤทธิ์ในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 5(10.07)
17	[07522] ผลของการใช้ระบบการสอนของเคลเลอร์ที่มีต่อผลสัมฤทธิ์ในการแก้โจทย์ปัญหา ทางคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6(10.02)
18	[01546] การเปรียบเทียบผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 เรื่องโจทย์ปัญหาร้อยละด้วยวิธีสอนแบบใช้ขั้นตอนชี้แนะความเข้าใจโจทย์ปัญหาตาม แนวของสติฟกับวิธีสอนแบบปกติ(9.974)

Query:4	[1502]
Orig Thai	ทักษะการบริหารของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษา
Orig Eng	ADMINISTRATIVE SKILL OF PRIMARY SCHOOL ADMINISTRATOR UNDER THE OFFICE OF PRIMARY EDUCATION

Relevance Docs

1	[06101] ความคิดเห็นของผู้บริหารและครูอาจารย์เกี่ยวกับทักษะการบริหารการศึกษาของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานการ ประถมศึกษา จังหวัดสมุทรปราการ(18.85)
2	[01375] การบริหารโครงการเสริมสร้างวินัยนักเรียนของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรปราการ(15.48)
3	[02199] การปฏิบัติงานของผู้บริหารโรงเรียนที่ผ่านการอบรมหลักสูตรเตรียม ผู้บริหารโรงเรียนประถม สังกัดสำนักงานการประถมศึกษาจังหวัด อุดรธานี(13.31)
4	[01409] กระบวนการบริหารงานของผู้บริหารโรงเรียนประถมศึกษาดีเด่น สังกัดสำนักงานคณะกรรมการ ประถมศึกษาแห่งชาติ ปีการศึกษา 2532(13.05)
5	[09442] ความต้องการการใช้ไมโครคอมพิวเตอร์ ในการบริหารงานโรงเรียน ประถมศึกษาของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานคณะกรรมการ การประถมศึกษา(13.03)
6	[01479] ปัญหาการนิเทศภายในโรงเรียนของผู้บริหารโรงเรียนประถมศึกษา สังกัด สำนักงานการประถมศึกษาจังหวัดสกลนคร(12.94)
7	[01512] ปัญหาการบริหารงานวิชาการของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงาน การประถมศึกษาจังหวัดลพบุรี(12.45)
8	[01518] ปัญหาการบริหารงานวิชาการ ของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	งาน การประถมศึกษาจังหวัดชลบุรี(12.45)
9	[09566] การศึกษาบทบาทที่ปฏิบัติจริงและบทบาทที่คาดหวังของผู้ช่วยผู้บริหารฝ่ายวิชาการใน โรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัด เขตการศึกษา 1 ตามการรับรู้ของผู้บริหารโรงเรียน ผู้ช่วยผู้บริหารฝ่ายวิชาการ และครูวิชาการโรงเรียน (12.38)
10	[10528] การศึกษาการใช้ทักษะการบริหารงานของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ เขตการศึกษา 1(12.14)
11	[08388] การศึกษาการบริหารงานพัฒนาสิ่งแวดล้อมของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรปราการ(12.08)
12	[10358] การนำเสนอแบบจำลองการพัฒนาภาวะผู้นำสำหรับผู้บริหาร โรงเรียนประถมศึกษา สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ(12.07)
13	[07743] การศึกษาสภาพและปัญหาการบริหารงานของผู้บริหารโรงเรียนประถมศึกษาขนาดเล็ก สังกัดสำนักงานการประถมศึกษา จังหวัดสุรินทร์(12.04)
14	[06024] การประเมินความต้องการจำเป็นเกี่ยวกับการนิเทศงานวิชาการ ในโรงเรียนของผู้บริหารและครูโรงเรียนประถมศึกษาที่เปิดสอน ตามโครงการขยายโอกาสทางการศึกษา สังกัดสำนักงานคณะกรรมการ การประถมศึกษาแห่งชาติ(11.99)
15	[01498] สภาพความต้องการในการพัฒนางานการเรียนการสอนระดับก่อนประถมศึกษาของผู้ บริหารและครู โรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดระยอง(11.97)
16	[01380] การศึกษาสภาพและปัญหาการจัดการศึกษาระดับก่อนประถมศึกษาในโรงเรียน ประถมศึกษาของ ผู้บริหารโรงเรียนและครู สังกัดสำนักงานการประถมศึกษาจังหวัด ปราจีนบุรี(11.96)
17	[09385] การประเมินความต้องการจำเป็นในการฝึกอบรมเกี่ยวกับการวางแผนปฏิบัติการ ประจำปีในโรงเรียนของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงาน คณะกรรมการ การประถมศึกษาจังหวัดอุดรธานี(11.96)
18	[05980] การศึกษาลักษณะการตัดสินใจสั่งการของผู้บริหารโรงเรียน ประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสระบุรี ที่นักเรียนมีผลสัมฤทธิ์ทางการเรียนสูง(11.79)
19	[01345] บทบาทการบริหารงานวิชาการของผู้บริหารโรงเรียนประถมศึกษา ตามทรรศนะของครูวิชาการ กลุ่มโรงเรียนและครูผู้สอน สังกัดสำนักงานการประถมศึกษาจังหวัด นครนายก(11.62)
20	[02198] การบริหารงานวิชาการของผู้บริหารโรงเรียนตามทัศนะของครู โรงเรียนชาวเขา สังกัดสำนักงานการประถมศึกษาจังหวัดลำพูน(11.58)

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการศึกษาวิจัยเท่านั้น ไม่สามารถนำข้อมูลไปใช้ประโยชน์อื่นใดได้

สงวนลิขสิทธิ์ © 2564 โดย สำนักพิมพ์มหาวิทยาลัยราชภัฏวชิราวุธวิทยาลัย

21	[07928] พฤติกรรมภาวะผู้นำของผู้บริหารโรงเรียนประถมศึกษา ตามการรับรู้ของข้าราชการครู สังกัดสำนักงานการประถมศึกษาจังหวัด เขตการศึกษา 1(11.50)
22	[01410] แรงจูงใจในการปฏิบัติงานของครูและผู้บริหารโรงเรียนสังกัดสำนักงาน การประถมศึกษาจังหวัดสระบุรี(11.49)
23	[03370] พฤติกรรมการบริหารงานวิชาการของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษาอำเภอแม่จรม จังหวัดน่าน(11.41)
24	[02471] การบริหารกิจกรรมเสริมหลักสูตรของโรงเรียนสังกัดสำนักงาน การประถมศึกษาอำเภอแม่แตง จังหวัดเชียงใหม่(11.37)
25	[08156] การศึกษาการปฏิบัติงานในหน้าที่ของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัด เขตการศึกษา 12(11.36)
26	[03484] ความรู้และทัศนคติเกี่ยวกับการวางแผนของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษาอำเภอแม่แตง จังหวัดเชียงใหม่(11.25)
27	[09022] ลักษณะที่พึงประสงค์ของผู้บริหารสถานศึกษาในโรงเรียนประถมศึกษา ในโครงการขยายโอกาสทางการศึกษาขั้นพื้นฐาน สังกัดสำนักงาน คณะกรรมการการประถมศึกษาแห่งชาติ เขตการศึกษา 6(11.24)
28	[01599] ปัญหาด้านการใช้กระบวนการวางแผน กระบวนการเรียนการสอนและกระบวนการนิเทศภายในของผู้บริหาร โรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัด เขตการศึกษา 12(11.22)
29	[02460] การรับรู้กระบวนการนิเทศการศึกษาของผู้บริหารโรงเรียน สังกัดสำนักงานการประถมศึกษาจังหวัดหนองคาย(11.18)
30	[01290] ปัญหาการประสานงานในการจัดการศึกษานอกระบบโรงเรียนของผู้บริหารสถานศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดอุดรธานี(11.17)
31	[01472] การบริหารเวลาของผู้บริหารโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษาจังหวัดสมุทรปราการ(11.03)
32	[02833] แนวทางพัฒนาการนิเทศการสอนของผู้บริหารโรงเรียนสังกัด สำนักงานการประถมศึกษาอำเภอเมืองลำปาง จังหวัดลำปาง(10.97)
33	[08149] การศึกษาพฤติกรรมการตัดสินใจสั่งการของผู้บริหารสตรี โรงเรียนประถมศึกษา สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ(10.67)
34	[10795] บทบาทในการสนับสนุนห้องสมุดโรงเรียนของผู้บริหารโรงเรียนประถมศึกษา สังกัดกรุงเทพมหานคร(10.45)
35	[02073] บทบาทที่เป็นจริงและบทบาทที่คาดหวังในการบริหารงานวิชาการ ของผู้บริหาร

เอกสารนี้เป็น 35 สาร

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	โรงเรียน สังกัดสำนักงานการประถมศึกษาจังหวัด อ่างทอง(10.44)
36	[02450] การปฏิบัติงานตามหน้าที่ของหัวหน้าสำนักงานกลุ่มโรงเรียน ประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดลำปาง(10.37)
37	[10736] ผลของการใช้ทักษะการสื่อสารทางวาจาตามแนวคิดกลุ่มปฏิสัมพันธ์ที่มีต่อ ความสามารถทางการพูดของนักเรียนชั้นประถมศึกษาปีที่ 4 ในโรงเรียนสังกัดสำนักงาน การประถมศึกษากรุงเทพมหานคร(10.32)
38	[10117] การเปรียบเทียบวิธีการตัดสินใจสั่งการในงานบริหารวิชาการระหว่างผู้บริหาร หญิง และชายโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา กรุงเทพมหานคร(10.29)
39	[03215] พฤติกรรมการบริหารของผู้บริหารโรงเรียนประถมศึกษา ที่เปิดสอนระดับมัธยม ศึกษาตอนต้น ในเขตการศึกษา 8(10.25)
40	[21354] การปฏิบัติงานนิเทศการศึกษาของผู้บริหารโรงเรียนประถมศึกษา สังกัดสำนัก งานการประถมศึกษาจังหวัดระยอง(10.11)
41	[01340] บทบาทหัวหน้าสำนักงานกลุ่มโรงเรียน สังกัดสำนักงานการประถมศึกษา จังหวัดสิงห์บุรี(10.05)
Query:5	[6209]
Orig Thai	การเพาะเลี้ยงสาหร่ายเกลียวทองเพื่อใช้เป็นอาหาร
Orig Eng	CULTIVATION OF SPIRULINA AS FEED
<i>Relevance Docs</i>	
1	[00545] การสกัดไฟโคไซยานินจากสาหร่ายเกลียวทองเพื่อใช้ เป็นสีผสม อาหาร(27.09)
2	[00468] การพัฒนาถึงปฏิกรณชีวภาพแบบท่อเพื่อเลี้ยงสาหร่าย เกลียวทอง(18.48)
3	[00326] การใช้ก๊าซคาร์บอนไดออกไซด์เป็นแหล่งคาร์บอนในการเลี้ยง สาหร่ายเกลียว ทอง(16.88)
4	[08282] สภาพที่เหมาะสมในการเพาะเลี้ยงสาหร่ายสไปรูไลน่าเพื่อผลิตเบต้าแคโรทีน (12.20)
5	[02154] การเพาะเลี้ยงและปริมาณโปรตีนของสาหร่าย Spirulina platensis ที่เลี้ยงในน้ำ เวย์เต้าหู้(9.794)
6	[02156] คุณค่าทางโภชนาการบางประการของสาหร่าย Spirulina platensis ที่เลี้ยงใน น้ำกากส่าเหล้า(5.079)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:6	[02799]
Orig Thai	อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง
Orig Eng	THE INFLUENCE OF SUPPLEMENTARY NITROGEN FERTILIZE APPLY AT EAR POD FORMATION ON THE QUALITY AN YIELD OF SOYBEAN

Relevance Docs

1	[02404] ความผันแปรของผลผลิตและคุณภาพผลผลิตถั่วเหลืองฝักสดพันธุ์ ต่าง ๆ ภายใต้ระยะเวลาปลูกและอัตราปุ๋ยไนโตรเจนที่แตกต่างกัน(22.55)
2	[02668] การจัดการปุ๋ยไนโตรเจนในถั่วเหลืองฝักสด(20.76)
3	[02227] ผลกระทบของการปลูกถั่วเหลือง และถั่วเขียวผิวมันที่มีต่อ การใช้ไนโตรเจน และผลผลิตของข้าวโพดที่ปลูกตามหลัง(11.22)
4	[03000] การตรึงและการสะสมไนโตรเจนของถั่วเหลือง ในแต่ละระดับการ ใส่ปุ๋ยไนโตรเจน และความหนาแน่นของต้นปลูก(10.50)
5	[02128] การตรึงไนโตรเจนของถั่วเหลืองพันธุ์ต่าง ๆ ภายในฤดูปลูก ที่แตกต่างกัน (8.900)
6	[02137] การตรึงไนโตรเจนในถั่วเหลืองที่ปลูกในระบบตามหลังข้าว(8.614)
7	[02136] การปรับตัวและการตรึงไนโตรเจนในถั่วเหลืองปลูกในสภาพดินอ้อมตัว ด้วยน้ำ (8.588)
8	[02402] ผลกระทบของวันปลูกที่มีต่อการตรึงไนโตรเจน การเจริญเติบโต และการให้ผลผลิตของถั่วเหลือง(8.234)

Query:7	[6007]
Orig Thai	การดูแลสุขภาพของผู้สูงอายุ
Orig Eng	HEALTH CARE OF THE AGE

Relevance Docs

1	[02261] ความรู้และการปฏิบัติตนเกี่ยวกับสุขภาพอนามัยของชาวบ้านสันโป่ง ตำบลบ้านกาด กิ่งอำเภอแม่วาง จังหวัดเชียงใหม่ : ความรู้และ การปฏิบัติในการดูแลสุขภาพอนามัยตนเองของผู้สูงอายุ(14.03)
2	[03445] พฤติกรรมการดูแลสุขภาพของผู้สูงอายุในชุมชนแออัดเขตเมือง(12.81)

เอกสารนี้เป็นเอกสารทสวงวนสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3	[02296] พฤติกรรมการดูแลตนเองของผู้สูงอายุในเขตเทศบาลเมือง จังหวัดพิษณุโลก (12.59)
4	[02714] ความสัมพันธ์ระหว่างพฤติกรรมสุขภาพและภาวะสุขภาพในผู้สูงอายุ(10.17)
5	[03004] การแก้ไขปัญหาสุขภาพผู้สูงอายุ โดยการจัดตั้งชมรมผู้สูงอายุ บ้านวังทอง ตำบลหนองหญ้าปล้อง อำเภอวังสะพุง จังหวัดเลย(9.578)
6	[07539] การศึกษาพฤติกรรมจริยธรรมทางสังคมของผู้สูงอายุในชมรมผู้สูงอายุและสถานสงเคราะห์คนชรา กรุงเทพมหานคร(9.517)
7	[06474] ผลของการมีส่วนร่วมในกลุ่มช่วยเหลือตนเองต่อการเห็นคุณค่า ในตนเองและการปฏิบัติกิจกรรมของผู้สูงอายุ(9.464)
8	[02910] ความสัมพันธ์ระหว่าง ความเชื่อในอำนาจด้านสุขภาพ ความเชื่อ ในประสิทธิภาพแห่งตนและพฤติกรรมสุขภาพ ของผู้สูงอายุ(8.732)
9	[06400] ภาวะสุขภาพและการรักษาพยาบาลของผู้สูงอายุไทย(8.359)
10	[09120] การสำรวจภาวะของครอบครัวในการดูแลผู้สูงอายุสมองเสื่อมที่บ้าน(8.319)
11	[02257] ความสัมพันธ์ระหว่างความว่าเหวและภาวะสุขภาพของผู้สูงอายุ(8.272)
12	[06044] การเลือกใช้สถานบริการสุขภาพอนามัยของผู้สูงอายุไทย(8.272)
13	[02295] ความสัมพันธ์ระหว่างความวิตกกังวลและภาวะสุขภาพของผู้สูงอายุ ในสถานสงเคราะห์คนชรา(8.096)

Query:8	[9251] , [2578]
Orig Thai	การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อวางแผนการใช้ที่ดิน
Orig Eng	THE APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM FOR LAND USE PLAN

Relevance Docs

1	[09947] การประเมินทางธรณีวิทยาสิงแวดล้อมโดยใช้ระบบสารสนเทศภูมิศาสตร์ เพื่อการวางแผนการใช้ที่ดินในพื้นที่จังหวัดสระบุรี ประเทศไทย(21.21)
2	[08972] การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ เพื่อการจำแนกความเหมาะสมของที่ดินสำหรับพืชเศรษฐกิจ : กรณีศึกษา จังหวัดชลบุรี(19.20)
3	[09366] การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในงานวางแผนโครงสร้างจังหวัด: กรณีศึกษาจังหวัดจันทบุรี(13.69)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4	[09405] การประยุกต์ใช้การจัดการจัดรูปที่ดินเพื่อพัฒนาที่ดินสำนักงานทรัพย์สินส่วนพระมหากษัตริย์: กรณีศึกษา ที่ดินบริเวณคลองพลับพลา พระรามที่ 9 กรุงเทพมหานคร(13.64)
5	[02576] การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการศึกษาระดับการ พัฒนาของพื้นที่บริเวณลุ่มน้ำปิงตอนบน(13.16)
6	[09043] การประยุกต์ใช้ระบบสารสนเทศทางภูมิศาสตร์ในการหาพื้นที่ที่ได้รับ ผลกระทบจากก๊าซซัลเฟอร์ไดออกไซด์ที่เกิดจากโรงงานไฟฟ้าแม่เมาะ(12.53)
7	[07922] การประยุกต์ระบบสารสนเทศภูมิศาสตร์เพื่อการวางแผนก่อนการลุ่มน้ำในประชากร:กรณีศึกษา-เทศบาลเมืองชลบุรี(10.82)
8	[08987] การใช้ระบบสารสนเทศภูมิศาสตร์ในการกำหนดเขตของหน่วยเลือกตั้ง ในเขตเทศบาลเมืองชลบุรี(10.15)
9	[02753] การสร้างระบบสารสนเทศเชิงพื้นที่ สำหรับการประเมินคุณภาพ ที่ดินเพื่อการเกษตร ในจังหวัดเชียงใหม่(9.793)
10	[09435] ระบบสารสนเทศภูมิศาสตร์แบบโต้ตอบสำหรับการท่องเที่ยว : กรณีศึกษา จังหวัดกาญจนบุรี(9.693)
Query:9	[2119]
Orig Thai	การย่อยสลายแป้งมันสำปะหลัง โดยใช้เอนไซม์
Orig Eng	HYDROLYSIS OF CASSAVA STARCH BY ENZYME
<i>Relevance Docs</i>	
1	[08632] การผลิตมอลโตเดกซ์ทรินจากแป้งมันสำปะหลังโดยใช้ เอนไซม์แอลฟา-อะมิเลสที่ทนความร้อน(13.91)
2	[21896] การดัดแปรสตาarch ในแป้งมันสำปะหลังและข้าวด้วยวิธีอะซิทิเลชัน(10.42)
3	[21457] การปรับปรุงคุณสมบัติของแป้งมันสำปะหลังโดยวิธีการแปรสภาพทางกายภาพและทางเคมีแบบเชื่อมข้าม (Cross-linking) ร่วมกับวิธีอะซิทิเลชัน (Acetylation) และพรีเจลลิตไนเซชัน (Pregelatinization)(10.22)
4	[10607] การผลิตน้ำตาลรีดิวซ์จากการมันสำปะหลังโดยการให้เอนไซม์และอัลตราฟิลเทรชัน(8.421)
5	[08123] การผลิตโปรตีนเซลล์เดียวจากมันสำปะหลังโดยยีสต์ผสมระหว่าง <i>Endomycopsis fibuligera</i> และ <i>Candida utilis</i> (7.171)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ทางการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query:10	[8247]
Orig Thai	พฤติกรรมของแผ่นพื้นคอนกรีตท้องเรียบชนิดอัดแรงบางส่วน
Orig Eng	BEHAVIOUR OF PARTIAL PRESTRESS CONCRETE FLAT PLATE
<i>Relevance Docs</i>	
1	[07905] กำลังของแผ่นพื้นคอนกรีตท้องเรียบทางเดียวชนิดอัดแรงบางส่วน(44.96)
2	[09999] พฤติกรรมการดัดของแผ่นพื้นท้องเรียบหล่อสำเร็จเทบหน้าคอนกรีตไฟเบอร์โพลีโพรพีลีน(27.38)
3	[02789] การประยุกต์ไมโครคอมพิวเตอริ์ในการวิเคราะห์และออกแบบพื้นไร้ คานอัดแรง ภายหลังแบบอัดแรงบางส่วนและอัดแรงเต็มกำลัง(24.87)
4	[06453] การวิเคราะห์โครงสร้างแผ่นพื้นท้องเรียบด้วยวิธีอย่างง่าย(21.50)
5	[10624] การศึกษาเพื่อเสนอแนวทางออกแบบชิ้นส่วนโครงสร้างคอนกรีตเสริมเหล็กโดยอาศัย แบบจำลองด้วยแรงอัดและแรงดึงภายใน(16.15)
6	[10520] การศึกษากำลังรับแรงเฉือนของแผ่นพื้นไร้คานที่มีช่องเปิดเสริมด้วยหมุนรับแรงเฉือน(14.71)
7	[06121] การใช้วิธีบาวดารีอินทิกรัลวิเคราะห์แผ่นพื้นรูปเหลี่ยมไร้คาน(9.828)
8	[00346] การวิเคราะห์แผ่นพื้นสำเร็จรูปโดยวิธีโครงสร้างถ่ายภายใต้้น้ำ หนักบรรทุกแบบจุด (9.442)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

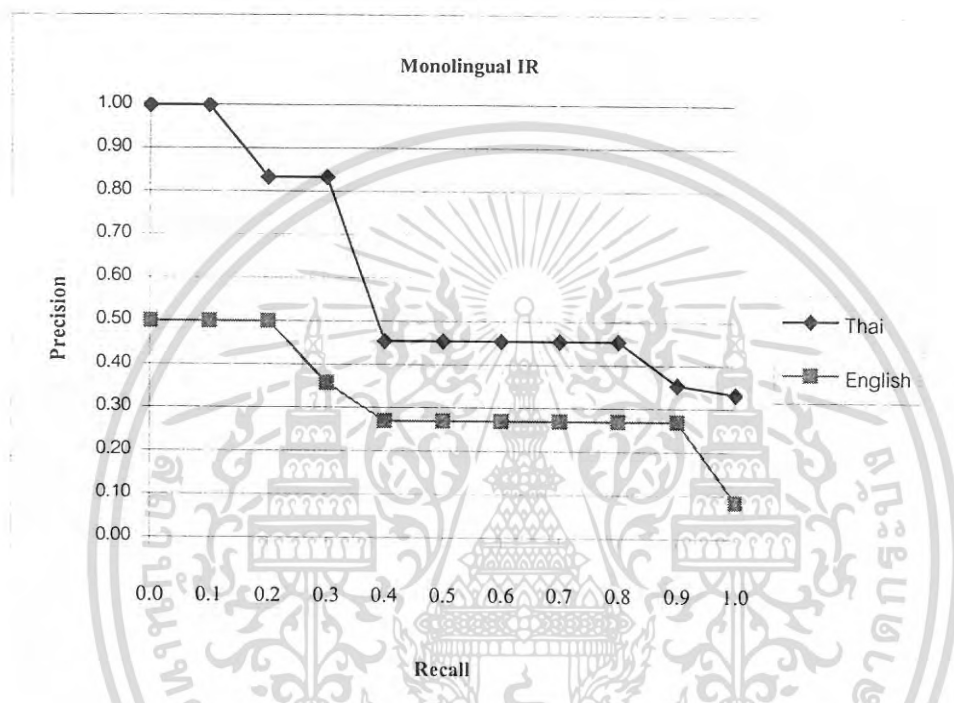
ภาคผนวก ค

ผลการทดสอบระบบค้นคืนสารสนเทศภาษาเดียว

Query1

ความรู้ ทักษะ และ การปฏิบัติ เกี่ยวกับการป้องกันโรคเอดส์
--

KNOWLEDGE, ATTITUDE AN PRACTICE IN AIDS PREVENTION
--



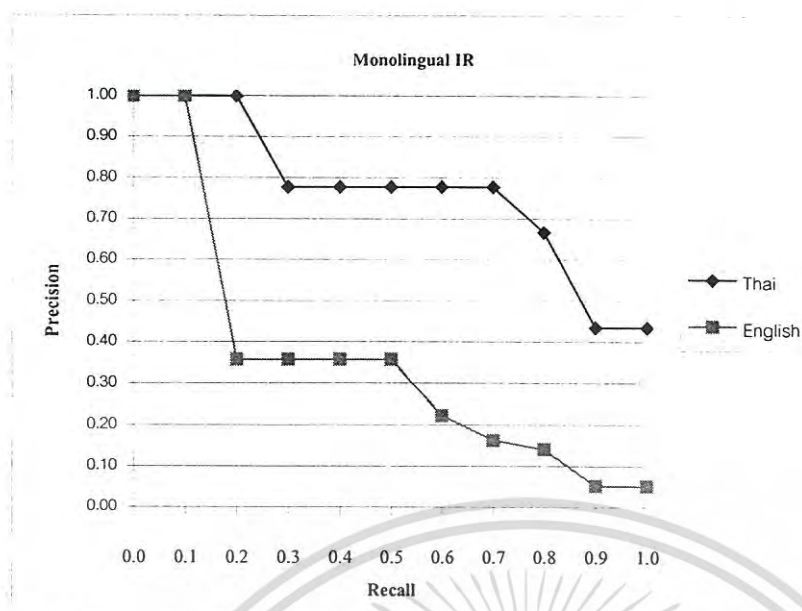
Query2

การใช้ประโยชน์และความพึงพอใจจากการเปิดรับรายการข่าวทางโทรทัศน์ ของประชาชนในกรุงเทพมหานคร
--

THE USE AN GRATIFICATION OF TELEVISION NEWS AMONG PEOPLE IN BANGKOK

METROPOLITAN AREA

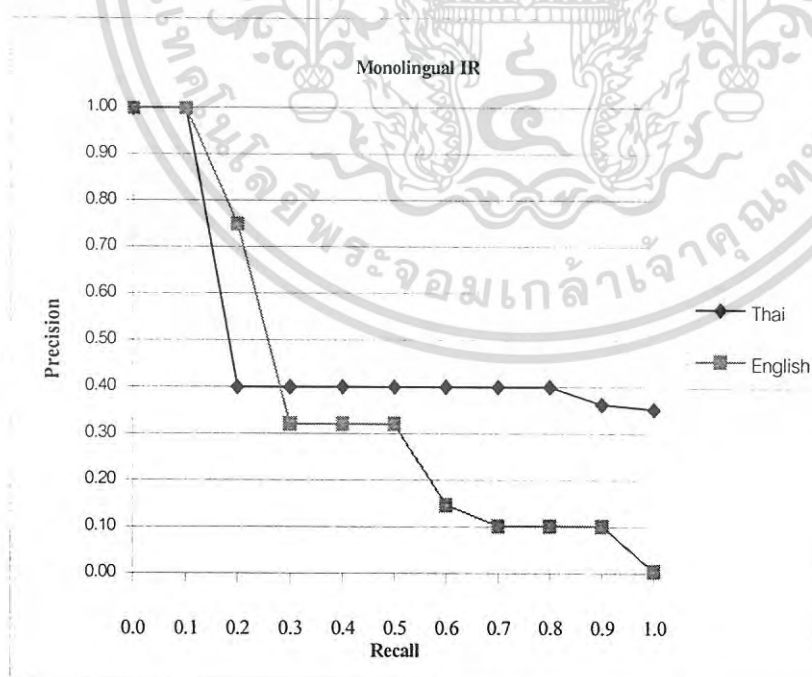
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Query3

ความสัมพันธ์ระหว่างความรู้ความเข้าใจ ทักษะการคิดคำนวณ และทักษะการแก้โจทย์
ปัญหาในวิชาคณิตศาสตร์ ในระดับชั้น ประถมศึกษา

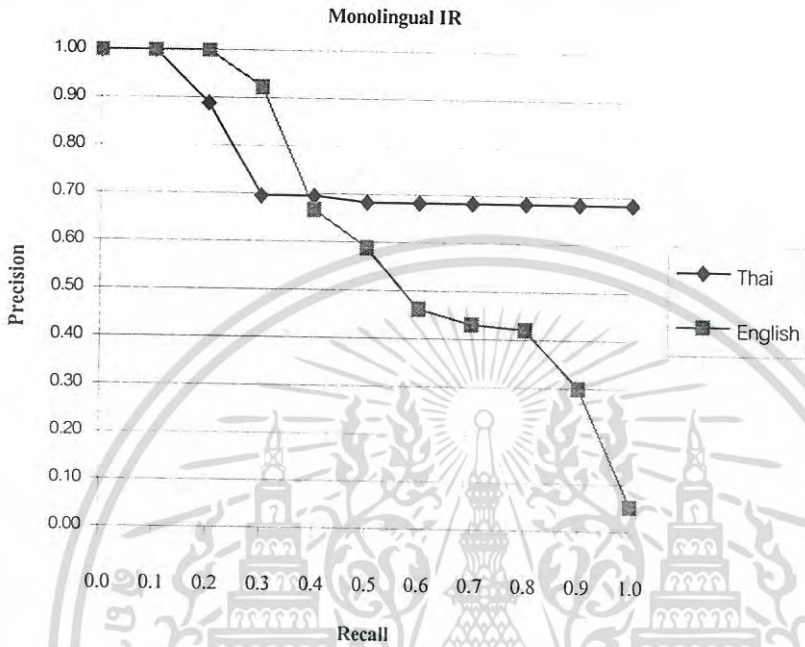
RELATIONSHIP AMONG UNDERSTAND COMPUTATIONAL AND WORD-
PROBLEM SOLVE SKILL IN PRIMARY SCHOOL MATHEMATICS



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

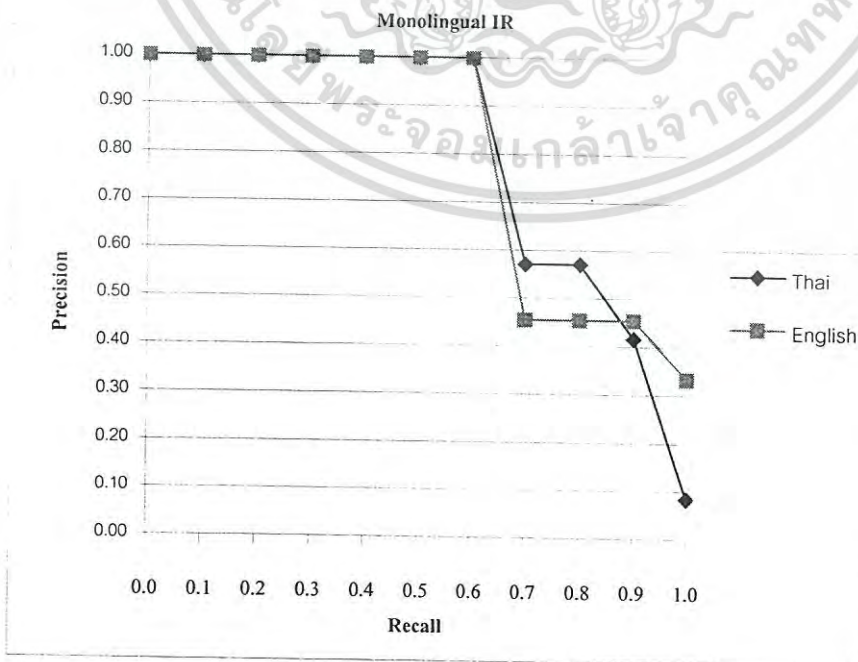
Query4

ทักษะการบริหารของผู้บริหารโรงเรียน สังกัดสำนักงานการศึกษา
 ADMINISTRATIVE SKILL OF PRIMARY SCHOOL ADMINISTRATOR UNDER THE OFFICE OF
 PRIMARY EDUCATION



Query5

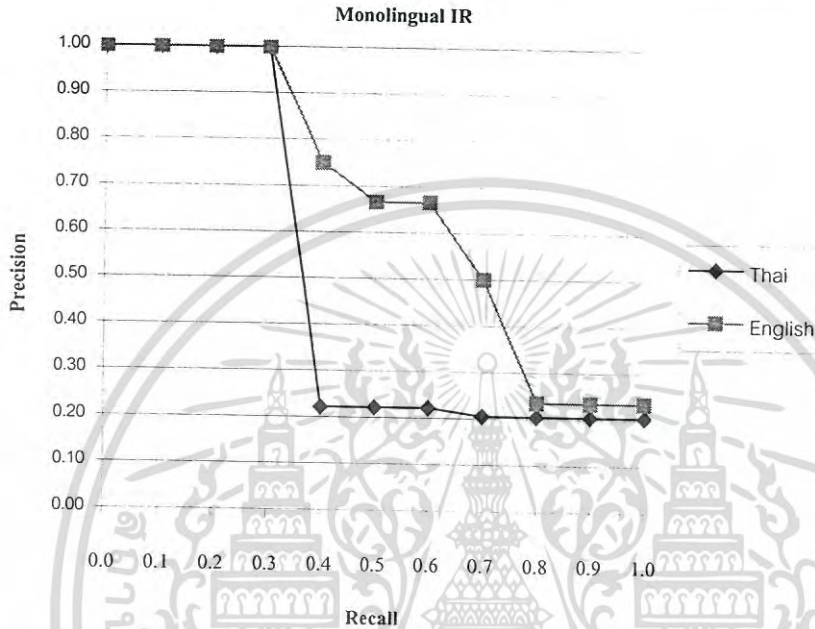
การเพาะเลี้ยงสาหร่ายเกลียวทองเพื่อใช้เป็นอาหาร
 CULTIVATION OF SPIRULINA AS FEED



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

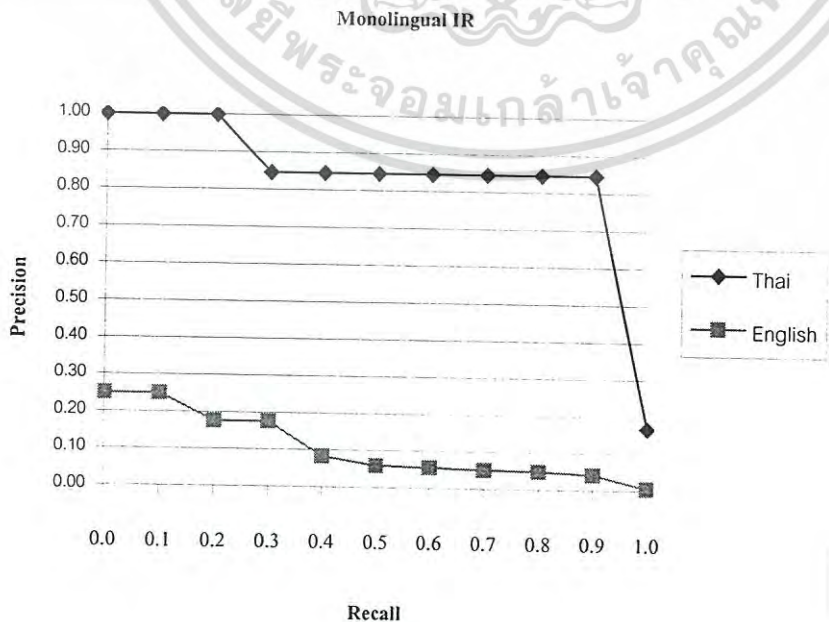
Query6

อิทธิพลของการเสริมปุ๋ยไนโตรเจนในระยะเริ่มติดฝักต่อคุณภาพ และผลผลิตของถั่วเหลือง
 THE INFLUENCE OF SUPPLEMENTARY NITROGEN FERTILIZE APPLY AT EAR POD
 FORMATION ON THE QUALITY AN YIELD OF SOYBEAN



Query7

การดูแลสุขภาพของผู้สูงอายุ
 HEALTH CARE OF THE AGE

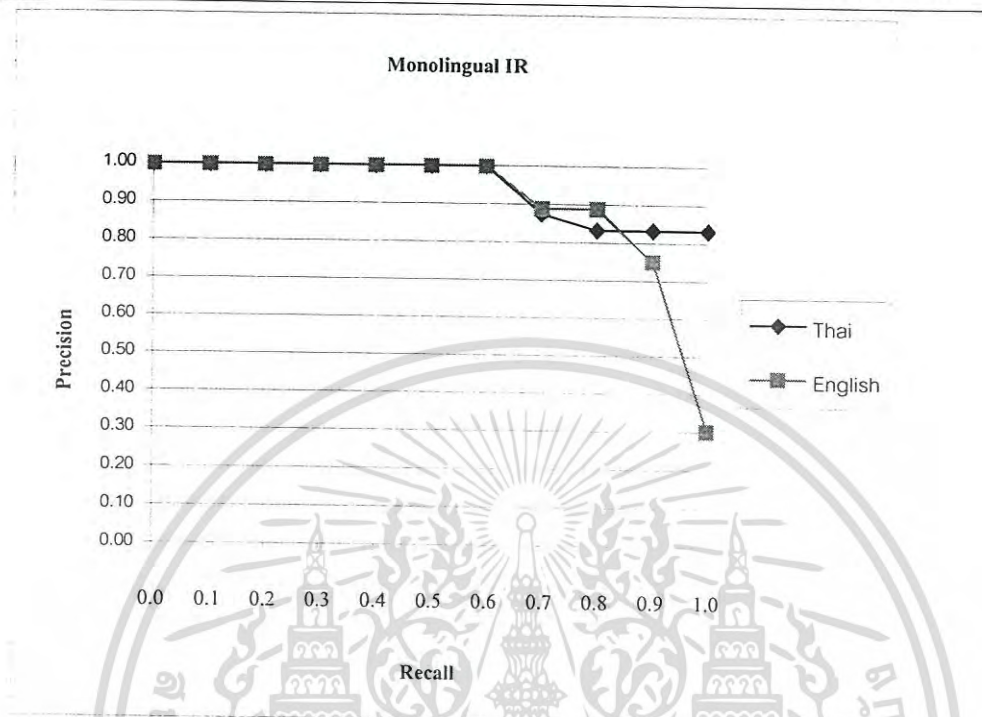


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query8

การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อวางแผนการใช้ที่ดิน

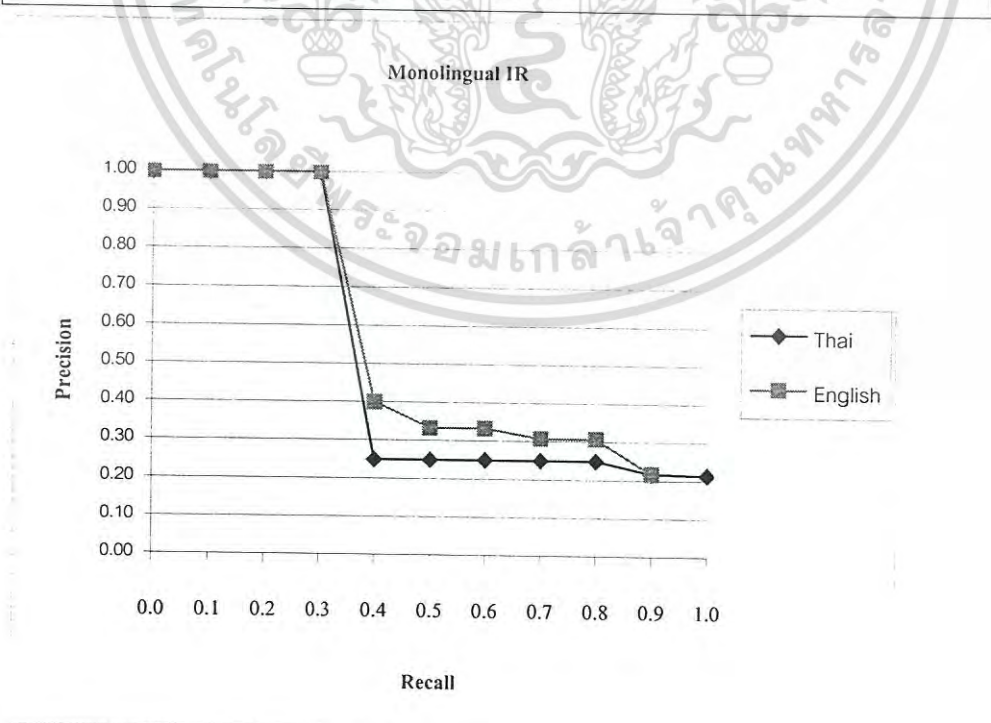
THE APPLICATION OF GEOGRAPHIC INFORMATION SYSTEM FOR LAND USE PLAN



Query9

การย่อยสลายแป้งมันสำปะหลัง โดยใช้เอนไซม์

HYDROLYSIS OF CASSAVA STARCH BY ENZYME

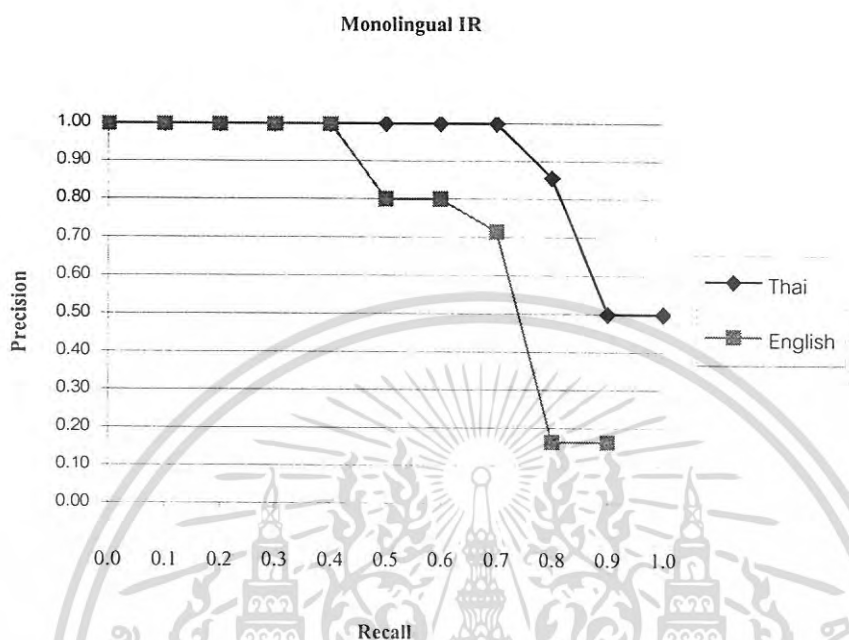


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Query10

พฤติกรรมของแผ่นพื้นคอนกรีตต้องเรียบชนิดอัดแรงบางส่วน

BEHAVIOUR OF PARTIAL PRESTRESS CONCRETE-FLAT PLATE



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นายณัฐนันท์ ศรีสะอาด เกิดเมื่อวันที่ 30 พฤษภาคม 2517 ที่จังหวัดปราจีนบุรี สำเร็จ การศึกษาวิทยาศาสตร์บัณฑิต (เคมี) จากมหาวิทยาลัยบูรพา ปีการศึกษา 2540

ปี 2541 ทำงานเป็นนักวิเคราะห์ที่บริษัท แอดวานซ์อะโกร จำกัดมหาชน ปี 2542 ทำธุรกิจ บริการอินเทอร์เน็ตในจังหวัดชลบุรี และเปิดบริการค้นหาข้อมูลภาษาไทยบนเครือข่ายอินเทอร์เน็ต เป็นรายแรกในชื่อไทยเสิร์ช ปี พ.ศ. 2543 ทำธุรกิจบริการอินเทอร์เน็ตในจังหวัดฉะเชิงเทราและเปิด บริการอีเมลฟรีภาษาไทยชื่อไชโยเมลล์ (www.chaiyo.com) ปี พ.ศ. 2544 ทำธุรกิจบริการอินเทอร์เน็ตในจังหวัดระยอง ปี พ.ศ. 2546 ทำธุรกิจให้คำปรึกษาและติดตั้งระบบเครือข่าย ชื่อ บ. เอมเบส เทคโนโลยี จก.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้