

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การสอบถามข้อมูลการท่องเที่ยวโดยใช้ภาษาธรรมชาติ

Natural Language Query for Tourism

โดย

นายกิตติวุฒิ กรอบพานิชวงศ์

รหัส 42067060

H001769

อาจารย์ที่ปรึกษา

ดร. รัฐการ อภิวัฒน์วาจา

วัน เดือน ปี..... 09 มิ.ย. 2550
เลขทะเบียน..... 01769
เลขเรียกหนังสือ..... อพ. ก ๒74ก 2543
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2543

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การสอบถามข้อมูลการท่องเที่ยวโดยใช้ภาษาธรรมชาติ
นักศึกษา	นายกิตติวุฒิ กรอบพานิชวงศ์
อาจารย์ที่ปรึกษา	ดร.รัฐการ อภิวัฒน์วาท
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2543

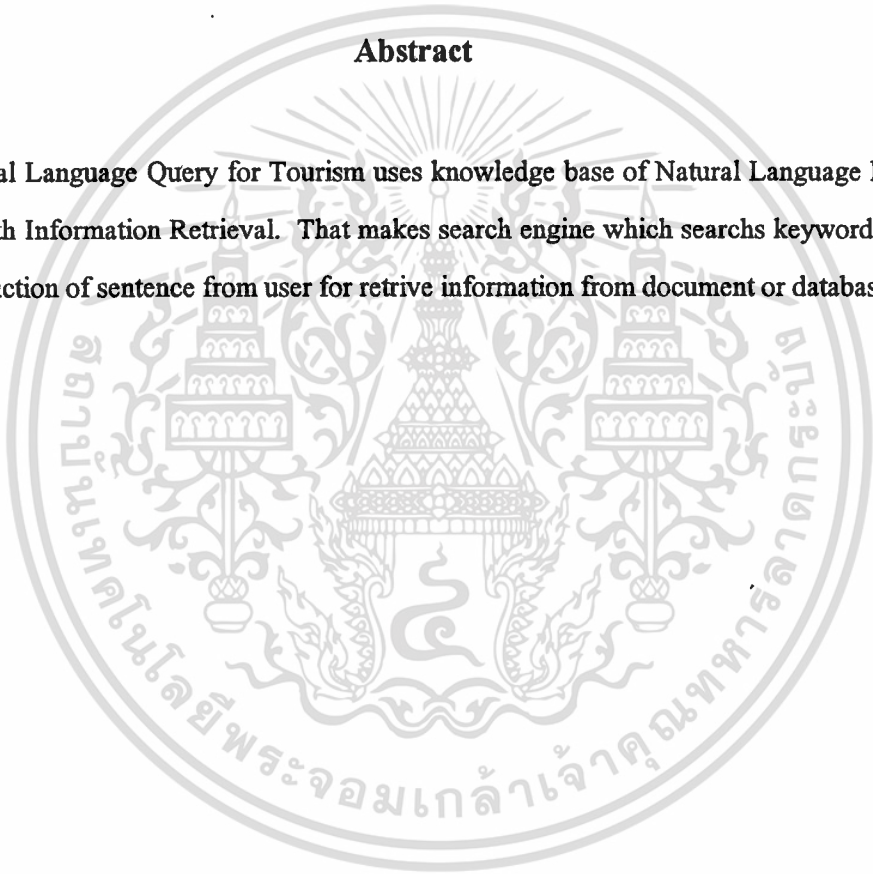
บทคัดย่อ

โครงการพัฒนาระบบงาน Natural Language Query for Tourism เป็นการนำความรู้ด้าน Natural Language Processing มาประยุกต์ใช้กับ Information Retrieval เพื่อสร้าง Search Engine ที่สามารถรับ Query ที่เป็นประโยคคำถามภาษาอังกฤษ หรือประโยคคำสั่งจากผู้ใช้ และนำไปวิเคราะห์เพื่อหา Keyword เพื่อค้นหาข้อมูลจาก Document หรือ Database ที่ผู้จัดทำรวบรวมไว้ซึ่งเป็นข้อมูลเกี่ยวกับการท่องเที่ยว และข้อมูลที่สำคัญสำหรับนักท่องเที่ยวเกี่ยวกับเมืองไทย

Title Natural Language Query for Tourism
Student Mr.Kittiwut Krobpanichwong
Advisor Rattakarn Apiwatwaja, Ph.D.
Level of Study Master of Science in Information Technology
Major Information Science
Academic Year 2000

Abstract

Natural Language Query for Tourism uses knowledge base of Natural Language Processing for applies with Information Retrieval. That makes search engine which searches keyword in english query or instruction of sentence from user for retrieve information from document or database of travel integration.



กิตติกรรมประกาศ

โครงการ Natural Language Query for Tourism จัดทำขึ้นเพื่อศึกษาการ Query โดยใช้ Natural Language ในการ Retrieve ข้อมูลจาก Documents ใน Search Engine แทน Boolean Logic Query ซึ่งทำให้ลดเวลาสำหรับผู้ใช้ในการค้นหาข้อมูลที่ต้องการ และลดจำนวนผลลัพธ์ที่ไม่ต้องการ

ทั้งนี้ผู้จัดทำโครงการต้องขอขอบพระคุณ อาจารย์รัฐกร อภิวัฒน์วาจา ซึ่งเป็นอาจารย์ที่ปรึกษาโครงการ ที่ได้กรุณาให้คำปรึกษา และชี้แนะให้โครงการดำเนินไปด้วยดี



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อ III ศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญภาพ	VI
บทที่	
1. บทนำ	1
1.1 ความเป็นมาของโครงการ	1
1.2 วัตถุประสงค์ของโครงการ	1
1.3 ปัญหา และอุปสรรค	2
1.4 ขอบเขตของโครงการ	2
2. ความรู้เบื้องต้นเกี่ยวกับการแจกแจงประโยค	3
2.1 Context Free Grammar	3
2.2 การแจกแจงประโยค	9
2.3 ข่ายงานและการเวียนซ้ำ	12
3. ระบบการสอบถามข้อมูลการท่องเที่ยวโดยใช้ภาษาธรรมชาติ	14
3.1 โครงสร้างของระบบ	14
3.2 แนวทางในการพัฒนาระบบ	16
3.3 การวิเคราะห์ และออกแบบระบบ	17
4. โปรแกรม	41
4.1 Hardware Requirements	41
4.2 System Requirements	41
4.3 Installation	41

5. การทดลอง และผลลัพธ์	47
5.1 วัตถุประสงค์ของการทดลอง	47
5.2 Environment	47
5.3 การทดลอง	48
6. บทสรุป	54
ข้อสรุป และผลที่ได้จากโครงการ	54



สารบัญภาพ

ภาพที่	หน้า
2-1 เครื่องจักรสถานะจำกัด	5
2-2 โครงสร้างวลีของประโยค	7
2-3 แสดงข่างานการเปลี่ยนแปลงของนามวลี	12
2-4 แสดงข่างานไวยากรณ์	13
3-1 Structure Chart	17
3-2 Context Diagram	18
3-3 DFD Level 1	19
3-4 DFD Level 2 of Process 1.0	20
3-5 DFD Level 2 of Process 2.0	21
3-6 DFD Level 2 of Process 3.0	22
3-7 DFD Level 2 of Process 4.0	23
3-8 DFD Level 2 of Process 5.0	24
4-1 แสดงหน้าจอแรกของระบบ	42
4-2 ส่วนรับข้อมูลจากผู้ใช้	43
4-3 แสดงตัวอย่างของการกรอกข้อความลงในฟอร์ม	44
4-4 ในกรณีที่ผู้ใช้ไม่ได้กรอกข้อความแต่กดปุ่มส่งข้อความระบบจะขึ้นข้อความเตือน	45
4-5 เมื่อทำการส่งข้อความเรียบร้อยแล้วโปรแกรมก็จะส่งค่าที่ประมวลผลได้กลับมา	46

บทที่ 1

บทนำ

1.1 ความเป็นมาของโครงการ

โครงการเรื่อง Natural Language Query เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงานโดยมุ่งเน้นเพิ่มประสิทธิภาพของการ Query เพื่อให้ได้ Information ที่ต้องการของผู้ใช้โดยใช้เวลาในการ Search น้อยที่สุดซึ่งจากเดิมการ Retrieve ข้อมูลจาก Documents ใน Search Engine ต้องใช้ Keyword และ Boolean logic ช่วยในการ Query ซึ่งผลลัพธ์ที่ Search Engine แสดงออกมาค่อนข้างมาก และมีทั้งข้อมูลที่ตรงตามความต้องการของผู้ใช้ และไม่ตรงตามความต้องการ ซึ่งทำให้ผู้ใช้อาจจะต้องใช้ keyword คำใหม่ในการ Query จนกว่าจะได้ Information ที่ต้องการ และจากการศึกษายังพบอีกว่ามีผู้ใช้บางกลุ่มยังไม่มีความรู้เกี่ยวกับ Boolean Logic เท่าที่ควร ในการ Query ถ้า Expression ที่ Query ผิด ผลลัพธ์ที่แสดงออกมาจะไม่ตรงความต้องการของผู้ใช้

ผู้จัดทำจึงเกิดแนวความคิดที่จะปรับปรุงประสิทธิภาพ Query และความง่ายในการใช้งานสำหรับผู้ใช้ โดยใช้ความรู้เกี่ยวกับ Natural Language Processing (NLP) ซึ่งเป็นสาขาหนึ่งของวิชา AI มาใช้กับ Query ของผู้ใช้ในการ Retrieve Documents ซึ่งผู้จัดทำกำหนดขอบเขตของเนื้อหาเกี่ยวกับข้อมูลการท่องเที่ยวในประเทศไทย เพื่อที่จะเป็นประโยชน์สำหรับนักท่องเที่ยว หรือชาวต่างประเทศที่ต้องการทราบข้อมูลเกี่ยวกับประเทศไทย

1.2 วัตถุประสงค์ของโครงการ

โครงการ Natural Language Query จัดทำขึ้นเพื่อวัตถุประสงค์หลักดังต่อไปนี้

1. เพื่อเพิ่มประสิทธิภาพในการ Query ข้อมูลจาก Search Engine โดยลดจำนวนผลลัพธ์ที่ไม่ต้องการให้น้อยที่สุด
2. เพื่อสร้าง Search Engine ที่ง่ายต่อการใช้งานโดยรูปแบบของการ Query ที่เป็นแบบ Natural Language ที่เป็นประโยคคำถามภาษาอังกฤษ
3. เพื่อรวบรวมข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทย สำหรับเผยแพร่การท่องเที่ยว และผู้ที่ต้องการข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทย

1.3 ปัญหาและอุปสรรค

โครงการ Natural Language Query for Tourism ได้เกิดจากปัญหาดังต่อไปนี้

1. รูปแบบของ Query ในปัจจุบันส่วนใหญ่จะอยู่ในรูปของ Keyword และ Boolean Logic ซึ่งผลลัพธ์ที่ได้จากการค้นหาค่อนข้างมาก ทำให้เสียเวลาในการค้นผลลัพธ์ที่ต้องการ
2. ใน Search Engine ที่สำคัญบน Internet ยังไม่มีข้อมูลเกี่ยวกับประเทศไทยรวบรวมไว้ใน Database มากพอ
3. Boolean Logic มีปัญหาในกรณีที่ผู้ใช้ไม่มีความรู้ด้าน Boolean Logic หรือ ในกรณีที่ผู้ใช้มีความรู้แต่ Boolean Query Statement มีความซับซ้อนอาจจะทำให้ผู้ใช้สับสนเองได้

1.4 ขอบเขตของโครงการ

ในโครงการ Natural Language Query ได้ตั้งขอบเขตของโครงการไว้ดังต่อไปนี้

1. พัฒนา Search Engine ที่สามารถรองรับ Query ที่มีลักษณะเป็น Natural Language ที่เป็นประโยคคำถามภาษาอังกฤษที่เป็น WH question และประโยคคำสั่ง ที่มีโครงสร้างไวยากรณ์ไม่ซับซ้อนมากนัก
2. โดยเนื้อหาที่ตั้งคำถาม หรือค้นหาจะเกี่ยวกับข้อมูลการท่องเที่ยวที่สำคัญสำหรับนักท่องเที่ยว เช่น สถานที่ท่องเที่ยวที่สำคัญ อาหาร ที่พัก วัฒนธรรม
3. รวบรวมข้อมูลที่สำคัญเกี่ยวกับประเทศไทยที่จำเป็นสำหรับนักท่องเที่ยว

บทที่ 2

ความรู้เบื้องต้นเกี่ยวกับการแจกแจงประโยค

2.1 Context Free Grammar

การประมวลผลภาษาธรรมชาติ (Natural Language Processing — NLP) เพื่อให้เครื่องจักรหรือคอมพิวเตอร์สามารถรับรู้และเข้าใจภาษามนุษย์ จำเป็นต้องประยุกต์หลักการ และโครงสร้างทางภาษาศาสตร์ไปใช้ในด้านต่าง ๆ หลายระดับเช่นระดับเสียง (Phonology) ระดับหน่วยคำ (Morphology) ระดับกลุ่มคำ (Syntax) ระดับความหมาย (Semantic) จะเป็นเรื่องสำคัญมากที่สุด

การจำแนกกลุ่มคำ

Zellig S. Harris. เป็นนักภาษาศาสตร์ที่ได้เสนอกฎเกณฑ์การแทนที่คำไว้ในปี ค.ศ. 1946 โดยแฮร์ริส เสนอว่า การศึกษาภาษาศาสตร์จำเป็นต้องมีการจำแนกคำและกลุ่มคำเพื่อที่จะได้นำคำที่อยู่ในกลุ่มคำชนิดเดียวกันมาแทนที่กัน ซึ่งชนิดของคำอาจจะแบ่งได้เป็น นาม, นามวลี, กริยา, คุณศัพท์ เป็นต้น

ลักษณะของกฎการแทนที่คำในกลุ่มเดียวกันนี้กล่าวว่า ในประโยคที่เกินจากการเรียงคำเป็น C A D เราสามารถดึงคำ A ออก แล้วแทนที่ด้วยคำ B ถ้าคำ A และ B ต่างเป็นสมาชิกของกลุ่มคำชนิดเดียวกัน ลองพิจารณาตัวอย่างจากประโยค

Where did the . . . go?

เราจะเห็นว่าคำที่อยู่ในช่องว่าง อาจจะเป็นคำนาม Child หรือ young boy ก็ได้เพราะต่างเป็นคำนามกลุ่มเดียวกัน

ดังนั้นในการศึกษาภาษาศาสตร์ เพื่อให้ประมวลผลได้จำเป็นต้องมีการจำแนกคำหรือกลุ่มของคำ ตัวอย่างเช่น

The monkey has eaten the banana.

N T V V T N

จะประกอบกลุ่มคำ Article (T) คำนาม (N) กริยา (V) เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้แฮริสยังเสนอว่าประโยคสามารถแปลงรูปโครงสร้างไปมาได้เช่น การเปลี่ยนรูประหว่าง Active voice กับ Passive voice

The dog bit the man. (Active)

Tha man was bitten by the dog. (Passive)

การเปลี่ยนรูประหว่างประโยคบอกเล่ากับประโยคปฏิเสธ

The man was bitten by the dog.

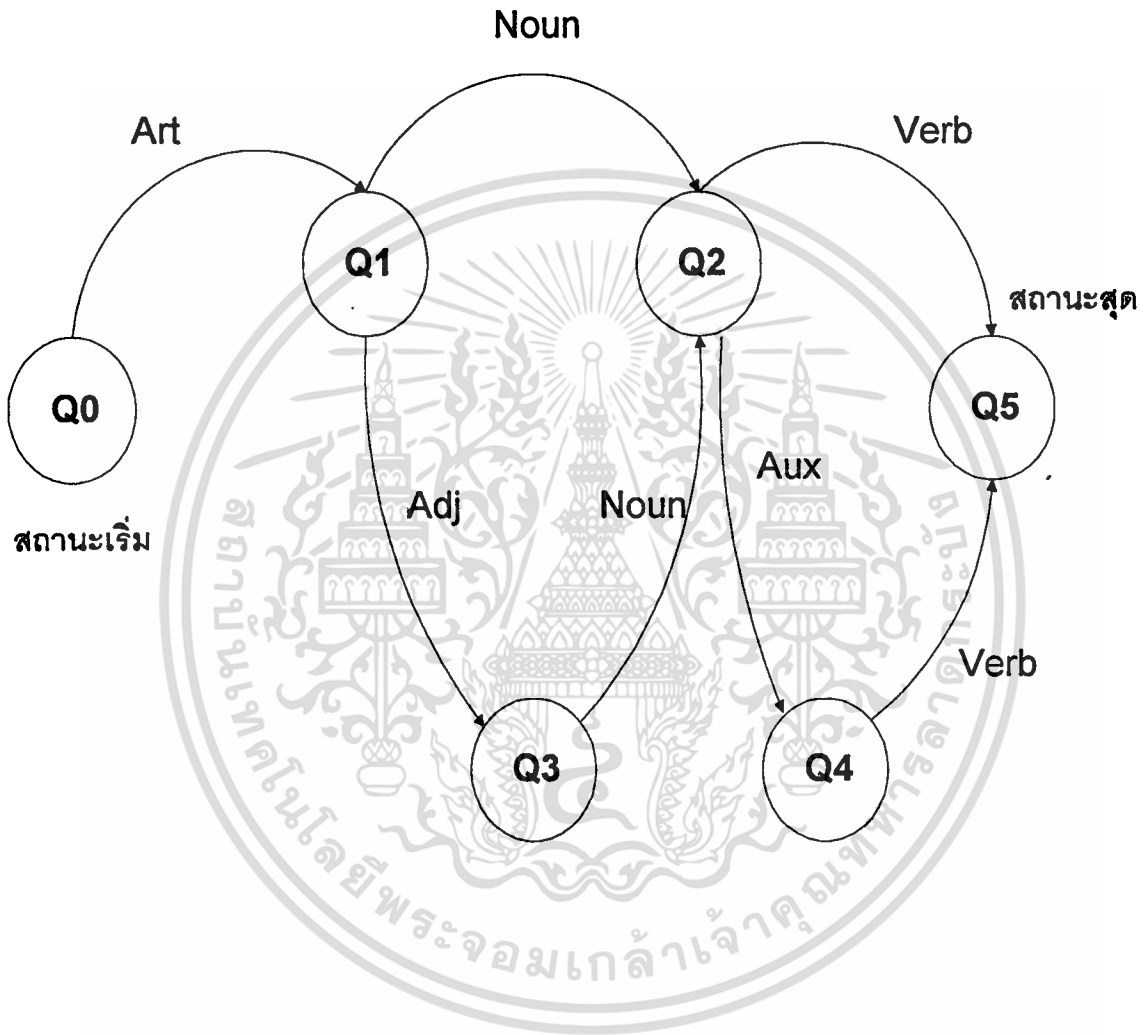
Was the man bitten by the dog?

ไวยากรณ์เพิ่มพูน

ในปี คศ. 1957 Noam Chomsky ได้เขียนหนังสือขึ้นมาเล่มหนึ่งชื่อ “Syntactic Structures” หนังสือเล่มนี้ถือเป็นจุดเริ่มต้นที่ชี้แนวทางการศึกษาศาสตร์ในแนวใหม่ และมีแนวทางสำหรับนำมาประยุกต์ช่วยในการประมวลผล ซอมสกี ได้เสนอวิธีแทนรูปไวยากรณ์ด้วยการสร้างเป็นกฎเกณฑ์ที่แน่นอน เพื่อให้การวิเคราะห์ และการสร้างประโยคเป็นไปอย่างชัดเจนจากโครงสร้างประโยคที่เด่นชัดนี้ ก็ทำให้เราสามารถสร้างและเข้าใจประโยคเพิ่มขึ้นมามาก

ลักษณะของไวยากรณ์ ซอมสกี ได้พัฒนานี้เรียกว่า ไวยากรณ์เพิ่มพูน (generative grammars) ซึ่งเป็นการสร้างประโยคตามโครงสร้างที่ชี้แนะอยู่ในกฎไวยากรณ์ กฎไวยากรณ์ที่กำหนดนี้ไม่สามารถชี้ว่าประโยคนั้นถูกต้องในแง่ของภาษาหรือไม่ แต่จะใช้อธิบายว่าประโยคนั้นมีโครงสร้างและความสัมพันธ์เป็นที่ยอมรับได้หรือไม่

รูปแบบที่ง่ายที่สุดของไวยากรณ์เพิ่มพูนจะเรียกว่าไวยากรณ์สถานะจำกัด (finite state grammar) คือในระบบหรือเครื่องจักรหนึ่งจะประกอบด้วยสถานะที่อาจแทนด้วยวงกลม จำนวนจำกัด การเคลื่อนย้ายจากสถานะหนึ่งไปอีกสถานะหนึ่งตามทิศทางที่แสดงจะให้สัญลักษณ์หรือคำออกมา โดยจะมีสถานะพิเศษเป็นสถานะเริ่มต้น (initial state) และสถานะสุดท้าย (final state) การเข้าสู่ระบบหรือเครื่องจักรจะต้องเริ่มจากสถานะเริ่มต้นและเคลื่อนย้ายไปจนถึงสถานะสุดท้าย



รูปที่ 2-1 เครื่องจักรสถานะจำกัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปเป็นตัวอย่างของระบบหรือไวยากรณ์สถานะจำกัดที่สามารถนำมาใช้แทนประโยคภาษาอังกฤษอย่างย่อ ๆ ได้หลายรูปแบบดังนี้คือ

	Art + Noun + Verb	(The dog runs.)
หรือ		
	Art + Adj + Noun + Verb	(The brown dog runs.)
หรือ		
	Art + Noun + Aux + Verb	(The children can sing.)
หรือ		
	Art + Adj + Noun + Aux + Verb	(The little children can sing.)

อย่างไรก็ตาม ภาษาธรรมชาติเป็นภาษาซับซ้อนไม่แน่นอน ดังนั้นจึงใช้หลักไวยากรณ์สถานะจำกัดไม่ได้ ขอมสก็ได้นิยามไวยากรณ์รูปใหม่ขึ้นมาเรียกว่าไวยากรณ์โครงสร้างวลี (Phase Structure grammars เรียกย่อว่า PS)

โดยองค์ประกอบของประโยคจะแบ่งเป็นส่วน ๆ ตามชนิดของคำหรือในทางภาษาศาสตร์จะเรียก ส่วนประกอบ (Constituents) ได้แก่ คำนาม กริยา เป็นต้น หลักไวยากรณ์แบบโครงสร้างวลีจะเริ่มต้นจากประโยคว่าประกอบด้วยส่วนสำคัญอะไรจากส่วนสำคัญที่แบ่งย่อย และส่วนย่อยก็แบ่งเล็กลงไปอีก

ตัวอย่างของไวยากรณ์ PS

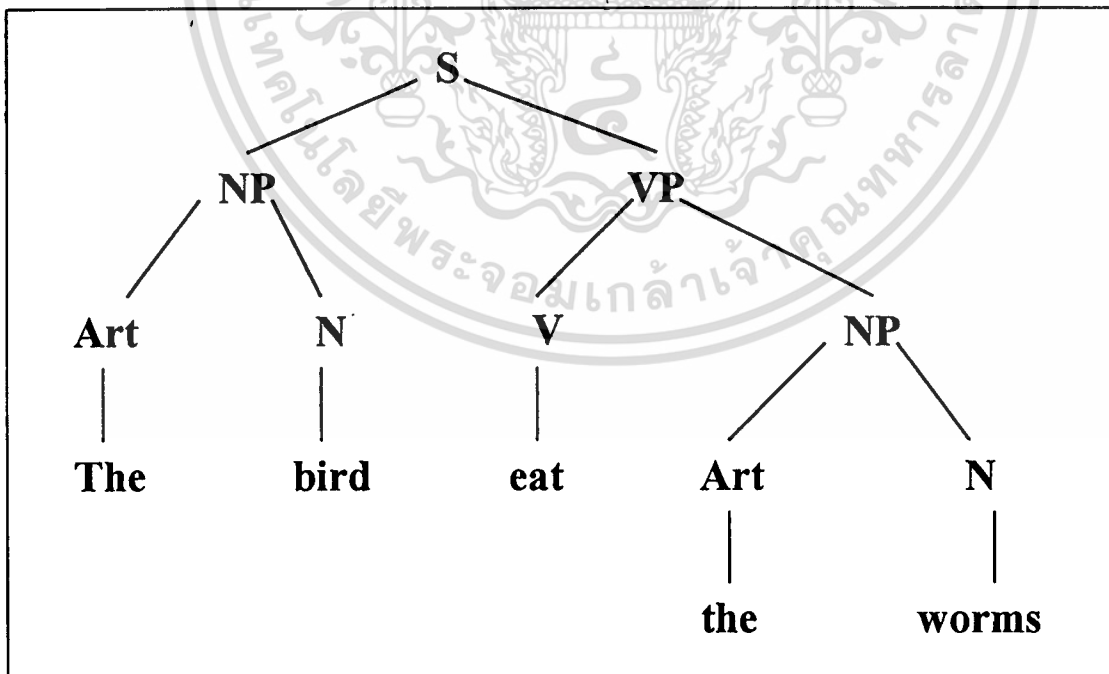
S	→	NP + VP
NP	→	Art + N
VP	→	V + NP
N	→	bird, worms, cars, rocks . . .
Art	→	the
V	→	eat, drive, learn

โดยรูปประโยค (S) จะประกอบด้วยส่วนที่เป็นนามวลี (NP) และกริยาวลี (VP) ส่วนนามวลีประกอบด้วยคำนำหน้า (article) เขียนย่อเป็น Art กับนาม (N) และส่วนกริยาวลีจะประกอบด้วยกริยา และกรรมซึ่งก็คือนามวลี (NP) นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อกำหนดประโยคมาให้เราสามารถนำไวยากรณ์โครงสร้างวลีมาตรวจสอบความถูกต้องทางไวยากรณ์ของประโยคที่เรานิยมเรียกกันว่าการแจงประโยค (Parsing) ซึ่งก็คือการบอกสัมพันธ์ของคำในประโยคนั้นเอง ต่อไปนี้จะเป็นการแสดงวิธีการแจงประโยคจากบนสู่ล่าง (top down parsing) หรือเราอาจแสดงโครงสร้างเป็นแผนภูมิต้นไม้

S → NP + VP
 → Art + N + VP
 → The + N + VP
 → The bird + VP
 → The bird + V + NP
 → The bird eat + NP
 → The bird eat art + N
 → The bird eat art worms



รูปที่ 2-2 โครงสร้างวลีของประโยค The birds eat the worms.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากหลักการไวยากรณ์ที่กล่าวมา เราสามารถสร้างประโยคที่ถูกต้องตามโครงสร้างวลีได้อีก เช่น

The boys drive the cars.

The worms eat the birds.

The birds drive the cars.

ซึ่งจะเห็นว่าบางประโยคมีความหมายที่เป็นไปไม่ได้

Context-Free Grammars (CFG)

ในการศึกษาโครงสร้างประโยคของภาษา นักภาษาศาสตร์จะมีการกำหนดกฎเกณฑ์ที่เรียกหลักไวยากรณ์มาช่วยตรวจวิเคราะห์ และสร้างประโยค ซึ่งก็คือ Context-Free Grammars (CFG) ซึ่งจัดเป็นส่วนหนึ่งของไวยากรณ์โครงสร้างวลี (PS) ซึ่งเป็นหลักไวยากรณ์ที่กล่าวถึงกันมาก เพราะใช้อธิบายโครงสร้างของภาษา และความสัมพันธ์ระหว่างโครงสร้างต่าง ๆ ที่ประกอบในประโยค โดยไม่พิจารณาถึงความหมายของประโยค Context Free Grammar มีชื่อเรียกชื่ออื่นอีก เช่น Backus Normal Form (BNF) หรือ Recursive Pattern

กฎไวยากรณ์ของ CFG จะมีรูปแบบดังนี้

$\langle \text{symbol} \rangle \quad \langle \text{symbol} \rangle^1 \dots \langle \text{symbol} \rangle^n \quad \text{For } n > 1$

2.2 การแจกแจงประโยค

การแจกแจงประโยคสำหรับไวยากรณ์ CFG อาจทำได้ 2 ทาง คือ การแจกแจงประโยคจากบนลงล่าง (Top-Down parsing) และการแจกแจงจากล่างขึ้นบน (Bottom-up parsing) การแจกแจงประโยคจากบนลงล่าง จะเริ่มจากสัญลักษณ์ประโยค S จะเขียนแทนด้วย สัญลักษณ์นามวลี (NP) ตามด้วยกริยาวลี (VP) สัญลักษณ์ NP และ VP จะเขียนแทนต่อไป จนถึงสัญลักษณ์จบท้ายที่จะใช้แทนคำศัพท์ต่างๆ ลักษณะของการแจกแจงประโยคบนลงล่างจะเป็นการใช้กฎไวยากรณ์ของสัญลักษณ์ทางด้านขวาไปเขียนใหม่แทนสัญลักษณ์ทางด้านซ้าย ดังตัวอย่างด้านล่าง

ตัวอย่าง การแจกแจงประโยคจากบนลงล่าง (Top down parsing)

S → NP + VP
 → Art + N + VP
 → The + N + VP
 → The bird + VP
 → The bird + V + NP
 → The bird eat + NP
 → The bird eat art + N
 → The bird eat art worms.

ตัวอย่าง การแจกแจงประโยคจากล่างขึ้นบน (Bottom-up parsing)

→ The birds eat the worms.
 → Art + birds eat the worms.
 → Art + N + eat the worms.
 → Art + N + V + the worms.
 → Art + N + V + Art + worms.
 → Art + N + V + Art + N
 → Art + N + V + Art + N
 → NP + V + Art + N
 → NP + V + NP
 → NP + VP
 → S

กฎไวยากรณ์ที่กล่าวมาแล้ว ยังไม่ได้รวมถึงส่วนที่เวียนเกิดซ้ำ (Recursive) ซึ่งส่วนที่เวียนเกิดซ้ำหมายถึงสัญลักษณ์ที่เวียนกลับมากำหนดตัวเองได้อีก กฎไวยากรณ์ที่สมบูรณ์จะต้องรวมถึงการเวียนเกิดซ้ำด้วย ตัวอย่างกฎไวยากรณ์ที่รวมการเวียนเกิดซ้ำของภาษาอังกฤษที่สมบูรณ์ยิ่งขึ้นจะแสดงตามตัวอย่างข้างล่าง โดยขอให้สังเกตว่าสัญลักษณ์นามวลี (NP) อาจจะประกอบด้วยบุพบทวลี (PP) ซึ่งจะย้อนกลับมาสู่นามวลี (NP) อีกครั้งหนึ่งเป็นลักษณะของการเวียนเกิดซ้ำอีก หลักไวยากรณ์ที่จะยกตัวอย่างด้านล่าง เป็นเพียงกฎง่าย ๆ ที่ไม่ได้ครอบคลุมภาษาอังกฤษทั้งหมด ยังไม่มีคำสันธานมาเชื่อมประโยค และประโยคเชิงซ้อน ไม่มีกฎว่าด้วยการกระจายของประธาน และกริยา

ตัวอย่าง กฎไวยากรณ์ และคำศัพท์ คำย่อ

กฎไวยากรณ์		
S	→	NP + VP
NP	→	Mod + N + (PP)
Mod	→	(Art) + (Adj)
VP	→	V + (ADV)
VP	→	Aux + V
V	→	Vi
V	→	Vt + NP
V	→	Vc + NP
ADV	→	PP
ADV	→	Adv
PP	→	Prp + NP

คำศัพท์

N	→	bird, tree, boy
Art	→	a, an, the
Adj	→	Dark, good, tall
Adv	→	fast, slowly, ...
Vi	→	cry, swim
Vt	→	hit, break, eat
Vc	→	is, become, ...
Aux	→	has, must, can, ...

คำอธิบายสัญลักษณ์ย่อ

S	Sentence	(ประโยค)
NP	noun phase	(นามวลี)
VP	Verb phase	(กริยาวลี)
M,d	modifier	(ส่วนขยาย)
Art	article	(คำนำหน้านาม)
Adj	adjective	(คุณศัพท์)
ADV	adverbial	(วิเศษณ์วลี)
Adv	adverb	(วิเศษณ์)
V	Verb	(กริยา)
Aux	auxiliary	(กริยาช่วย)
Vi	intransitive verb	(อกรรมกริยา)
Vt	transitive	(สกรรมกริยา)
Vc	copulative verb	(กริยาต่อคำ)
PP	preposition phase	(บุพบทวลี)
Prep	Preposition	(บุพบท)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 ข่ายงาน และการเวียนเกิดซ้ำ

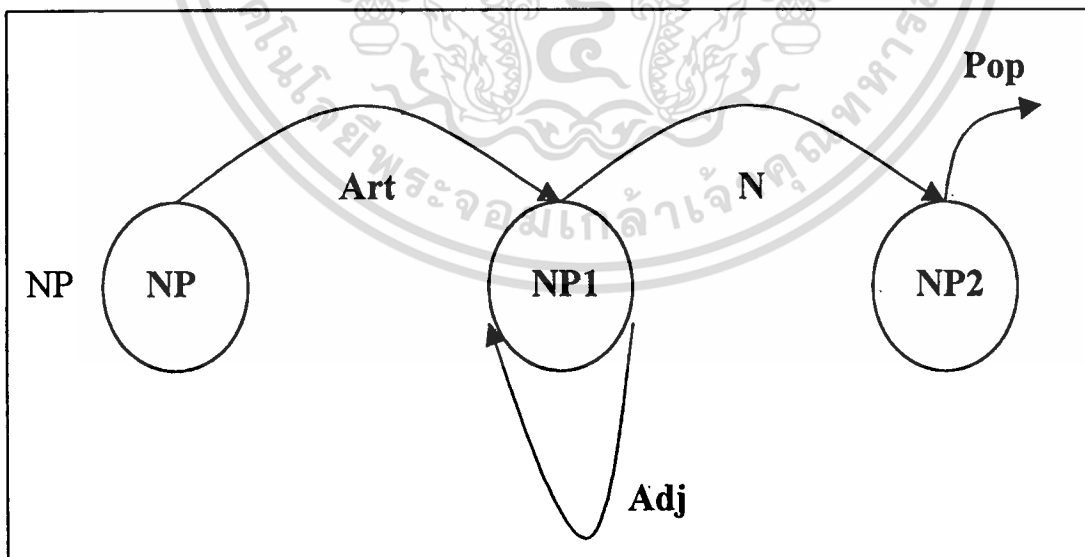
กฎไวยากรณ์ CFG ที่กล่าวมาแล้วเราสามารถนำมาเขียนในรูปกราฟในลักษณะของสถานะต่าง ๆ ดังรูปที่ 2-3 เรียกว่า ข่ายงานการเปลี่ยน transition networks ที่เป็นเครือข่ายประกอบด้วยโหนด (node) แทนด้วยวงกลม และส่วนโค้ง (arc) ที่มีอักษรสัญลักษณ์กำกับพร้อมลูกศรบอกทิศทางดังรูป ข่ายงานการเปลี่ยนทางเดินของนามวลีที่มาจากกฎไวยากรณ์ CFG

NP → Art + NP1

NP1 → Adj + NP1

NP1 → N + NP2

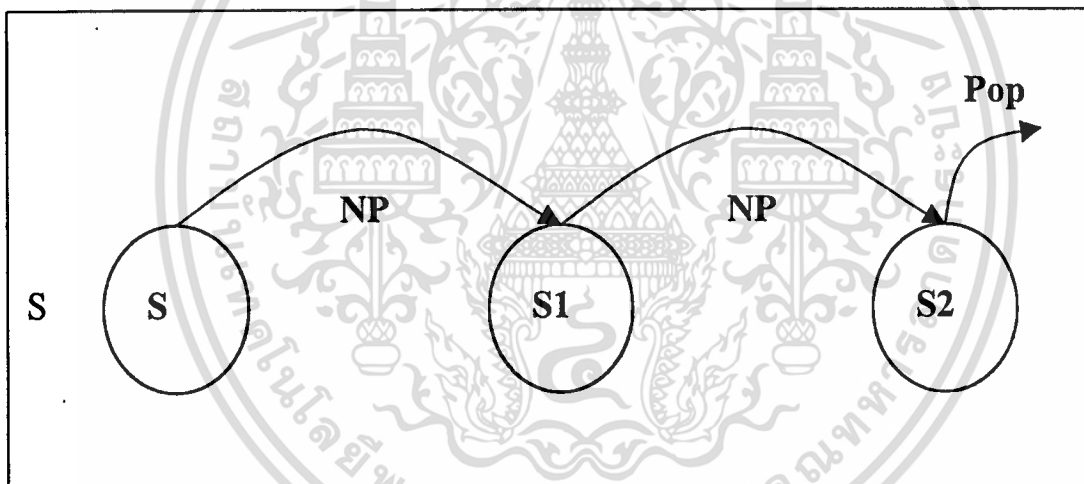
โดยแต่ละโหนดจะมีสัญลักษณ์กำกับภายใน ขณะเดียวกันเส้นโค้งทุกเส้นก็จะมีชนิดของคำกำกับไว้เมื่อเริ่มต้นจากโหนด เราสามารถเดินทางตามเส้นโค้งไปสู่โหนดอื่นตามทิศทางที่กำหนดไว้ หากคำที่กำลังพิจารณาในขณะกระจายประโยชน์ตรงกับชนิดของคำที่กำกับเส้นโค้งนั้น ถ้า NP หรือเครือข่ายนั้นถูกต้อง เราก็สามารถออกจากเส้นโค้ง Pop ที่เป็นสัญญาณบอกว่าประสบความสำเร็จในการเดินทางตามเครือข่าย



รูปที่ 2-3 แสดงข่ายงานการเปลี่ยนแปลงของนามวลี NP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบของข่ายงานข้างต้นเป็นแบบง่าย ๆ ซึ่งไม่เหมาะสำหรับการแทนภาษาธรรมชาติ รูปแบบข่ายงานที่สมบูรณ์มากขึ้นจะต้องประกอบด้วยส่วนการเวียนเกิดซ้ำ ดังนั้นจึงเกิดข่ายงานการเปลี่ยนแบบเวียนเกิดซ้ำ (Recursive Transition Networks เรียกย่อ ๆ ว่า RTN) ที่มีการกระโดดข้ามระหว่างข่ายงานหนึ่งไปหาอีกข่ายงานหนึ่ง โดยแต่ละข่ายงานมีชื่อกำกับไว้ดังรูปที่ 2-4 เป็นข่ายงาน S หรือประโยค S เมื่อมีการเดินทางจากโหนด S ไป S1 ตามเส้นโค้ง NP ตามรูปที่ 2-3 และถ้าเดินทางภายในข่ายงาน NP ได้สำเร็จจนออกมาที่เส้นโค้ง pop ได้ก็จะเป็นการกลับมาสู่โหนด S1 ของข่ายงานเดิมที่มีการเรียกไปยังข่ายงาน NP เพื่อเดินทางไปยังโหนด S2 ผ่านเส้นโค้ง Verb ต่อไปเรื่อย ๆ จนจบข่ายงาน S ทางเส้นโค้ง pop



รูปที่ 2-4 แสดงข่ายงานไวยากรณ์ RTN

บทที่ 3

ระบบ Natural Language Query for Tourism

1.1 โครงสร้างของระบบ

การทำงานของระบบแบ่งออกเป็น 5 ส่วน คือ

1. Morphological (Lexicon Analysis)

มีหน้าที่วิเคราะห์ประโยคที่ผู้ใช้ Query เข้ามาว่าเป็นคำชนิดใดในภาษาอังกฤษ แล้ว word ที่ได้ไปยังส่วนของ syntactic Analysis มีขั้นตอนการทำงานดังนี้

- นำประโยคที่ผู้ใช้ Input เข้า sprite เป็นคำ
- ทำการเปลี่ยนรูปของคำที่ได้มาให้อยู่ในรูปที่แท้จริง
- ทำการค้นหา และเปรียบเทียบกับคำใน English lexicon
- กำหนดชนิดของคำ ในกรณีทีคำนั้นใน Lexicon โดยแทนชนิดของคำด้วยตัวเลข 0 – 9 ดังนี้

Noun	=	0
Verb	=	1
Adjective	=	2
Adverb	=	3
Auxiliary	=	4
Article	=	5
Question	=	6
Pronoun	=	7
Preposition	=	8

2. Syntactic Analysis

ทำหน้าที่แจกแจงประโยค ว่าถูกต้องตามหลักไวยากรณ์ของประโยคหรือไม่โดยใช้ไวยากรณ์ (Grammar rule) ที่มีอยู่ในระบบ จากนั้นนำมาเรียงตามลำดับของคำในประโยค แล้วแทนคำ

ตัวอย่าง How about climate of Thailand? จะได้ค่า → 33080
และค่าของประโยคที่เป็นไปได้ในรูปแบบอื่นดังนี้

→ 38080

→ 63080

→ 68080

แล้วทำการตรวจสอบประโยคที่ถูกต้องตามหลักไวยากรณ์ โดยใช้การแจกแจงประโยค (Parsing Grammar) ของ CFG แล้วใช้ Algorithm แบบบนลงล่าง (Top Down Parsing)

3. Keyword Extract

ทำหน้าที่หา Keyword จากประโยค Query ที่ผ่าน Syntactic Analysis และแปลงค่าเป็นตัวเลขตามชนิดของคำแล้วนำมาเปรียบเทียบกับ Pattern หรือ Feature ที่ระบบมีอยู่ แล้วทำการตีความว่าประโยคที่มี Feature ลักษณะนี้ควรมีค่าในตำแหน่งใดเป็น keyword

ตัวอย่าง ขั้นตอนการทำงานของ Keyword Extract Process

ประโยค How about climate of Thailand

Step 1 → climate of Thailand

Step 2 → ได้ Query statement (Thailand and climate)

ประโยค What is the famous temple in Bangkok?

Step 1 → Bangkok famous temple

Step 2 → ได้ Query statement (Bangkok and famous and temple)

ประโยค What is the highest mountain in Thailand

Step 1 → highest mountain in Thailand

Step 2 → ได้ Query statement (highest mountain and Thailand)

ประโยค Show me the map of Thailand

Step 1 → map of Thailand

Step 2 → ได้ Query statement (map and Thailand)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. Output Format

ทำหน้าที่รับ Query Statement ที่ได้จาก Keyword Extract Process นำมา Query ใน Internal Document และ WWW Document เพื่อหา Document ที่ match กับ Query Statement ของผู้ใช้

5. Document Operate

ในส่วนนี้จะเกี่ยวกับการจัดการเกี่ยวกับข้อมูลการท่องเที่ยวที่เป็น Internal Document ในส่วนนี้จะมี Content Editor เป็นผู้คอยจัดการกับข้อมูลหรือเอกสารที่เตรียมไว้ให้ผู้ใช้ค้นหา

3.2 แนวทางในการพัฒนาระบบ

โครงการ Natural Language Query for Tourism มีลักษณะขั้นตอนในการออกแบบ และเทคนิคการทำงานซึ่งแบ่งออกเป็น ส่วน ๆ ดังต่อไปนี้

3.2.1 Database

Microsoft Access 2000 เป็นฐานข้อมูลในการจัดเก็บคำศัพท์ และฐานข้อมูลการท่องเที่ยวสำหรับนักท่องเที่ยว

3.2.2 Software Tools อื่น ๆ

ตัวแปลภาษา Active Perl 5 for Win32

Web Server Software OmniHTTPd Professional 2.07

Adobe PhotoShop 5.5 ใช้ในการตกแต่งรูปภาพ และสร้าง Icon

Microsoft Windows 98 หรือ Windows NT เป็น Operating System

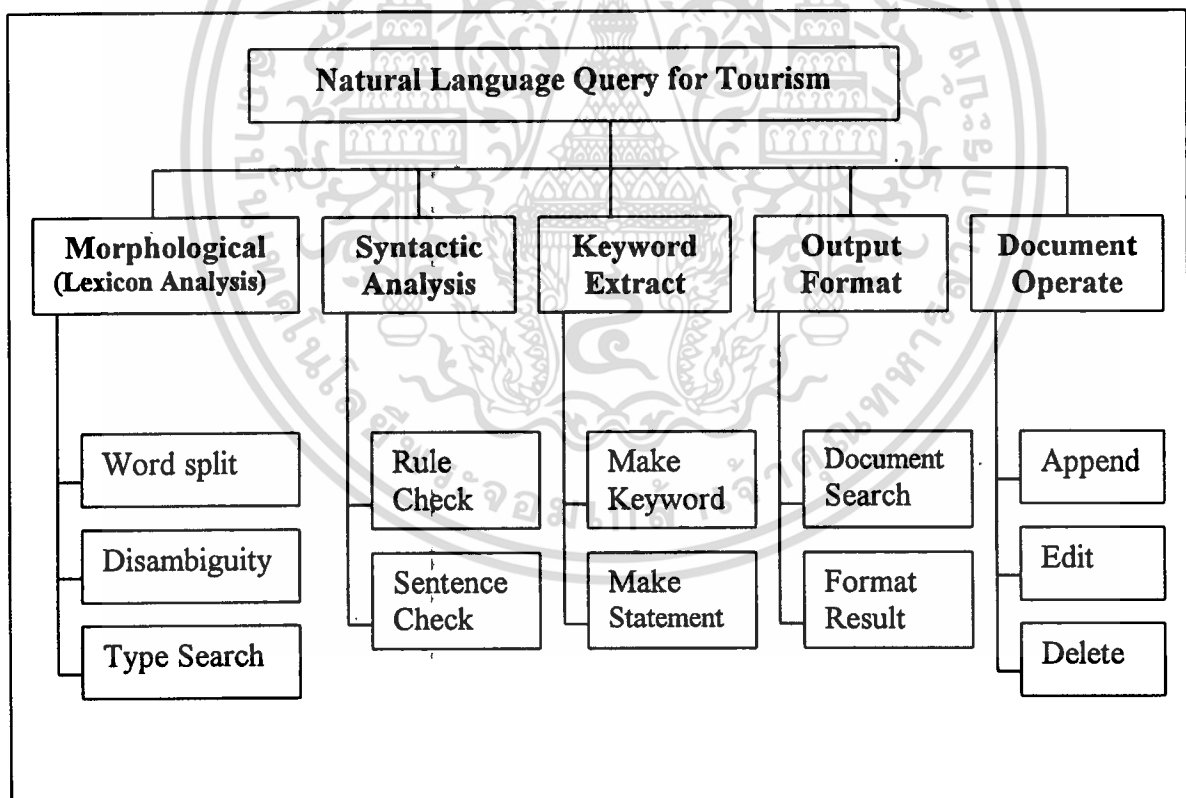
3.3 การวิเคราะห์ และออกแบบระบบ

ขั้นตอนวิเคราะห์ และออกแบบโครงการวิจัย Natural Language Query for Tourism ได้นำเสนอหัวข้อต่อไปนี้

- Structure Chart
- Data Flow Diagram
- Data Dictionary
- File Structure

3.3.1 Structure Chart

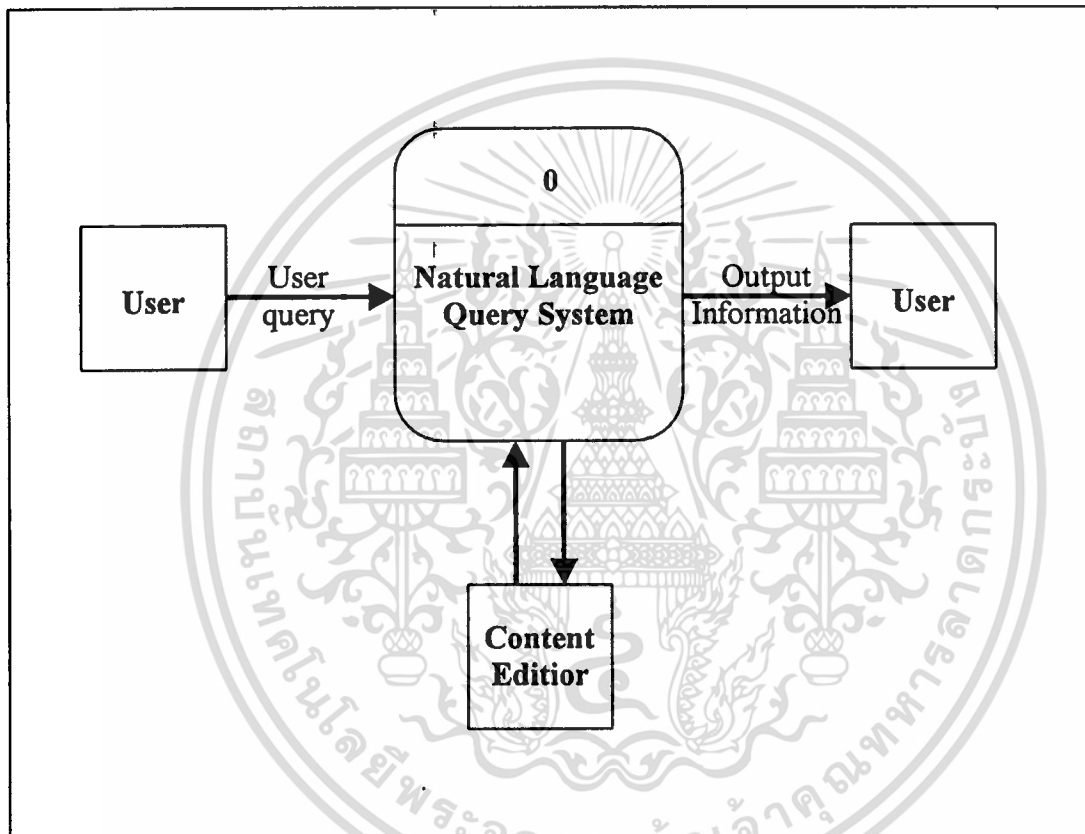
เป็น Chart ที่แสดงโครงสร้างของระบบ Natural Language Query โดยแบ่งตาม Module ดังที่แสดงตามรูป 3-1



รูปที่ 3-1 Structure Chart Natural Language Query for Tourism System

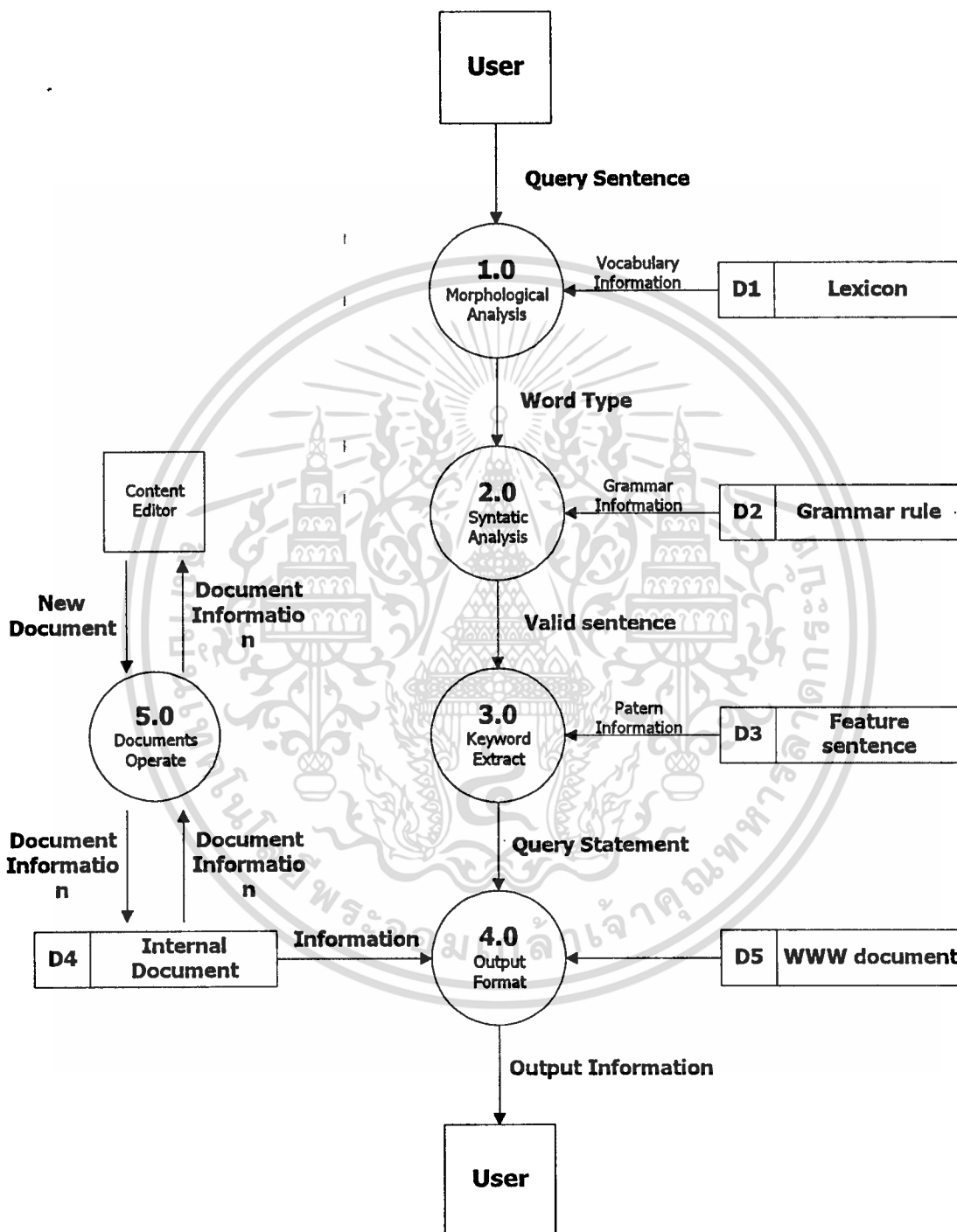
3.3.2 Data Flow Diagram

เป็น Diagram ซึ่งแสดงการ Flow ของข้อมูลในระบบ โดยจะแบ่งเป็น Level จำนวน 2 Level ดังรูปที่ 3-2 ถึงรูปที่ 3-8



รูปที่ 3-2 Context Diagram Level 0 of Natural Language Query for Tourism

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

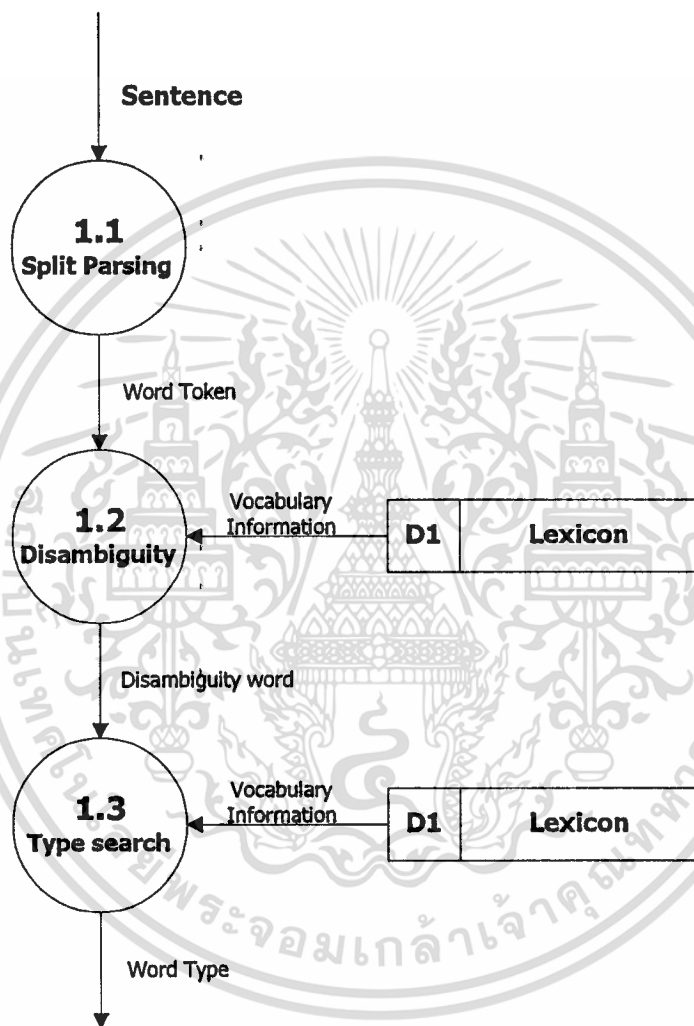


รูปที่ 3-3 Level 1 DFD Natural Language Query for Tourism

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์ของบริษัทเอกชนเพื่อใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

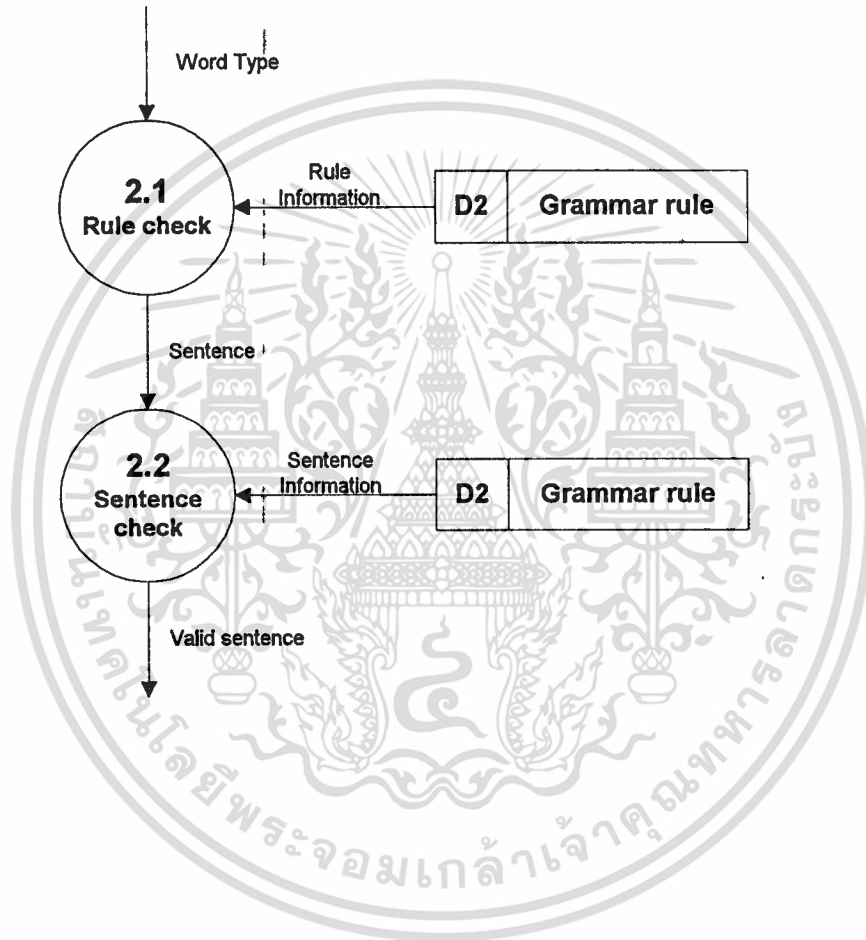
Morphological Analysis



รูปที่ 3-4 DFD Natural Language Query for Tourism Level 2 Process 1.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

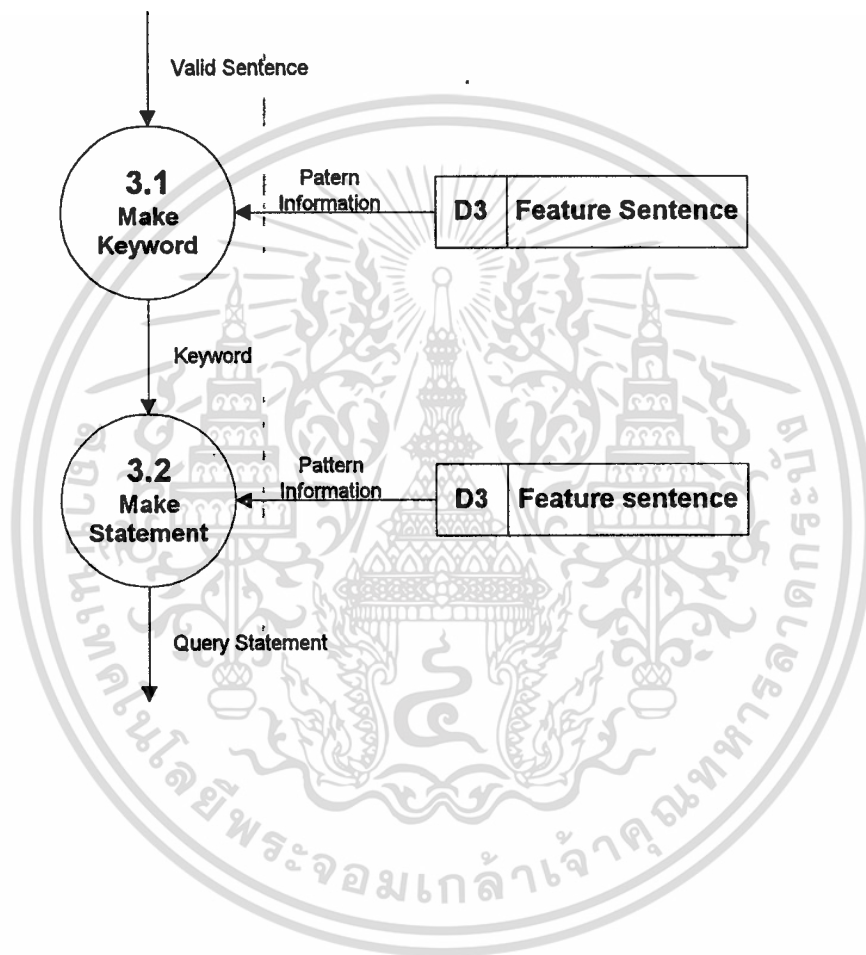
Syntactic Analysis Process



รูปที่ 3-5 DFD Natural Language Query for Tourism Level 2 Process 2.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

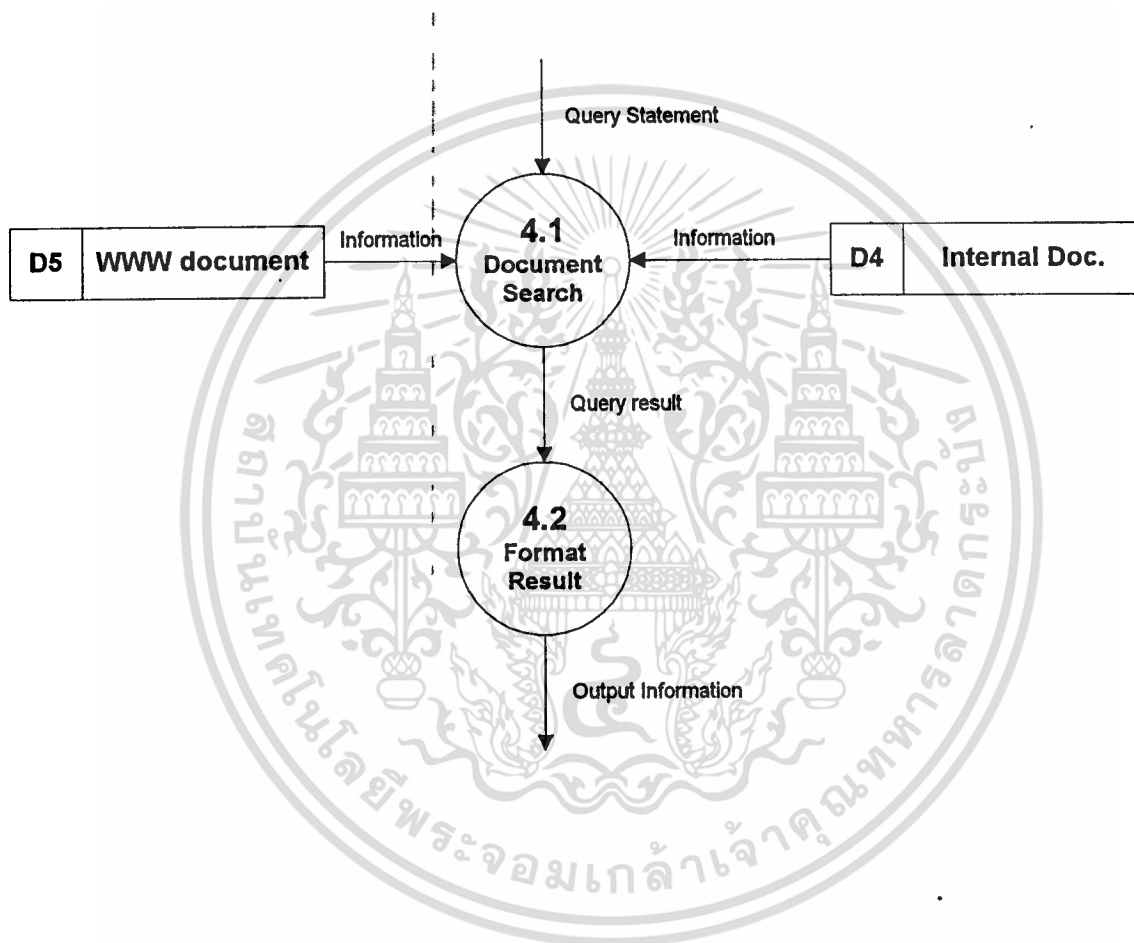
Keyword Extract



รูปที่ 3-6 DFD Natural Language Query for Tourism Level 2 Process 3.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

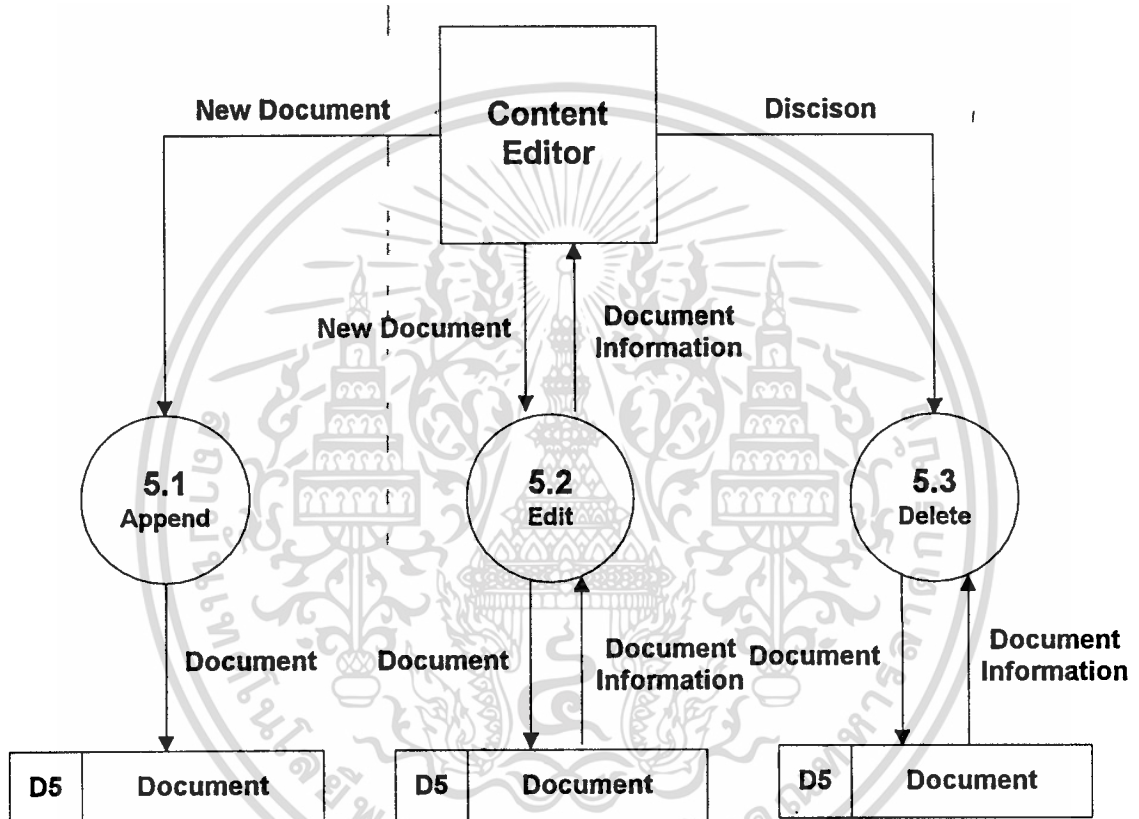
Process Output Format



รูปที่ 3-7 DFD Natural Language Query for Tourism Level 2 Process 4.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Document Operate



รูปที่ 3-8 DFD Natural Language Query for Tourism Level 2 Process 5.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3 Data Dictionary

ในส่วนนี้จะอธิบายรายละเอียดของ Data Flow Diagram ทุก Level โดยแยกอธิบายเป็น 3 ส่วน ดังนี้

- Data Dictionary of Data Flow Diagram ในส่วนนี้จะอธิบายรายละเอียดของ Data Flow หรือทิศทางกรไหลของข้อมูลในส่วนต่าง ๆ ของ Data Flow Diagram
- Data Dictionary of Process ในส่วนนี้จะอธิบายรายละเอียดของ Process ต่าง ๆ ของ Data Flow Diagram
- Data Dictionary of Data Store ในส่วนนี้จะอธิบายรายละเอียดของ Data Store แต่ละตัวว่ามีข้อมูลอะไร

Data Dictionary of Data Flow	
Level 0 of Natural Language Query for Tourism	
Data Flow Name	: User Query
Description	: ประโยคคำถามภาษาอังกฤษที่ผู้ใช้ป้อนเข้ามาในระบบ NLQ ซึ่งประโยคส่วนใหญ่จะเป็นประโยคคำถามที่เป็นลักษณะของ WH Question และประโยคคำสั่ง ซึ่งจะสอบถามข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทย เช่น Where is the popular place in Thailand, Show me a map of Thailand
From Source	: User
To Source	: Natural Language Query for Tourism (NLQ)
Data Flow Name	: Query Result
Description	: ผลลัพธ์ที่ได้จากการ Query ของผู้ใช้ซึ่ง Output จะอยู่ในรูปของ HTML link และคำอธิบายสั้น ๆ เกี่ยวกับเนื้อหาของ Home Page ที่ Match กับประโยคที่ผู้ใช้ Query เข้ามา
From Source	: Natural Language Query for Tourism (NLQ)
To Source	: User

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Flow Name : Tourism Information
Description : ข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทยที่จัดเก็บลงในระบบ โดย Content Editor
From Source : Content Editor
To Source : Natural Language Query for Tourism

Data Flow Name : Document Information
Description : ข้อมูลเกี่ยวกับการท่องเที่ยวที่มีอยู่ในระบบ ซึ่ง Content Editor จะอ่าน และทำการแก้ไขตามความเหมาะสม
From Source : Natural Language Query for Tourism (NLQ)
To Source : Content Editor

Level 1 of Natural Language for Tourism

Data Flow Name : Query Sentence
Description : ประโยคคำถามภาษาอังกฤษที่ผู้ใช้ป้อนเข้ามาในระบบ NLQ ซึ่งประโยคส่วนใหญ่จะเป็นประโยคคำถามที่เป็นลักษณะของ WH Question และประโยคคำสั่ง ซึ่งจะสอบถามข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทย เช่น Where is the popular place in Thailand, Show me a map of Thailand
From Source : User
To Source : 1.0 Morphological

Data Flow Name : Vocabulary Information
Description : รายละเอียดของรายการศัพท์ที่เก็บอยู่ใน Lexicon Data Store
From Source : D1: Lexicon
To Source : 1.0 Morphological

Data Flow Name	: Word token
Description	: คำที่ได้จากการ Split จากประโยคที่ผู้ใช้ Query เข้ามาพร้อมทั้งกำหนดชนิดของคำ (Part of Speech)
From Source	: 1.0 Morphological
To Source	: Syntactic Analysis
.	
Data Flow Name	: Grammar Information
Description	: กฎเกณฑ์ทางภาษาที่ใช้ในการแจกแจงประโยค
From Source	: D2: Grammar rule
To Source	: 2.0 Syntactic Analysis
.	
Data Flow Name	: Valid Sentence
Description	: ประโยคที่รับการแจกแจง (Parsing) จากระบบเรียบร้อยแล้ว
From Source	: Syntactic Analysis
To Source	: Keyword Extract
.	
Data Flow Name	: Keyword
Description	: Keyword ที่ได้จาก Query Sentence ของผู้ใช้
From Source	: Keyword Extract
To Source	: Output Format
.	
Data Flow Name	: Information
Description	: ข้อมูลเกี่ยวกับการท่องเที่ยวที่อยู่ในระบบ ลักษณะของข้อมูลจะเป็นข้อมูลที่จำเป็นสำหรับนักท่องเที่ยว เช่น ที่พัก สถานที่ท่องเที่ยวที่สำคัญในประเทศไทย ตารางเวลาการเดินทาง
From Source	: D4: Tour Information
To Source	: 4.0 Output Format

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Flow Name : Output Result

Description : ผลลัพธ์ที่ได้จากการ Query ของผู้ใช้

From Source : D4: Output Format

To Source : User

Data Flow Name : Document Input

Description : ข้อมูลการท่องเที่ยวที่ Content Editor พิจารณาเห็นว่าควรเพิ่มเข้าไปในระบบ

From Source : Content Editor

To Source : Document Operate

Data Flow Name : Document Out

Description : ข้อมูลการท่องเที่ยวเกี่ยวกับประเทศไทยที่ Content Editor อ่านขึ้นมาเพื่อแก้ไข หรือปรับปรุง

From Source : Tour Information, 5.0 Document Operate

To Source : Content Editor

Level 2 of Morphological Process

Data Flow Name : Sentence

Description : ประโยคที่ผู้ใช้ Query เข้ามาในระบบ

From Source : User

To Source : Split parsing

Data Flow Name : Word Token

Description : กลุ่มของคำที่ได้จากประโยคที่ผู้ใช้ Query เข้ามาในระบบ

From Source : 1.1 Sprit parsing

To Source : 1.2 Disambiquity

Data Flow Name : Vocabulary Information
Description : รายละเอียดของคำศัพท์ที่อยู่ในรูปต่างๆ
From Source : D1: Lexicon
To Source : 1.2 Disambiguity

Data Flow Name : Disambiguity word
Description : คำที่แปลงอยู่ในรูปที่แท้จริง สามารถนำไปค้นหาชนิดของคำเพื่อ
กำหนด Part of Speech ของคำ
From Source : 1.2 Disambiguity
To Source : 1.3 Type Search

Data Flow Name : Word Type
Description : คำที่ระบุชนิดของคำ (Part of Speech) เรียบร้อยแล้ว
From Source : 1.3 Type Search
To Source : 2.0 Syntactic Analysis

Level 2 of Syntactic Analysis Process

Data Flow Name : Word Token
Description : คำศัพท์ที่ระบุชนิดของคำเรียบร้อยแล้ว
From Source : 1.3 Type Search
To Source : Rule Check

Data Flow Name : Sentence
Description : ประโยคที่ได้แจกแจงเรียบร้อยแล้วตาม Grammar rule
From Source : 2.1 Rule Check
To Source : 2.2 Sentence Check

Data Flow Name : Pattern Information
Description : รูปแบบของประโยคที่ระบบสามารถรับได้
From Source : D3: Feature Sentence
To Source : 3.2 Sentence Check

Data Flow Name : Valid Sentence
Description : ประโยคที่ทำการตรวจสอบความถูกต้องเรียบร้อยแล้วพร้อมที่จะนำไปหา Keyword
From Source : 2.2 Sentence Check
To Source : 3.1 Make Keyword

Level 2 of Keyword Extract Process

Data Flow Name : Valid Sentence
Description : ประโยคที่ได้รับการตรวจสอบความถูกต้องตามหลักไวยากรณ์แล้ว
From Source : 2.2 Sentence Check
To Source : 3.1 Make Keyword

Data Flow Name : Pattern Information
Description : ข้อมูลรูปแบบของประโยคที่ใช้สำหรับหา Keyword ของประโยค
From Source : D3: Feature Sentence
To Source : 3.1 Make Keyword

Data Flow Name : Keyword
Description : คำสำคัญ (Keyword) ของประโยคที่ผู้ใช้ Query
From Source : 3.1 Make Keyword
To Source : 3.2 Make Statement

Data Flow Name : Vocabulary Information
Description : ข้อมูล หรือรายละเอียดของคำศัพท์ เช่น ตำแหน่งการวางในประโยคชนิดของคำศัพท์ (Part of Speech)
From Source : D1: Lexicon
To Source : 3.2 Make Statement

Data Flow Name : Query Statement
Description : ประโยคที่อยู่ในรูปของ Boolean Logic ที่ระบบสร้างขึ้น สามารถใช้ค้นหาใน Internal Database หรือส่งต่อไปให้ยัง Search Engine ในกรณีที่ไม่มีพบ Document ที่ Match กับ Query Expression
From Source : 3.2 Make Keyword
To Source : 4.1 Document Search

Level 2 of Output Format Process

Data Flow Name : Internal Information
Description : ข้อมูลการท่องเที่ยว และข้อมูลที่เกี่ยวข้องกับประเทศไทย
From Source : D4: Internal Information
To Source : 4.1 Document Search

Data Flow Name : WWW Information
Description : ข้อมูลที่ได้จาก Search Engine เพื่อนำมาเช็คว่า Match กับ Query Expression ของผู้ใช้หรือไม่
From Source : D5: WWW Information
To Source : 4.1 Document Search

Data Flow Name : Result
Description : ผลลัพธ์ที่ได้จากการ Query ของผู้ใช้
From Source : 4.1 Document Search

To Source : 4.2 Format Result

Level 2 of Document Operate Process

Data Flow Name	: Decision
Description	: ข้อมูลการพิจารณา หรือตัดสินใจว่าควรลบ Document ออกจากระบบหรือไม่ ซึ่งบางครั้ง Document บางตัวถ้าสมัสมควรที่จะลบออกจากระบบ
From Source	: Content Editor
To Source	: 5.3 Delete
Data Flow Name	: Document Information
Description	: ข้อมูลของ Document ที่เก็บอยู่ใน Database
From Source	: D4: Internal Document
To Source	: 5.1, 5.2, 5.3
Data Flow Name	: Document
Description	: Document ที่ Content Editor ทำการลบ แก้ไข เพิ่ม เข้าไปยัง Internal Document
From Source	: 5.1, 5.2, 5.3
To Source	: D4: Internal Document
Data Flow Name	: New Document
Description	: Document ใหม่ที่ content Editor พิจารณาแล้วควรเพิ่มเข้าไปใน Internal Document สำหรับนักท่งเที่ยวหรือผู้สนใจ
From Source	: Content Editor
To Source	: 5.1, 5.2, 5.3

Data Dictionary of Process

Level 0 of Natural Language Query for Tourism

Process name : Natural Language Query for Tourism
Description : ระบบ Natural Language Query for Tourism
Input Data Flow : User Query
Output Data Flow : Query Result

Level 1 of Natural Language Query for Tourism

Process name : Morphological Analysis
Description : Process วิเคราะห์ชนิดของคำ
Input Data Flow : Query Sentence, Vocabulary Information
Output Data Flow : Word type
Data Flow Name : Syntactic Analysis
Description : Process วิเคราะห์ และแจกแจงประโยคตามหลักไวยากรณ์ (Grammar rule)
From Source : Word type, Grammar rule
To Source : Valid sentence
Data Flow Name : Keyword Extract
Description : Process หา คำสำคัญ (Keyword) ในประโยคของผู้ใช้
From Source : Valid sentence, Pattern information
To Source : Keyword

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Flow Name : Output Format
Description : Process ที่ค้นหา Document ที่ Match กับ Query ของผู้ใช้
From Source : Query Statement
To Source : Query Result

Data Flow Name : Document Operate
Description : Process ที่จัดการเกี่ยวกับ Documents ใน Internal Documents
From Source : New Documents, Document information
To Source : Document information

Level 2 of Natural Language Query for Tourism Process 1.0

Process name : Split Parsing
Description : Process ในการแยกประโยคให้เป็นคำ
Input Data Flow : Sentence
Output Data Flow : Word Token
Data Flow Name : Disambiguity
Description : Process จัดการกับคำที่อยู่ในรูปที่แท้จริงของคำ เพื่อสะดวกต่อการหาชนิดของคำ
From Source : Word Token
To Source : Disambiguity word

Data Flow Name : Type Search
Description : Process ที่ค้นหาชนิดของคำ (Part of Speech)
From Source : Disambiguity word, Vocabulary information
To Source : Word Type

Level 2 of Natural Language Query for Tourism Process 2.0

Process name	: Rule Check
Description	: Process ตรวจสอบตำแหน่งของคำศัพท์ที่วางในประโยค ชนิดของคำ ตามกฎไวยากรณ์
Input Data Flow	: Word Token, Rule Information
Output Data Flow	: Sentence
Data Flow Name	: Sentence Check
Description	: ตรวจสอบประโยคว่ามีรูปแบบที่ใช้ได้หรือมีความหมายหรือเปล่า
From Source	: Sentence
To Source	: Valid Sentence

Level 2 of Natural Language Query for Tourism Process 3.0

Process name	: Make Keyword
Description	: Process ในการหาคำสำคัญ (Keyword) ของประโยค
Input Data Flow	: Valid Sentence, Pattern Information
Output Data Flow	: Keyword
Data Flow Name	: Make Statement
Description	: Process ในการสร้าง Boolean Statement จาก keyword
From Source	: Keyword, Vocabulary Information, Pattern information
To Source	: Query Statement

Level 2 of Natural Language Query for Tourism Process 4.0

Process name : Document Search
Description : Process ค้นหา document ที่ Match กับ Query ของผู้ใช้
Input Data Flow : Query Statement, Internal Information, External Information
Output Data Flow : Query Result

Data Flow Name : Format Output
Description : Process จัดรูปแบบของผลลัพธ์ที่ได้ให้อยู่ในรูปแบบที่ผู้ใช้สามารถเข้าใจได้ง่าย และเป็นระเบียบ
From Source : Query Result
To Source : Output Information

Level 2 of Natural Language Query for Tourism Process 5.0

Process name : Delete
Description : Process ที่จัดการลบ Document ที่ล้าสมัยออกจากระบบ
Input Data Flow : Discision, Document Information
Output Data Flow : Document

Data Flow Name : Edit
Description : Process ที่จัดการแก้ไข Document ในระบบ
From Source : New Document Information, Document Information
To Source : Document Information

Data Flow Name : Append
Description : Process ที่จัดการเพิ่ม Document เข้าในระบบ
From Source : New Document Information
To Source : Document

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Dictionary of Data Store

- Data Store Name** : Lexicon
- Description** : เก็บข้อมูลเกี่ยวกับคำศัพท์ในภาษาอังกฤษ เช่น เป็นคำชนิดใด
- Data Store Name** : Grammar Rule
- Description** : เก็บกฎเกณฑ์ทางภาษาที่ใช้ในการวิเคราะห์ประโยค
- Data Store Name** : Feature Sentence
- Description** : เก็บรูปแบบของประโยคและตำแหน่งของคำที่จำเป็นในการวิเคราะห์ประโยคเพื่อหา Keyword
- Data Store Name** : Internal Document
- Description** : เก็บข้อมูลการท่องเที่ยว และข้อมูลที่เป็นสำหรบนักท่องเที่ยว เช่น สถานที่ท่องเที่ยวที่สำคัญ ตารางเวลาเดินรถ แหล่งการค้า อยู่ในรูปของ Keyword และ URL ของ Web site ที่มีข้อมูลเหล่านี้
- Data Store Name** : WWW Document
- Description** : Document ที่เก็บอยู่ใน Search Engine ชื่อต่าง ๆ เช่น Yahoo, Infoseek

3.3.4 File Structure

File Name : Noun

Description : เก็บคำศัพท์ที่เป็น คำ Noun ไว้

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Number	Number	4	ลำดับคำที่
	Meaning	Number	10	ความหมาย

File Name : Verb

Description : เก็บคำศัพท์ที่เป็น คำประเภท Verb ไว้

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Number	Number	4	ลำดับคำที่
	Meaning	Number	10	ความหมาย

File Name : Adj

Description : เก็บคำศัพท์ที่เป็น คำ Adjective ไว้

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Number	Number	4	ลำดับคำที่
	Meaning	Number	10	ความหมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

File Name : Adv

Description : เก็บคำศัพท์ที่เป็น คำประเภท Adverb ไว้

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Number	Number	4	ลำดับคำที่
	Meaning	Number	10	ความหมาย

File Name : Aux

Description : เก็บคำศัพท์ที่เป็น คำที่เป็นประเภท Article, Auxiliry Verb, Wh Question, Pronoun, Preposition, Quantifier

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Type	Number	2	ชนิดของคำ
	Number	Number	4	ลำดับคำที่
	Meaning	Number	10	ความหมาย

File Name : Rule

Description : เป็นตารางที่ใช้เก็บกฎต่าง ๆ

Structure

Key	Field Name	Type	Length	Description
*	Type	Text	20	ชนิด
	Compound	Number	4	องค์ประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

File Name : Exchange

Description : เก็บความหมายของศัพท์ที่ใช้ถามกับสิ่งที่ผู้
ใช้ต้องการ

Structure

Key	Field Name	Type	Length	Description
*	Word	Text	20	คำศัพท์
	Exchange	Number	20	คำที่ใช้แทนที่



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

โปรแกรม

4.1 Hardware Requirements

อุปกรณ์คอมพิวเตอร์ที่ต้องการเมื่อใช้ระบบนี้ประกอบด้วย

1. IBM PC Compatible
2. CPU Pentium 100 MHz. or higher
3. Ram 32 MB.
4. Mouse
5. Harddisk space 100 MB
6. Network Interface Card

4.2 System Requirements

โปรแกรมที่ใช้พัฒนาระบบมีดังนี้

1. Software Requirements
2. Windows 98
3. Web Server Software
4. Microsoft Internet Explorer Version 5.5
5. Active Perl 5.x for Windows

4.3 Installation

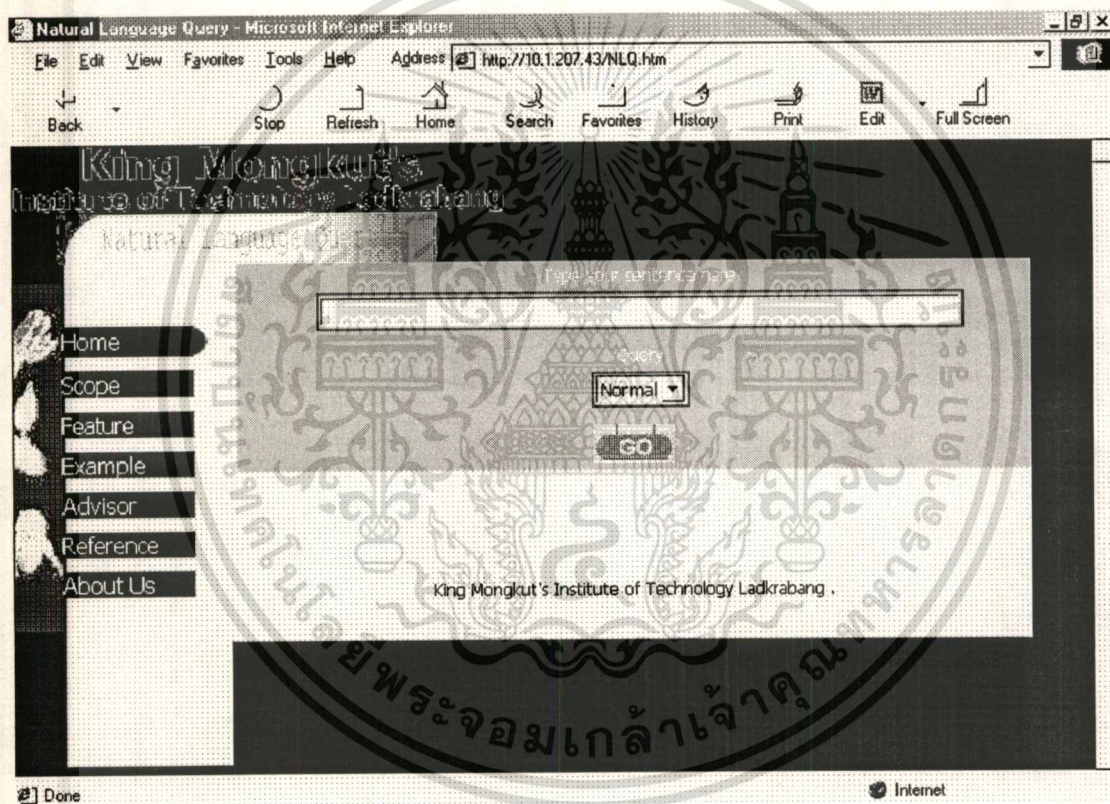
4.3.1 Server Installation

ในส่วนของ Server จะเป็นตัวให้บริการกับ User ที่ต้องการสอบถามข้อมูลจากระบบมีขั้นตอนการติดตั้งดังนี้

1. ติดตั้ง Software ที่ทำให้เครื่องคอมพิวเตอร์เป็น Web Server ที่ Support Perl CGI
2. Copy Program ลง ไปยัง Directory ที่จะถูก Map เป็น Directory cgi-bin
3. Copy Database แล้ว Set ODBC ให้มี Data Source Name ชื่อ Dictionary

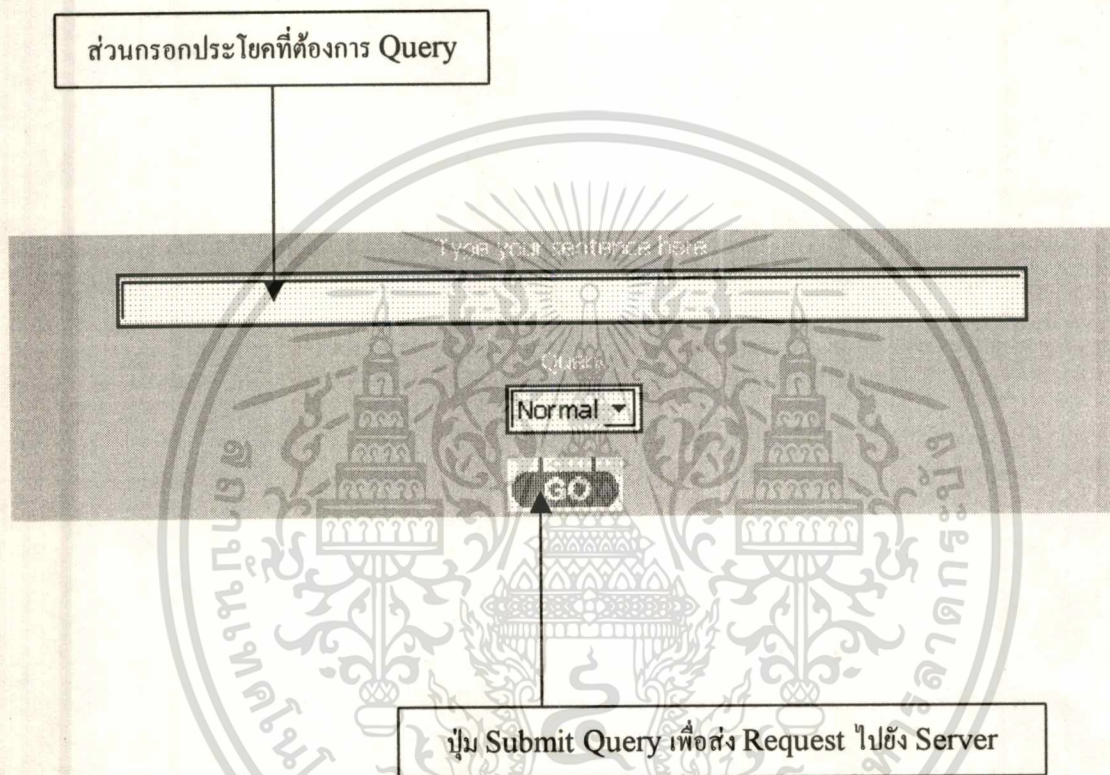
4.3.2 Client Installation

ในส่วนของ Client หรือส่วนของผู้ใช้มี Browser และเชื่อมต่อ Network โดยใช้ Protocol TCP/IP เมื่อต้องการเรียกใช้ระบบ Natural Language Query ทำได้โดย Run Browser โดยเรียกมายัง URL ของ Web Server ที่ให้บริการจะแสดงหน้าจอ ดังรูปที่ 4-1



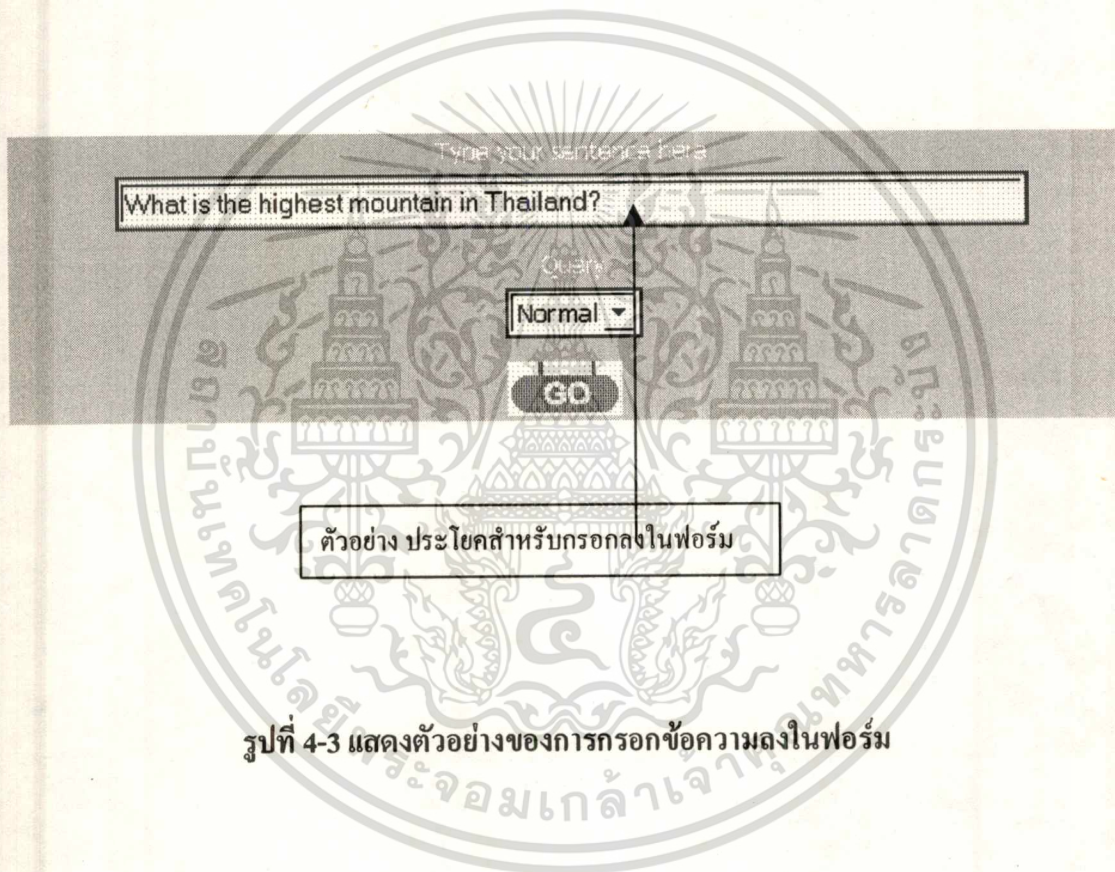
รูปที่ 4-1 แสดงหน้าจอแรกของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

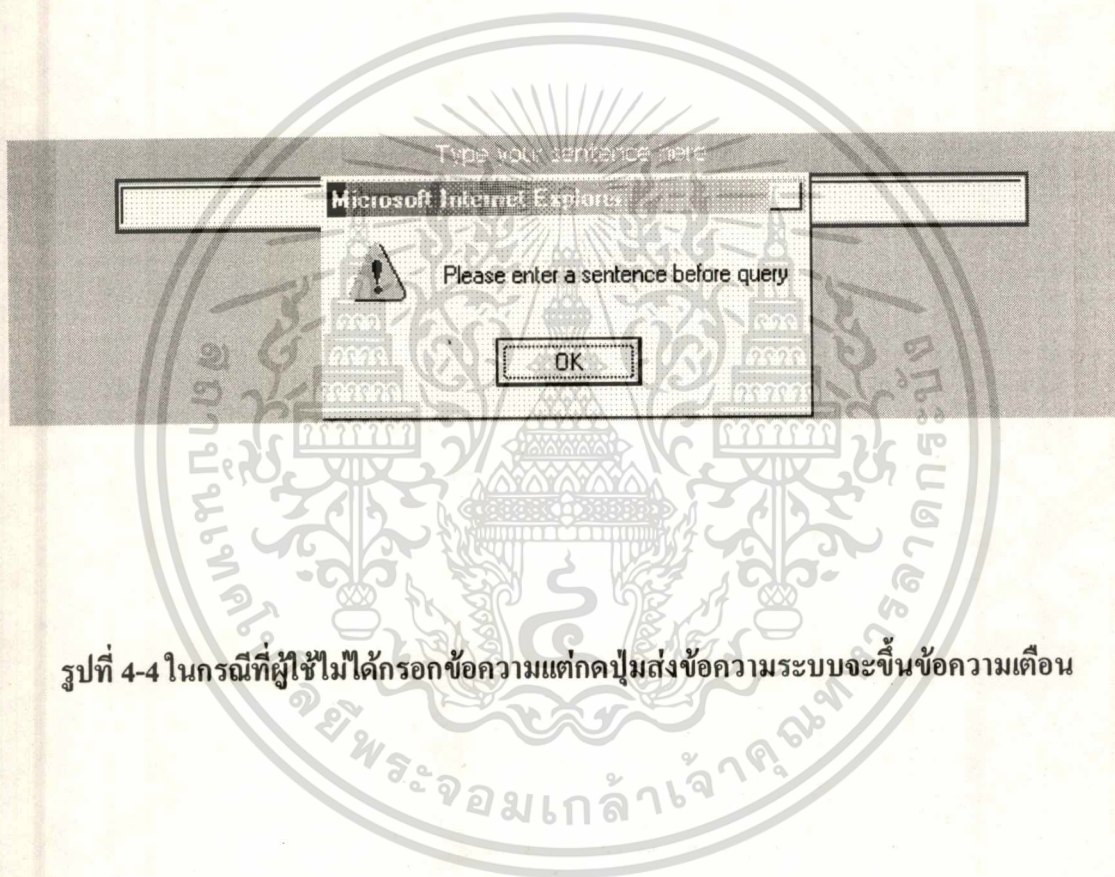


รูปที่ 4-2 ส่วนรับข้อมูลจากผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-3 แสดงตัวอย่างของการกรอกข้อความลงในฟอร์ม



รูปที่ 4-4 ในกรณีที่ผู้ใช้ไม่ได้กรอกข้อความแต่กดปุ่มส่งข้อความระบบจะขึ้นข้อความเตือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-5 เมื่อทำการส่งข้อความเรียบร้อยแล้วโปรแกรมก็จะส่งค่าที่ประมวลผลได้กลับมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การทดลอง และผลลัพธ์

2.1 วัตถุประสงค์ของการทดลอง

1. เพื่อทดสอบผลลัพธ์ที่ได้จากการ Query Sentence ของผู้ใช้งานว่าได้ผลลัพธ์ตามต้องการหรือไม่
2. เพื่อตรวจสอบ Query Sentence ของผู้ใช้งานถูก Convert เป็น Boolean Logic Statement ถูกต้องและสอดคล้องตามลักษณะของประโยคหรือไม่
3. เพื่อหาข้อมูลจำกัดของระบบ Natural Language Query เพื่อจะได้ปรับปรุงแก้ไขให้มีประสิทธิภาพมากขึ้น
4. เพื่อทดสอบประโยคที่ระบบสามารถรับรู้ได้

2.2 Environment

สภาพแวดล้อมที่ใช้ในการทดลองประกอบด้วย

1. เครื่อง Server

Hardware

- PC Computer Pentium II 350 MHz
- Ram 128 MB
- Harddisk 4 GB
- SVGA Monitor และ Mouse

Software

- Windows 98
- Web Server Software OmniHTTPd Professional 2.07
- ตัวแปลภาษา Perl Version 5 for Windows

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เครื่อง Client

Hardware

- PC Computer Pentium II 350 MHz
- Ram 128 MB
- Harddisk 4 GB
- SVGA Monitor และ Mouse

Software

- Web Browser Internal Explorer 5.5

5.3 การทดลอง

กำหนดให้

0	=	Noun
1	=	Verb
2	=	Adjective
3	=	Adverb
4	=	Aux
5	=	Article
6	=	Question Word
7	=	Pronoun
8	=	Preposition
9	=	Quantifier

S = NP + VP

S = Q + NP

S = Q + NP + VP

NP = NP + NP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อแทนด้วยตัวเลข

Q	=	[64], [604], [68], [17]
NP	=	[50], [080], [0800], [0], [7], [20], [520], [80]
VP	=	[1], [31], [310], [530]

5.3.1 ตัวอย่างของประโยคที่สามารถวิเคราะห์ได้อย่างถูกต้อง

-What is the name of Thai Currency

แจกคำศัพท์แต่ละคำ

What	Q.[6]
is	aux.[4]
the	art.[5]
name	n.[0] vt.[1]
of	pre.[8]
Thai	n.[0] adj.[2]
currency	n.[0]

เมื่อนำมารเรียงประโยคจะได้รูปแบบที่เป็นไปได้ดังนี้

6450800

6450820

6451800

6451820

นำแต่ละประโยคมาวิเคราะห์ตามกฎไวยากรณ์ได้ดังนี้

→ [64] [50] [800]	=	Q + NP + NP
→ [64] [50] 820	=	ไม่ใช่ประโยค
→ [64] 51 800	=	ไม่ใช่ประโยค
→ [64] 51 820	=	ไม่ใช่ประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบประโยคที่ถูกต้อง คือ

What is the name of Thai currency

6 4 5 0 8 0 0

q aux art noun prep noun noun

keyword จะวิเคราะห์เอาคำ noun เป็นหลัก จะได้คำว่า

name, Thai, currency

-What is Thailand locate?

[64] [0] [1]

Keyword = Thailand

-How about climate of Thailand?

[68] [080]

Keyword = climate, Thailand

-What are the customer regulations

[64] [50] [0]

Keyword = customer, regulations

-What is The religion of population?

[64] [50] [80]

Keyword = religion, population

-What is the name of Thai currency?

[64] [50] [80] [0]

Keyword = name, Thai, currency

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.2 ตัวอย่างรูปแบบประโยคที่มีปัญหาในการวิเคราะห์

-What is the best time to visit Thailand?

รูปแบบประโยคที่น่าจะเป็นไปได้สำหรับประโยคนี้

→ [64] [50] [0] [310] = Q + NP + N + VP

→ [64] [50] [0800] = Q + NP + NP

→ [64] [50] [1] [310] = Q + NP + VP

→ [64] [520] [310] = Q + NP + VP

→ [64] [520] [80] [0] = Q + NP + NP + NP

ประโยคนี้มีความเป็นไปได้ที่จะเป็นประโยคได้หลายรูปแบบ จากกฎที่ได้กล่าวมาข้างต้น การพิจารณาว่าตำแหน่งใดเป็นคำ noun หรือ verb ที่แน่นอนเป็นเรื่องยากการกรอง keyword ถ้าเลือกประโยคผิดจะทำให้ได้ คำออกมาผิดเช่นในประโยคแรก

[64] [50] [0] [310]

Key word อาจจะเป็น best, time, Thailand

[64] [50] [0800]

Key word อาจจะเป็น best, time, visit, Thailand

[64] [50] [1] [310]

Key word อาจจะเป็น best, Thailand

[64] [520] [310]

Key word อาจจะเป็น time, Thailand

[64] [520] [80] [0]

Key word อาจจะเป็น time, visit, Thailand

-What is the highest mountain in Thailand ?

[64] [520] [0] [0]

Key word **จะเป็น** mountain, in, Thailand

[64] [520] [20]

Key word **จะเป็น** mountain, Thailand

[64] [520] [80]

Key word **จะเป็น** mountain, Thailand

-What is the famous temple in Bangkok?

[64] [520] [0] [0]

Key word **จะเป็น** temple, in, Bangkok

[64] [520] [20]

Key word **จะเป็น** temple, Bangkok

[64] [520] [80]

Key word **จะเป็น** temple, Bangkok

-Show me the map of Thailand

[0] [7] [50] [80]

Key word **จะเป็น** Show, map, Thailand

[17] [50] [80]

Key word **จะเป็น** map, Thailand

-Where is Jatujak park?

[64] [0] [0]

Key word จะเป็น Jatujak, park

[64] [0] [1]

Key word จะเป็น Jatujak

-What is the famous market in Bangkok?

[64] [520] [0] [0]

Key word จะเป็น market, in, Bangkok

[64] [520] [80]

Key word จะเป็น market, Bangkok



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

บทสรุป

ข้อสรุป และผลที่ได้รับจากโครงการ

ระบบ Natural Language Query for tourism เป็นระบบที่ช่วยให้นักท่องเที่ยว และชาวต่างประเทศที่ต้องการข้อมูลการท่องเที่ยว หรือข้อมูลเกี่ยวกับประเทศไทยเบื้องต้นเพื่อประโยชน์ในการติดต่อกับประเทศไทย หรือประกอบการตัดสินใจเพื่อเข้ามาเที่ยวเมืองไทย ซึ่งการสอบถามข้อมูลไม่จำกัดอยู่ที่การค้นหาโดยใช้ Keyword แต่ยังสามารถสอบถามโดยใช้ประโยคคำถาม หรือประโยคคำสั่งเพื่อสอบถามข้อมูลจาก Database หรือ Search Engine ที่มีข้อมูลเกี่ยวกับการท่องเที่ยว

ข้อสรุปจากโครงการมีดังนี้

1. ได้ใช้ความรู้ทางวิชาการดังนี้
 - 1.1 นำความรู้ Artificial Intelligent (AI) ในสาขา Natural Language Processing มาใช้ในการวิเคราะห์ประโยคที่ผู้ใช้ป้อนเข้ามาเพื่อหา Keywords ที่ผู้ใช้ต้องการค้นหา
 - 1.2 นำความรู้ด้าน Information Retrieval มาใช้สร้าง Lexicon และ Search Engine เพื่อรับ Keyword ที่ได้จากการวิเคราะห์มา Search ใน Documents ที่เก็บไว้
 - 1.3 นำความรู้ไวยากรณ์ภาษาอังกฤษเบื้องต้นมาใช้ในการวิเคราะห์ประโยค (Sentence)
2. รวบรวมข้อมูลที่เป็นสำหรับนักท่องเที่ยวไทย เช่น แหล่งท่องเที่ยวในประเทศไทย, ตารางเวลาการบิน, ตารางเวลาการเดินทาง, ร้านอาหาร, กวดคาร, โรงแรม
3. เพื่อเป็นแนวทางในการพัฒนา และเพิ่มประสิทธิภาพในการค้นหาข้อมูล Search Engine ให้ได้ผลลัพธ์ตรงความต้องการของผู้ใช้ โดยใช้เวลาน้อยที่สุด
4. ทำให้ผู้ใช้สามารถป้อน Query ข้อมูลโดยใช้ประโยคคำถาม หรือประโยคง่ายๆ ที่นิยมใช้ในชีวิตประจำวันได้

บรรณานุกรม

ชินขง ภู่วรรณ และชัยขงศ์ วงศ์ชัยสุวัฒน์. การประมวลผลภาษาธรรมชาติ.

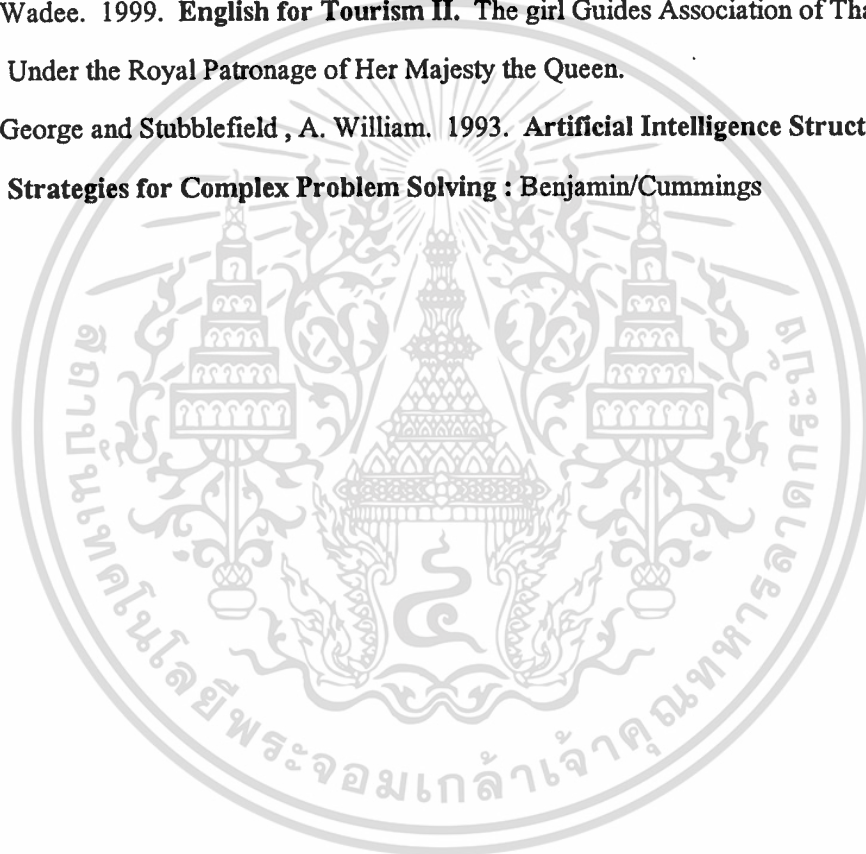
ชินขง ภู่วรรณ และชัยขงศ์ วงศ์ชัยสุวัฒน์. 2531. **Electronic Dictionary for Thai Data Processing.** ปทุมธานี : สถาบันเทคโนโลยีราชมงคล.

อัศนีย์ ก่อตระกูล. 2535. **ปัญญาประดิษฐ์.** กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์

<http://member.tripod.com/nlq/doc/index.html>. Natural Language Query (NLQ)

Kheourai Wadee. 1999. **English for Tourism II.** The girl Guides Association of Thailand
Under the Royal Patronage of Her Majesty the Queen.

Luger, F. George and Stubblefield , A. William. 1993. **Artificial Intelligence Structures and Strategies for Complex Problem Solving : Benjamin/Cummings**



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้