

ระบบฐานความรู้เกี่ยวกับ

ประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา

KNOWLEDGE BASE SYSTEM
OF THAI AYUDDHAYA'S HISTORY

นายคมศักดิ์	สุขเกษม	รหัส 41360231
นายวิวัฒน์	อริยะสุนทร	รหัส 41360447
นายสุรชาติ	โอปุ่น	รหัส 41360520

5894045 ๑/๒
มร.
ต/๕๒๕
๖๖๕๕

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิศวกรรมคอมพิวเตอร์ ภาควิชาไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

ปีการศึกษา ๒๕๔๔

ห้องสมุดคณะวิศวกรรมศาสตร์	ร.ร. ๒๕๔๔
วันที่รับ...../...../.....	
เลขทะเบียน.....	๗๖.๙.๐.๑๒
เลขเรียกหนังสือ.....	๑๕
มหาวิทยาลัยนเรศวร	๗๖.๙.๐.๑๒
	๑๕๒๕

๖๕๔๔ C.๒



ใบรับรองโครงการงานวิจัย

หัวข้อโครงการ ระบบฐานความรู้เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา
ผู้ดำเนินโครงการ นายคมศักดิ์ สุขเกษม รหัส 41360231
นายวิทวัส อริยะสุนทร รหัส 41360447
นายสุรชาติ โอบั่น รหัส 41360520
อาจารย์ที่ปรึกษา อาจารย์วัชรวีร์ พิชพันธ์
สาขา วิศวกรรมคอมพิวเตอร์
ภาควิชา วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ปีการศึกษา 2544

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี อนุมัติให้โครงการฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ คณะกรรมการสอบโครงการงานวิจัย

.....กรรมการ
(อาจารย์สิทธิโชค เขาวกุล)

.....กรรมการ
(อาจารย์ประทีป ตริณโสภาส)

.....กรรมการ
(อาจารย์วัชรวีร์ พิชพันธ์)

.....กรรมการ
(อาจารย์ศิริพร เดชะศิริรักษ์)

หัวข้อโครงการ ระบบฐานความรู้เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา

ผู้ดำเนินโครงการ นายคมศักดิ์ สุขเกษม รหัส 41360231

นายวิหวัธ อริยะสุนทร รหัส 41360447

นายสุรชาติ โอปิ่น รหัส 41360520

อาจารย์ที่ปรึกษา อาจารย์วัชรวีร์ พิษพันธ์

สาขา วิศวกรรมคอมพิวเตอร์

ภาควิชา วิศวกรรมไฟฟ้าและคอมพิวเตอร์

ปีการศึกษา 2544

บทคัดย่อ

โครงการระบบฐานความรู้เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาเป็นโปรแกรมที่พัฒนาโดยมีวัตถุประสงค์เพื่อเผยแพร่ความรู้เกี่ยวกับประวัติศาสตร์ไทย ในสมัยกรุงศรีอยุธยาในลักษณะของการถามคำถามไปยังคอมพิวเตอร์ แล้วคอมพิวเตอร์จะตอบคำถามที่เราต้องการกลับมายังผู้ถามคำถามนั้น โดยการตอบคำถามของคอมพิวเตอร์นั้นจะไม่ได้เป็นลักษณะของการจับคู่กันของคำถามและคำตอบ แต่จะเป็นลักษณะของการที่คอมพิวเตอร์สร้าง (Generate) คำตอบนั้นออกมาเองโดยอาศัยการตัดคำ การสร้างประโยค ฐานความรู้เรื่องประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา และความรู้เรื่องปัญญาประดิษฐ์ (Artificial Intelligence) การพัฒนาโปรแกรมนี้ออกมาและทำงานบนระบบปฏิบัติการไมโครซอฟท์วินโดวส์ 98 ใช้ภาษาวิซวลซี พลัสพลัสเวอร์ชัน 6.0 (Visual C++ version 6.0) ในการพัฒนา

ผลที่ได้จากการทำโครงการนี้คือได้โปรแกรมที่มีความสามารถสร้างประโยคขึ้นเองได้ ลักษณะการทำงานของโปรแกรมนี้นี้คือ ผู้ใช้จะถามคำถามที่เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา แล้วคอมพิวเตอร์มีกระบวนการสร้างคำตอบแล้วตอบออกมายังผู้ใช้ เสมือนว่าคอมพิวเตอร์เป็นผู้ที่มีความรู้เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาและตอบคำถามเกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาได้

Project Title Knowledge Base System of Thai Ayuddhaya's History

Name Mr. Komsak Sookkasame ID. 41360231

Mr. Vittawat Ariyasontorn ID. 41360447

Mr. Surachat Opun ID. 41360520

Project Advisor Mr. Watchawee Phuchpan

Major Computer Engineering

Department Electrical and Computer Engineering

Academic Year 2001

.....

ABSTRACT

This Question and Answer system of Thai Ayuddhaya's history is developed to spread the knowledge of Thai Ayuddhaya's history. We can ask questions to computer and then it will answer the question. The method doesn't be a matching between question and answer. Computer will generate the answer by the knowledge of parsing word, sentence's generation, and knowledge base of history Thai Ayuddhaya's and knowledge of Artificial Intelligence. The way to develop this program is tested and run on Microsoft Windows 98 Operating System. Use Visual C++ version 6.0 language to develop this program.

The result of this project is a program that able to generate sentence by itself. The utilization of this program is using for the user ask the question about Thai Ayuddhaya's history. And then computer will have process to generate answer sentence and answer back to user.

กิตติกรรมประกาศ

ในการทำโครงการครั้งนี้ คณะผู้จัดทำได้รับความอนุเคราะห์เป็นอย่างดีจากบุคคลหลายฝ่าย หลายหน่วยงานของคณะวิศวกรรมศาสตร์ อันทำให้สามารถทำงานได้สำเร็จลุล่วงไปด้วยดี คณะผู้จัดทำจึงขอกราบขอบพระคุณมา ณ ที่นี้

คณะผู้จัดทำขอกราบขอบพระคุณ อาจารย์ยัชรวีร์ พิษพันธ์ ซึ่งเป็นอาจารย์ที่ปรึกษาโครงการ ตามที่ท่านได้ให้ความกรุณาแนะนำวิธีในการทำงาน ให้เข้าใจถึงการศึกษารูปแบบเป็นระบบ ขั้นตอน รวมทั้งระยะเวลาเพื่อตรวจสอบการทำงานและชี้แนวทางแก้ไขในทุกขั้นตอนของการศึกษา

ขอขอบพระคุณ อาจารย์ศิริพร เจริญศิริรักษ์ ที่ช่วยตรวจทานความถูกต้องของรายงาน และให้คำปรึกษาในด้านต่างๆ ในการทำโครงการครั้งนี้

ขอขอบคุณอาจารย์ทุกท่านที่ได้กล่าวนามไว้ ณ ที่นี้ได้ประสิทธิ์ ประศาสตร์ความรู้ให้แก่คณะผู้จัดทำตลอดมา

นายคมศักดิ์ สุขเกษม

นายวิทวัส อริยะสุนทร

นายสุรชาติ โอบั่น



สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	๒
บทคัดย่อภาษาอังกฤษ.....	๓
กิตติกรรมประกาศ.....	๔
สารบัญ.....	๖
สารบัญตาราง.....	๗
สารบัญรูป.....	๗

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 ขอบข่ายของโครงการ.....	1
1.4 ขั้นตอนการดำเนินงาน.....	2
1.5 ผลที่คาดว่าจะได้รับ.....	3
1.6 งบประมาณที่ใช้.....	3

บทที่ 2 การพัฒนาโปรแกรมด้วย Microsoft Visual C++ และกระบวนการภาษาธรรมชาติ

2.1 การเขียนโปรแกรมด้วย Microsoft Visual C++.....	4
2.2 การติดต่อฐานข้อมูลด้วย DAO.....	6
2.3 Augmented Translation Network สำหรับโครงสร้างประโยคภาษาไทย.....	9
2.4 ATN สำหรับประโยคคำถาม.....	12
2.5 การตัดคำ.....	14
2.6 วิธีการตัดคำ.....	14
2.7 โครงสร้างข้อมูลแบบทรี(Trie).....	20

บทที่ 3 วิธีการดำเนินงาน

3.1 วิธีการศึกษาค้นคว้าวิจัย.....	22
3.2 ศึกษาลักษณะของข้อมูล การเลือกข้อมูล และเหตุผลในการเลือก.....	22
3.3 การออกแบบฐานข้อมูล.....	24
3.4 เขียนโปรแกรมระบบฐานความรู้.....	29
3.5 ทำการติดตั้งและทดลองโปรแกรม.....	29

บทที่ 4 การออกแบบโปรแกรม

4.1 ส่วนประกอบของโปรแกรม.....	30
4.2 ศักยภาพงานแสดงการทำงานของโปรแกรม.....	30
4.3 ผังโปรแกรมแสดงการทำงานของโปรแกรม.....	31
4.4 ตัวอย่างการใช้งานโปรแกรม.....	32
บทที่ 5 สรุปและวิเคราะห์ผล	
5.1 สรุป.....	34
5.2 ข้อเสนอแนะ.....	35
เอกสารอ้างอิง.....	36
ประวัติผู้เขียนโครงการ.....	37



สารบัญตาราง

ตารางที่	หน้า
1.1 ขั้นตอนการดำเนินงาน.....	2
1.1(ต่อ) ขั้นตอนการดำเนินงาน.....	3
2.1 การตัดค่าแบบเลือกค่ายาวที่สุด.....	17
5.1 ตัวอย่างประโยคคำถามที่ได้สุ่มมาจากผู้ใช้.....	34



สารบัญรูป

รูปที่	หน้า
2.1	10
2.1(ต่อ)	11
2.2	12
2.2(ต่อ)	13
2.3	20
2.4	20
3.1	23
3.2	25
3.3	29
4.1	30
4.2	31
4.3	32
4.4	32
4.5	33

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของโครงการ

ในปัจจุบันเราจะเห็นการถามตอบ ปัญหาในคอมพิวเตอร์นั้นจะมีลักษณะของการที่เราต้องใส่คำถาม (Keyword) เพื่อให้ตรงกับคำตอบที่เตรียมเอาไว้ ตัวอย่างที่เห็นได้ชัดเจน เช่น ระบบช่วยเหลือที่มีอยู่ในโปรแกรมต่างๆ เช่นระบบช่วยเหลือใน โปรแกรมจัดทำเอกสาร(Word) เป็นต้น ระบบช่วยเหลือเหล่านี้จะมีไว้เวลาเราเกิดข้อข้องใจหรือต้องการคำแนะนำเวลาเราต้องการ แต่ว่าการใช้งานของระบบช่วยเหลือเหล่านี้ เราต้องพิมพ์คำหลัก (Keyword) ที่เป็นคำถามให้ตรงกับคำตอบที่เก็บไว้หรือที่โปรแกรมเตรียมไว้นั่นเอง ถ้าเราพิมพ์คำถามที่เป็นภาษาธรรมชาติ (Natural Language) ระบบช่วยเหลือดังกล่าวก็ไม่สามารถที่จะตอบคำถามที่เราต้องการได้

ดังนั้นเราน่าจะมีการสื่อสารกับคอมพิวเตอร์ได้โดยตรงโดยใช้ภาษาธรรมชาติ ได้ตอบปัญหา กับคอมพิวเตอร์ได้ โดยอินพุทข้อความถามเข้าไปเป็นภาษาธรรมชาติ แล้วคอมพิวเตอร์สามารถเข้าใจว่าผู้ใช้ต้องการทราบอะไร จากนั้นคอมพิวเตอร์ก็จะมีกระบวนการสร้างประโยคคำตอบแล้วตอบกลับมาให้ผู้ใช้ ซึ่งจะทำให้เราไม่ต้องกังวลว่าเราจะไม่ได้คำตอบที่เราถามเข้าไปดังเช่นที่เราเห็นกันอยู่บ่อยๆ

ลักษณะของโครงการชิ้นนี้จึงจัดทำขึ้น โดยจัดทำเป็นโปรแกรมที่คอมพิวเตอร์สามารถเข้าใจคำถาม และสามารถตอบคำถามที่ถามไปได้ ผู้จัดทำจัดทำเป็นโปรแกรมที่มีการถามตอบเกี่ยวกับประวัติศาสตร์ไทยในสมัยอยุธยา ทั้งนี้ทั้งนั้นเนื่องจากว่านอกจากเราจะได้รับความรู้ในเรื่องของการตัดคำ การสร้างประโยคแล้ว เรายังได้ความรู้ในเรื่องของประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาด้วย

1.2 วัตถุประสงค์ของโครงการ

1. เพื่อสร้าง โปรแกรมที่มีความสามารถตอบคำถามเกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาได้
2. เพื่อศึกษาการเก็บฐานความรู้และค้นหาความรู้ของประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา
3. เพื่อประยุกต์ใช้กระบวนการเข้าใจภาษาธรรมชาติ (Natural Language Understanding) มาใช้การสร้างประโยค
4. สามารถค้นหาความรู้ที่เก็บในฐานความรู้ได้โดยง่าย

1.3 ขอบข่ายของโครงการ

1. ผู้ใช้สามารถติดต่อกับโปรแกรม โดยการพิมพ์ตามคำถามที่ต้องการได้
2. โปรแกรมสามารถตอบคำถามของผู้ใช้ได้อย่างรวดเร็ว โดยตรงตามคำถามที่ผู้ใช้พิมพ์ถามเข้ามา
3. ฐานความรู้เป็นการเก็บความรู้ที่เก็บจากหนังสือเส้นทางประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา [6]

4. เมื่อผู้ใช้งานคำถามที่ไม่เกี่ยวข้องกับความรู้ที่มีอยู่โปรแกรมจะไม่สามารถตอบได้
5. โปรแกรมสามารถติดตั้งได้ง่าย และสะดวกในการใช้งาน ซึ่งผู้ใช้สามารถศึกษาได้จากคู่มือการใช้งาน

1.4 ขั้นตอนการดำเนินงาน

1. วางแผนการทำงาน เก็บรวบรวมข้อมูลที่เกี่ยวข้องทั้งหมด พร้อมทั้งศึกษาความรู้ที่จะนำมาใช้ในการทำงาน

2. ศึกษาทฤษฎีที่เกี่ยวข้อง
3. ออกแบบฐานข้อมูลและเก็บข้อมูล
4. ทำการออกแบบ โปรแกรมและเขียน โปรแกรม
5. ทดลองใช้งาน โปรแกรม หาข้อบกพร่องของโปรแกรม และทำการแก้ไขโปรแกรม
6. สรุปผลการทำงาน

รวมระยะเวลาในการทำโครงการทั้งสิ้น 9 เดือน

ตารางที่ 1.1 ขั้นตอนการดำเนินงาน

กิจกรรม	เดือน-ปี							
	ม.ค.2544	ก.พ.2544	มี.ค.2544	พ.ค.2544	มิ.ย.2544	ก.ค.2544	ธ.ค.2544	ก.ย.2544
1.วางแผนการทำงาน และเก็บรวบรวมข้อมูล	←————→							
2.ศึกษาทฤษฎีที่เกี่ยวข้อง		←————→						
3.ออกแบบฐานข้อมูลและเก็บข้อมูล			←————→					
4.ออกแบบโปรแกรมและเขียนโปรแกรม		←————→						

5.ทดลองใช้ งานจริงและ ปรับปรุงแก้ไข								
6.สรุปผล								

ตารางที่ 1.1(ต่อ) ขั้นตอนการดำเนินงาน

1.5 ผลที่คาดว่าจะได้รับ

1. ทราบถึงวิธีการสร้างประโยคโดยใช้กระบวนการภาษาธรรมชาติ(Natural Language Processing)
2. ได้รับความรู้เกี่ยวกับปัญญาประดิษฐ์ การตัดคำ การค้นหาคำ การติดต่อฐานข้อมูล ตลอดจนทราบถึงวิธีการสร้างประโยคภาษาไทย
3. ได้โปรแกรมที่สามารถตอบคำถามเกี่ยวกับประวัติศาสตร์สมัยกรุงศรีอยุธยาได้ โดยที่ลักษณะของการตอบคำถามไม่ได้เป็นลักษณะของการจับคู่กันระหว่าง คำถามและคำตอบแต่เป็นคำตอบที่เกิดจากการสร้างประโยคโดยคอมพิวเตอร์เอง
4. สามารถพัฒนาโปรแกรมต่อไปเพื่อใช้กับบุคคลพิการตาบอดได้ โดยพัฒนาให้โปรแกรมเป็นลักษณะของการถามตอบโดยใช้เสียงพูดแทนการพิมพ์

1.10 งบประมาณที่ใช้

งบประมาณที่ใช้เฉลี่ยทุกรายการเป็นเงิน 3,000 บาท

บทที่ 2

การพัฒนาโปรแกรมด้วย Microsoft Visual C++ และกระบวนการ ภาษาธรรมชาติ

2.1 การเขียนโปรแกรมด้วย Microsoft Visual C++

Microsoft Visual C++ เป็นโปรแกรมประเภท Visual อีกตัวของบริษัทไมโครซอฟต์ผู้ผลิตระบบปฏิบัติการวินโดวส์ Visual C++ ได้รับการพัฒนาให้มีความยืดหยุ่นและมีประสิทธิภาพสูงขึ้นมา จากภาษา C++ และได้สนับสนุนการพัฒนาโปรแกรมในหลายๆ ด้าน

ประวัติของ Visual C++ และ MFC

Visual C++ ได้รับการพัฒนาขึ้นมาจาก Microsoft C/C++ ให้เป็น IDE ที่ทำงานบนระบบปฏิบัติการวินโดวส์ได้อย่างเต็มที่ รองรับการพัฒนาโปรแกรมบนวินโดวส์โดยมี MFC (Microsoft Foundation Class) เป็นไลบรารีที่ช่วยอำนวยความสะดวกในการพัฒนาโปรแกรมบนวินโดวส์

ในยุคที่ MFC ยังไม่ถือกำเนิด เมื่อเราจะพัฒนาโปรแกรมบนวินโดวส์จะต้องใช้ SDK (Software Development Kit) และคอมไพเลอร์ภาษาซี เช่น Microsoft C++, Borland C++ ช่วยในการเขียนโปรแกรม โค้ดโปรแกรมที่เขียนขึ้นด้วย SDK นี้ค่อนข้างซับซ้อนและสร้างความลำบากในการศึกษาทำความเข้าใจ จึงได้มีการสร้างคลาสขึ้นมาชุดหนึ่ง เขียนขึ้นโดยใช้โครงสร้างของ OOP (Object Oriented Programming) ด้วยภาษา C++ ชื่อว่า MFC หรือ Microsoft Foundation Class ขึ้นมา โค้ดโปรแกรมที่เขียนขึ้นโดยใช้ MFC นั้นจะมีขนาดเล็กและไม่ซับซ้อน ทำให้การพัฒนาโปรแกรมบนวินโดวส์สามารถทำได้ง่ายกว่าการใช้ SDK

แนวคิดของ OOP

โอโอพี (OOP) หรือ Object Oriented Programming เป็นแนวคิดในการเขียนโปรแกรมแบบหนึ่ง ซึ่งนิยามว่าเป็น การเขียนโปรแกรมเชิงวัตถุ

ถ้าเราไม่มองในแง่มุมมองของการเขียนโปรแกรมเพียงอย่างเดียว ให้เรามองภาพรวมมองไปในสิ่งรอบๆ ตัวเรา เราบอกได้ว่า แนวคิดของ OOP ก็คือ “ธรรมชาติของวัตถุ” หมายความว่า OOP จะมองสิ่งแต่ละสิ่งเป็น “วัตถุชิ้นหนึ่ง” มันจะยาวหรือสั้น หรือมีสีอะไรก็ตาม มันก็คือวัตถุชิ้นหนึ่งเหมือนกัน ซึ่งวัตถุแต่ละสิ่งนั้น ย่อมมีคุณสมบัติที่ต่างกัน แต่อาจมีอย่างเหมือนกัน เรายังสามารถคิดได้อีกว่า “วัตถุแต่ละอย่างนั้น ต่างก็มีลักษณะและวิธีการใช้เป็นตัวของมันเอง” ซึ่งหมายความว่า วัตถุแต่ละชนิดหรือ

แต่ละชิ้น ต่างก็มีรูปร่างลักษณะ และการใช้งาน ที่แตกต่างกันไป เราเรียกคุณลักษณะของวัตถุว่า คุณลักษณะของวัตถุ (Attribute) และเรียกการใช้งานของวัตถุว่า เมธอด(Method) เช่น

“คินสอเป็นวัตถุที่มีลักษณะที่เรียวยาว, ภายในเป็นไส้ถ่านใช้สำหรับเขียน การใช้งานคินสอทำได้ โดยใช้มือจับและเขียนบนวัตถุรองรับ”

จากประโยคข้างต้น เราสามารถจับใจความได้ว่า คุณลักษณะของวัตถุ (Attribute) ก็คือ “ ยาว เรียว, ภายในเป็นไส้ถ่าน” ส่วนการใช้งาน (Method) ก็คือ “ใช้มือจับและเขียนลงบนวัตถุรองรับ”

จากตัวอย่างข้างต้น คราวนี้เราสรุปได้แล้วว่า ถ้าเกิดวัตถุใดมีลักษณะยาวเรียว, มีไส้เป็นถ่าน เมื่อใช้งานต้องใช้มือจับและเขียนบนวัตถุรองรับ-เราก็สามารถบอกได้ว่าวัตถุนั้นคือ “คินสอ”

คลาสและออบเจกต์ในภาษา C++

เรานำแนวคิด OOP มาเขียนโปรแกรมด้วยภาษา C++ โดยเราสมมติคลาสขึ้นมา 1 คลาส โดยเป็นคลาสของคินสอ ใช้ชื่อว่าคลาส Cpencil เราจะได้โค้ดดังนี้

```
class Cpencil {
    int length;
    void write (char * what);
};
```

จากโปรแกรมข้างต้นจะเห็นได้ว่า การประกาศคลาส Cpencil มีลักษณะคล้ายกับการประกาศโครงสร้างโดยใช้ struct โดยภายในคลาสจะประกอบไปด้วยคุณลักษณะ (Attribute) และวิธีการ (Method) โดยตัวแปร length จะเป็นคุณลักษณะจะใช้สำหรับเก็บค่าความยาวของคินสอ และฟังก์ชัน write() เป็นวิธีการใช้งานของคินสอ เพราะฉะนั้นเมื่อเราต้องการสร้างคินสอ 1 แท่ง ต้องเขียน โค้ดดังนี้

```
Cpencil pencil1;
```

การประกาศค่าในบรรทัดข้างต้น เราเรียกว่า “การประกาศออบเจกต์ (Object)” ซึ่งกำหนดให้ pencil1 เป็นตัวแปรคลาส Cpencil โดยมี length เก็บค่าความยาวของ pencil1 และมีฟังก์ชัน write () เป็นวิธีการใช้คินสอตัวเอง

คราวนี้เราสรุปได้ว่า

คลาส(Class) ก็คือการรวมคุณลักษณะและการทำงานของวัตถุอย่างน้อยหนึ่งอย่างมาไว้ในกลุ่มเดียวกัน

ออบเจกต์(Object) วัตถุที่เป็นตัวแปรคลาส เป็นรูปแบบคลาสที่มีตัวตน และสามารถใช้งานได้

การสืบทอดของคลาส(Inheritance)

คลาสแต่ละคลาส(คลาสแม่) สามารถสืบทอดมาเป็นคลาสใหม่ได้(คลาสลูก) โดยคลาสใหม่ที่สืบทอดมานี้ จะยังมีคุณสมบัติเหมือนคลาสเดิมทุกประการ ดังโปรแกรมต่อไปนี้

```

Class C-pencil {
    int length;
    void write (char * what);
};

class CpushPencil:C-pencil {
};

```

จากโปรแกรมในข้างต้น เป็นการสร้างคลาสใหม่ที่ชื่อว่า CpushPencil โดยสืบทอดคุณสมบัติมาจากคลาสเดิมสองคลาสคือ C-pencil และ C-pencil ไม่ได้สลายหายไป แต่จะยังอยู่ และสามารถใช้งานได้เหมือนเดิมทุกประการ และเรายังสามารถเพิ่มคุณสมบัติใหม่ๆ ลงในคลาส CPushPencil ได้

2.2 การติดต่อฐานข้อมูลด้วย DAO

ในการเขียนโปรแกรมเพื่อติดต่อฐานข้อมูลของ Visual C++ นั้นจะมีอินเตอร์เฟสในการติดต่ออยู่ ทั้งนี้ DAO ก็เป็นอินเตอร์เฟสหนึ่งที่ใช้เข้าถึงข้อมูล ซึ่งข้อดีของมันคือ เมื่อใช้ฐานข้อมูลที่เป็น Access การติดต่อกับฐานข้อมูลเมื่อใช้ DAO จะมีประสิทธิภาพดีกว่าตัวอื่นๆ สามารถใช้ได้หลายภาษา และยังเขียนได้ดั่งง่าย DAO ประกอบไปด้วยคลาสต่างๆ ที่สำคัญหลายคลาส เพื่อความเข้าใจการทำงานในการติดต่อฐานข้อมูล เรามาดูความเข้าใจกับคลาสที่เกี่ยวข้อง ดังนี้

คลาส CDaoWorkspace

คลาสนี้ใช้เพื่อเข้าถึงฟังก์ชันในคำสั่งเบสเอ็นจิน ด้วยคุณสมบัติที่เด่นในการทำทรานแซกชัน ทำให้คลาสนี้นำมาใช้ในการทำทรานแซกชันเป็นส่วนใหญ่ นอกจากนี้แล้วการทำทรานแซกชันสามารถเป็นอิสระจากทรานแซกชันอื่นๆ ที่เกิดขึ้นพร้อมกันได้

ฟังก์ชันที่สำคัญของคลาสและหน้าที่ของฟังก์ชันสรุปได้ดังนี้

CDaoWorkspace	สร้างออบเจกต์ Workspace หลังจากใช้คำสั่ง Create หรือ Open
RollBack	ทำให้การทรานแซกชันปัจจุบันสิ้นสุดลง และไม่มีกรบันทึกการเปลี่ยนแปลงข้อมูล
Create	สร้างออบเจกต์ DAOWorkspace ใหม่
Open	เปิดออบเจกต์ Workspace ที่ร่วมกับ Workspace ดิฟอลต์ของ DAO ภายหลังจากออบเจกต์ CDaoWorkspace ได้ถูกสร้างขึ้น

คลาส CDaoDatabase

เป็นคลาสหนึ่งของ MFC ที่ถูกใช้สำหรับการเข้าถึงข้อมูล Access เป็นส่วนใหญ่ แต่คลาส CDaoDatabase ไม่ได้รวมฟังก์ชันทรานแซกชันที่เหมือนกับคลาส CDaoWorkspace ไว้

คลาส CDaoDatabase มีฟังก์ชันที่สำคัญๆ ดังนี้

Create	สร้างออบเจ็กต์ฐานข้อมูล และเปิดการติดต่อกับฐานข้อมูลใหม่ด้วยออบเจ็กต์ CDaoDatabase
CDaoDatabase	เป็นคอนสตรัคเตอร์ (Constructor)
Execute	รันคิวรี SQL สเตทमेंต์
m_pWorkspace	เป็นพอยน์เตอร์ชี้ไปยังออบเจ็กต์ CDaoDatabase ซึ่งประกอบด้วยออบเจ็กต์ฐานข้อมูล และกำหนดการทรานเซกชัน
M_pDAODatabase	เป็นพอยน์เตอร์ชี้ไปยังออบเจ็กต์ DAODatabase

คลาส CDaoRecordset

คลาสนี้จะแสดงเซตของเรคคอร์ดจากฐานข้อมูล ซึ่งเรียกกันว่า “เรคคอร์ดเซต” (Recordset) MFC AppWizard จะสร้างคลาสเรคคอร์ดนี้ให้แก่แอปพลิเคชันโดยอัตโนมัติ คลาส CDaoRecordset มีฟังก์ชันและค่าตัวแปรต่างๆ ที่สำคัญในการติดต่อกับฐานข้อมูลดังนี้

AddNew	ใช้สำหรับการเพิ่มเรคคอร์ดใหม่ และยังไม่มีการเพิ่มเรคคอร์ดจนกว่าจะมีการเรียกใช้ฟังก์ชัน UpDate
CDaoRecordset	สร้างออบเจ็กต์ CDaoRecordset
GetDefaultDBName	บรรจุสตริงที่เป็น Path และชื่อฐานข้อมูลสำหรับเรคคอร์ดเซตที่สร้างขึ้น
GetDefaultSQL	ใช้เพื่อเรียกสตริงที่เป็น SQL สเตทमेंต์
Move	เลื่อนเรคคอร์ดโดยการผ่านพารามิเตอร์ ถ้าพารามิเตอร์เป็นศูนย์ จะไม่เลื่อน ถ้าเป็นบวกจะไปข้างหน้า แต่ถ้าเป็นลบก็จะเลื่อนถอยหลัง
Open	สร้างเรคคอร์ดใหม่จากตาราง Dynaset หรือ Snapshot
m_nParams	เก็บจำนวนค่าตัวแปรพารามิเตอร์ในคลาสเรคคอร์ดเซต
m_pDatabase	เก็บค่าพอยน์เตอร์ชี้ไปยังออบเจ็กต์ CDaoDatabase
m_strFilter	เก็บสตริงที่ถูกใช้เพื่อ WHERE ซึ่งเป็น SQL สเตทमेंต์

ในการติดต่อกับฐานข้อมูลของ DAO นั้นเราสามารถที่จะกำหนดคุณสมบัติของ Recordset ได้ โดยมี 3 คุณสมบัติ คือ Snapshot, Dynaset และ Table ทั้งนี้จะกล่าวถึงรายละเอียดของคุณสมบัติที่เราใช้ในการติดต่อกับฐานข้อมูลเป็นหลัก คือ Snapshot และ Dynaset

Dynaset

ไดนาเซต คือเรคคอร์ดเซตที่มีคุณสมบัติไดนามิก ถ้าเราเซตเรคคอร์ดเซตให้เป็นไดนาเซตแล้ว ในช่วงระหว่างที่โปรแกรมทำงาน ออบเจกต์เรคคอร์ดเซตจะทำการซิงโครไนซ์กับฐานข้อมูล ทำให้เราทราบการเปลี่ยนแปลงของข้อมูลทันทีเมื่อมีการเปลี่ยนแปลงเกิดขึ้น เพราะในแวลูอิมที่ผู้ใช้หลายๆ คน ผู้ใช้บางคนอาจแก้ไข ลบ หรือเพิ่มเรคคอร์ดได้

ลักษณะของไดนาเซตมีดังนี้

1. เมื่อเราเพิ่มหรือลบเรคคอร์ดในแอปพลิเคชัน จะมีผลกับไดนาเซตของเราทันที ซึ่งก็คือเราจะเห็นเรคคอร์ดเพิ่มขึ้นใหม่ทันที
2. เรคคอร์ดที่ผู้ใช้คนอื่นเพิ่มในตารางจะ ไม่มีผลต่อไดนาเซตของเรา ซึ่งก็คือ เราจะไม่เห็นเรคคอร์ดเหล่านั้นจนกว่าจะเรียกฟังก์ชัน Query เพื่อคิวรีข้อมูลใหม่อีกครั้ง
3. เมื่อผู้ใช้ได้ลบเรคคอร์ดแล้ว ใค้ค MFC จะข้ามเรคคอร์ดเซตที่ถูกลบไป
4. การแก้ไขเปลี่ยนแปลงข้อมูล โดยผู้ใช้คนอื่นกับเรคคอร์ดที่ปรากฏอยู่ จะมีผลกับไดนาเซตของเราทันที เมื่อเราเลื่อนเคอร์เซอร์ไปยังเรคคอร์ดที่แก้ไขข้อมูลแล้ว

Snapshot

Snapshot เป็นเรคคอร์ดเซตที่มีผลต่อวิวดิติต ซึ่งข้อมูลบนวิวจะอยู่ชั่วขณะหนึ่งในระหว่างที่ Snapshot ถูกสร้าง สำหรับลักษณะการทำงานของ Snapshot มีดังนี้

1. ถ้าเราเปิด Snapshot และเลื่อนเรคคอร์ดเซตทั้งหมด เซตของเรคคอร์ดที่มีข้อมูลอยู่จะไม่มี การเปลี่ยนแปลงใดๆ จนกว่าจะสร้าง Snapshot อีกครั้งด้วยการเปิดและปิดเรคคอร์ดเซต หรือเรียกใช้ ฟังก์ชัน Query อีกครั้ง
2. เราสามารถสร้าง Snapshot ที่สามารถอัปเดตข้อมูลหรืออ่านอย่างเดียว Snapshot ไม่เหมือนกับ ไดนาเซตดังนี้ เมื่อเราใช้ Snapshot ที่อัปเดตได้ ผู้ใช้คนอื่นๆ แก้ไขข้อมูลในเรคคอร์ดเซต ข้อมูลบน หน้าจอเรายังคงเหมือนเดิม แต่ข้อมูลบนหน้าจอเราจะเปลี่ยน ถ้าเราทำการแก้ไข หรือลบเรคคอร์ดด้วย โปรแกรมของเรา
3. สำหรับการเพิ่มเรคคอร์ดใน Snapshot แล้ว เราจะ ไม่เห็นจนกว่าจะเรียกใช้ฟังก์ชัน Query อีกครั้ง Snapshot จึงมีประโยชน์มากเมื่อเราต้องการให้ข้อมูลบนหน้าจอคงอยู่ระหว่างที่ปฏิบัติงานอยู่

2.2.1 ฟังก์ชันสำคัญที่ใช้ในการใช้ DAO ติดต่อฐานข้อมูล

ฟังก์ชันไดอะล็อกเดต้าเอ็กซ์เชนจ์ (Dialog Data Exchange: DDX)

ฟังก์ชัน DDX อยู่ในคลาสวิวที่สืบทอดจากคลาส CDaoRecordView ภายในแอปพลิเคชันที่ สร้างจาก MFC ฟังก์ชัน DDX นี้ให้เราสามารถ โอนย้ายข้อมูลระหว่างคอนโทรลไดอะล็อกหรือฟอร์มวิว

กับเคต้าเมมเบอร์ในคลาสเดียวกัน โดย Foreign Object สามารถแลกเปลี่ยนข้อมูลระหว่างคอนโทรลกับเคต้าเมมเบอร์ของออบเจกต์ที่แยกออกมา

เมื่อเราระบุคลาส CDaoRecordset หรือ CRecordset ที่สืบทอดมาร่วมกับ CDaoRecordView หรือ CRecordView เราสามารถตั้งชื่อคลาสเรคคอร์ดเซตที่มีอยู่แล้วหรือสร้างใหม่ได้ คลาสวิซาร์ดได้เพิ่มตัวแปรนี้เป็นคลาสเรคคอร์ดเซต เราสามารถแสดงความสัมพันธ์ระหว่างเรคคอร์ดบนหน้าจอออบเจกต์เรคคอร์ดวิว และกับออบเจกต์เรคคอร์ดเซตได้

AppWizard สร้างเรคคอร์ดวิว และเรคคอร์ดเซต ซึ่งเรคคอร์ดเซตคือ Foreign Object จากนั้น AppWizard จะแนบคอนโทรลเรคคอร์ดมายังเคต้าเมมเบอร์เรคคอร์ดเซตด้วย m_pSet เมื่อมีการเรียกฟังก์ชัน DoFieldExchange มันก็จะแลกเปลี่ยนข้อมูลกันระหว่างฟิลด์เคต้าเมมเบอร์ในเรคคอร์ดเซต กับคอลัมน์ของฐานข้อมูล

ฟังก์ชัน DAO Record Field Exchange: DFX และ Record Field Exchange: RFX

เมื่อมีการเรียกฟังก์ชัน DoFieldExchange ฟังก์ชันทั้ง DFX และ RFX จะทำหน้าที่แลกเปลี่ยนข้อมูลกันระหว่างฟิลด์ของออบเจกต์เรคคอร์ดเซตกับคอลัมน์ของฐานข้อมูล ซึ่งฟังก์ชัน DFX นี้จะใช้กับคลาสเรคคอร์ดเซต DAO และฟังก์ชัน RFX จะใช้กับคลาสเรคคอร์ดเซตที่เป็น ODBC

2.3 Augmented Translation Network สำหรับโครงสร้างภาษาไทย

Augmented Translation Network (ATN) นั้นเป็นกระบวนการกระจายประโยค หรือที่เรียกว่า top-down parsing นั่นเอง ลักษณะของ ATN นั้นจะเหมือนกับ finite state machine ซึ่งใช้แสดงการขั้นตอนการเปลี่ยนสถานะ หลังที่มีการตรวจสอบเงื่อนไขของประโยคนั้นๆแล้ว โดยที่ ATN นั้นจะทำการตรวจสอบประโยคที่มีลักษณะดังต่อไปนี้ คือ

1. จะต้องมีคำที่เป็นคำเฉพาะ เช่น กิน ไป เป็นต้น
2. สามารถแยกชนิดของคำได้
3. สามารถที่จะนำไปเข้า Network อื่นได้เช่น ต้องการตรวจสอบ NP ก็สามารถที่จะเอาเข้าไปใน Network ที่ใช้ตรวจสอบของ NP ได้
4. กระบวนการที่แสดงการตรวจสอบทั้งที่เป็น อินพุท และที่เป็นองค์ประกอบของประโยคนั้น ต้องเป็นอันเดียวกัน

5. กระบวนการที่สร้างโครงสร้างจะจัดรูปแบบให้เป็นไปตามไวยากรณ์

การกำหนด ATN ให้กับภาษาไทยนั้น อันดับแรกเราต้องเข้าใจโครงสร้างของภาษาไทยก่อนว่าเป็นเช่นไร เราได้ศึกษามาแล้วว่าประโยคภาษาไทยสามารถแบ่งออก เป็น 5 ลักษณะด้วยกัน คือ

1. ประโยคที่มีประธานเป็นผู้กระทำ ลักษณะประโยคคือ
ประธาน + กริยา + กรรม

2. ประโยคที่มีกรรมอยู่หน้าประโยค ลักษณะประโยคคือ

กรรม + กริยา + ประธาน

3. ประโยคที่เรากล่าวกริยาขึ้นก่อนประธาน กริยาพวกนี้มักมี 3 คำ คือ เกิด มี ปรากฏ
ลักษณะประโยคคือ

กริยา + วลี

4. ประโยคที่ผู้รับใช้เพิ่มขึ้น ลักษณะประโยคคือ

ประธาน + ให้ + กรรม + กริยา

5. ประโยคที่นำเอาคำกริยามาทำหน้าที่อย่างคำนาม โดยวางไว้หน้าประโยค ซึ่งการวางประโยคจะเหมือนกันกับลักษณะของประโยคในข้อ 1 แต่ประธานจะเป็นกริยาที่ทำหน้าที่อย่าง นามนั่นเอง

ลักษณะของการสร้าง ATN ของภาษาไทยนั้นเราจะอาศัยโครงสร้างหลักในการสร้าง คือเราจะเรียงประโยคในลักษณะ ที่เป็น ประธาน + กริยา + กรรม โดยประธาน กริยา และ กรรม จะมีลักษณะเป็นวลี (Noun Phrase) ในที่นี้จะใช้ตัวย่อภาษาอังกฤษเพื่อความสะดวก โดยมีเงื่อนไขคือ

NP แทน Noun Phrase

VP แทน Verb Phrase

Det แทน คำที่นำหน้านามเช่น คุณ นาย นาง นางสาว

PP แทนบุรพบท

V แทน กริยา

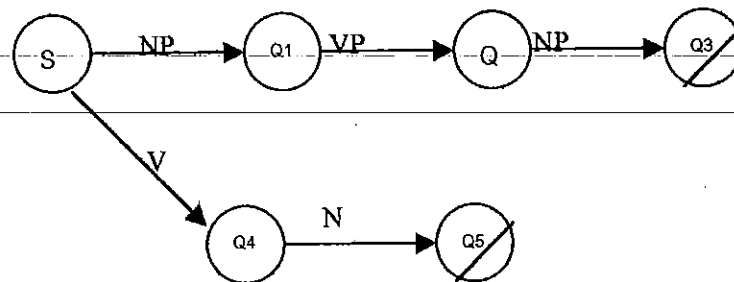
Aux. แทน คำช่วย เช่น ทรง การ

QN แทน ลักษณะนาม เช่น ค้าง เรือ น อัน

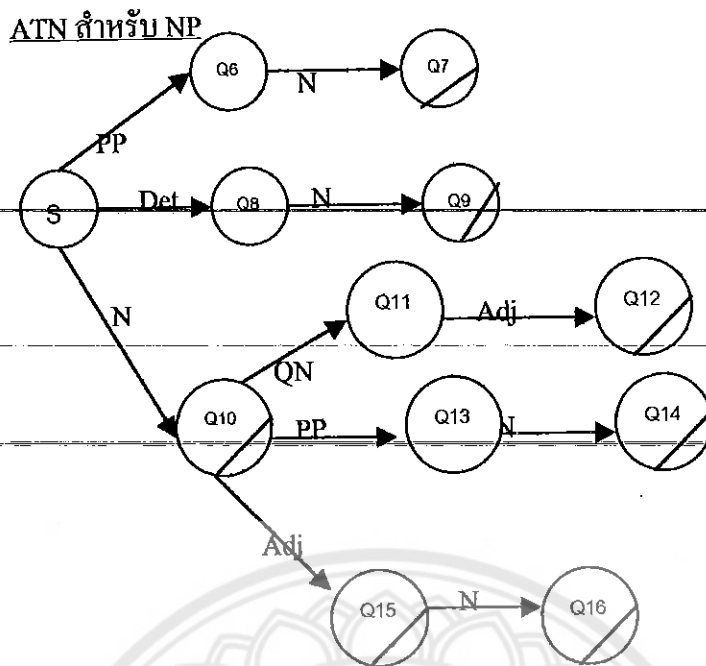
Adj. แทน คุณศัพท์ เช่น สูง ใหญ่

เราสามารถเขียน ATN ได้ดังนี้

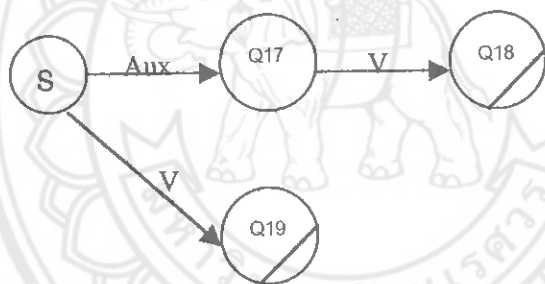
ATN สำหรับ ประโยคภาษาไทย



รูปที่ 2.1 ATN สำหรับ โครงสร้างของประโยคภาษาไทย



ATN สำหรับ VP



รูปที่ 2.1 (ต่อ) โครงสร้าง ATN สำหรับประโยคภาษาไทย

ตัวอย่างประโยคเช่น

สุนัขสีดำกินหนุตัวโต

1. เริ่มต้นที่ State S ของกระบวนการตรวจสอบ NP
2. เข้าสู่ NP
3. ตรวจสอบว่า สุนัขเป็นคำนามใช่หรือไม่
4. ไปที่ State Q10 เพราะการตรวจสอบถูกต้องพบว่าสุนัขเป็นคำนาม
5. ตรวจสอบว่า สีดำเป็น คุณศัพท์ใช่หรือไม่
6. ตรวจสอบแล้วพบว่า สีดำเป็น คุณศัพท์ ไปที่ State Q15
7. ตรวจสอบ คำว่า กินว่าเป็น คำนามหรือไม่
8. กินไม่ใช่คำนาม การทดสอบจึงไม่ได้ผล จึงต้องมีการย้อนกลับ

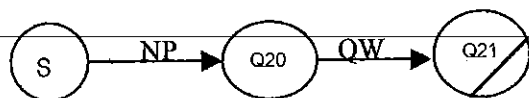
9. ไม่มีคำที่ต้องตรวจสอบแล้ว และที่ State Q15 สามารถครบโครงสร้างได้เพราะอยู่ที่ Accepting State เราจึง มีการย้ายไปที่ Q1 แล้วก็ตรวจสอบคำว่ากิน ว่าเป็น กริยา หรือเปล่า
10. กระบวนการมาที่การตรวจสอบ VP เริ่มที่ State S ของ VP
11. คำว่า กินเป็นคำกริยาการตรวจสอบถูกต้อง จึงย้ายไปที่ State Q19
12. ตรวจสอบคำว่า หนูว่าเป็นคำกริยาหรือไม่
13. การตรวจสอบผิด หนูไม่ใช่ คำกริยา จึงหยุดอยู่ที่ State Q19 ครบ โครงสร้างเพราะ อยู่ที่ Accepting State ของ VP แล้ว ไปที่ State Q2
14. จากนั้นตรวจสอบว่า หนูเป็นคำนามใช่หรือไม่
15. ไปที่ State Q10 เพราะการตรวจสอบถูกต้องพบว่าหนูเป็นคำนาม
16. ตรวจสอบว่า สีขาวเป็น คุณศัพท์ใช่หรือไม่
17. ตรวจสอบแล้วพบว่า สีขาวเป็น คุณศัพท์ ไปที่ State Q15
18. หลังคำว่า สีขาว ไม่มี อินพุตใดๆแล้วและ State Q15 เป็น Accepting Stateจากนั้น กระโดดไปที่ State Q3 ได้ เป็นอันว่า ได้เอาที่พุดออกมาเป็นประโยคที่ต้องการ แล้ว

2.4 ATN สำหรับประโยคคำถาม

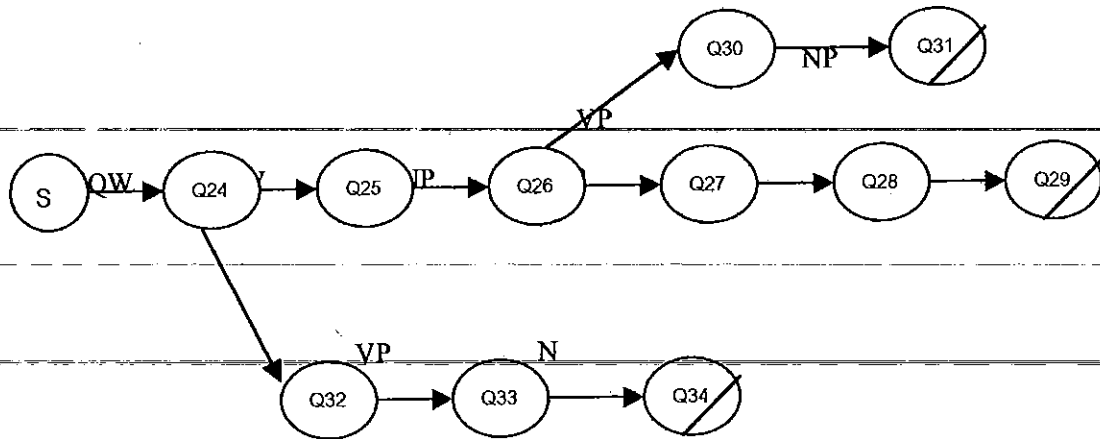
ก่อนอื่นเราต้องมาทราบโครงสร้าง ของประโยคคำถามก่อน ซึ่งในการสังเกตประโยคคำถาม สำหรับภาษาไทยนั้นเราจะดูได้จาก Question Words ซึ่งแบ่งได้เป็น 2 แบบ คือ

1. ปดจฉาสรพนาม ซึ่งเป็นสรรพนามที่ใช้แทนนาม มีเนื้อความเป็นคำถาม ได้แก่คำว่า อะไร ไหน ผู้ใด ชาวอะไร ชาวไหน ฯลฯ
2. ปดจฉาวิเศษณ์ คือ วิเศษณ์ ที่ประกอบคำอื่น เพื่อบอกคำถาม เป็นคำลักษณะเดียวกันกับ ปดจฉาสรพนาม แต่ใช้ต่างกัน ได้แก่คำว่า ไฉน อย่างไร ทำไม ทำไร ไฉน อย่างไร ฯลฯ

ATN สำหรับ ประโยคคำถาม



รูปที่ 2.2 โครงสร้าง ATN สำหรับโครงสร้างประโยคคำถาม



รูปที่ 2.2(ต่อ) โครงสร้าง ATN สำหรับโครงสร้างประโยคคำถาม

ตัวอย่างของการตรวจสอบประโยค

เราจะยกตัวอย่างประโยคดังนี้คือ

ช้างของพระนเรศวรชื่ออะไร

สำหรับประโยคแรกเราสามารถตรวจสอบได้ดังนี้

1. เริ่มที่ S State มีการอ่าน อินพุท แรก คือ คำว่า "ช้าง"
2. เข้าสู่ NP
3. ตรวจสอบว่า "ช้าง" เป็น คำนาม ใช่หรือไม่
4. ตรวจสอบแล้ว "ช้าง" เป็นคำนาม ไปที่ State Q10
5. อ่านอินพุท คำว่า "ของ" ตรวจสอบว่าเป็นคำประเภทใด
6. คำว่า "ของ" เป็นคำบุรพบท ไปที่ State Q13
7. ตรวจสอบคำว่า "พระนเรศวร" ว่าเป็นคำประเภทใด โดย เริ่มที่มีการอ่านอินพุท "พระ"
8. คำว่า พระ เป็นบุรพบท ใช่หรือไม่
9. ตรวจสอบแล้ว "พระ" คำว่า "พระ" เป็น Det ไปที่ State Q8
10. อ่านอินพุท "นเรศวร" ว่าเป็นคำนาม ใช่หรือไม่
11. "นเรศวร"เป็นคำนาม ไปที่ State Q9
12. อ่านอินพุท ชื่อ ตรวจสอบว่าเป็นคำนาม ใช่หรือไม่
13. ตรวจสอบแล้ว "ชื่อ" เป็นคำนามเพราะฉะนั้นยังอยู่ที่ State Q9
14. อ่านอินพุท คำว่า "อะไร" ตรวจสอบว่าเป็นคำนาม ใช่หรือไม่
15. อะไร ไม่ใช่คำนามแต่เป็น QW
16. ไปที่ State Q21

17. ไม่มีอินพุท เข้ามา และอยู่ที่ Accepting State
18. สามารถประกอบเป็นประโยค ได้สำเร็จ

2.5 การตัดคำ

ปัจจุบันได้มีการนำคอมพิวเตอร์เข้ามาใช้งานด้านต่างๆ อย่างแพร่หลายไม่ว่าจะเป็นทางด้าน การคำนวณ ด้านกราฟฟิก การจัดเก็บฐานข้อมูล รวมกรนำไปใช้งานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing) การประมวลผลภาษาธรรมชาติคือ กระบวนการที่ทำให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ได้ ตัวอย่างเช่น การแปลภาษาไทย(Thai-English-Machine-Translation) การสังเคราะห์เสียงภาษาไทย(Thai speech Synthesis) เป็นต้น สำหรับภาษาที่ไม่มีการเว้นวรรคระหว่างคำ เช่น ภาษาจีน ภาษาญี่ปุ่น และภาษาอื่นๆ รวมทั้งภาษาไทยด้วย การหาขอบเขตของคำหรือการตัดคำเป็นสิ่งที่ต้องทำเป็นอันดับแรกสำหรับงานด้านการประมวลผลภาษาธรรมชาติ และประสิทธิภาพการตัดคำ ก็จะส่งผลถึงความถูกต้องของการประมวลผลภาษาธรรมชาติในระบบงานต่างๆ ด้วย ดังนั้นจะเห็นได้ว่าการตัดคำเป็นสิ่งที่สำคัญอย่างยิ่ง ในงานด้านการประมวลผลภาษาธรรมชาติ

2.6 วิธีการตัดคำ

เนื่องจากการตัดคำได้มีการพัฒนาติดต่อกันมาเป็นเวลายาวนาน ทำให้มีงานวิจัยด้านการตัดคำเกิดขึ้นมากมายหลายวิธี ซึ่งช่วงแรกได้มีการพัฒนาตัดพยางค์ขึ้นมาก่อน หลังจากนั้นได้มีการพัฒนาการตัดคำต่อตามมา ซึ่งวิวัฒนาการการตัดคำได้แบ่งตามลักษณะฐานข้อมูลที่น่ามาใช้ในการตัดคำ แบ่งได้ 3 ยุค คือ 1. ยุคการใช้กฎ 2. ยุคการใช้พจนานุกรม 3. ยุคการใช้คลังข้อความ

1. ยุคการใช้กฎ

ในยุคนี้คอมพิวเตอร์ยังไม่มีความสามารถในการประมวลผลสูงมากนัก และหน่วยความจำมีขนาดเล็ก จึงมีการพัฒนาการตัดพยางค์ขึ้นมาก่อน เนื่องจากพยางค์มีกฎเกณฑ์ที่แน่นอนมากกว่าคำ ยุคนี้จึงได้ใช้กฎในการตัดพยางค์ จากการนำกฎมาช่วยในการตัดพยางค์ทำให้สามารถแบ่งพยางค์ได้ถูกต้องจำนวนมาก โดยงานการแบ่งพยางค์ที่ผ่านมามีดังต่อไปนี้

- งานของ ยูพิน ไทยรัตนานนท์

งานของ ยูพิน ไทยรัตนานนท์ เป็นงานวิจัยในการตัดพยางค์ โดยการใช้กฎในการตัดพยางค์ โดยกฎที่สร้างขึ้นมา อาศัยหลักไวยากรณ์ภาษาไทย แต่มีปัญหาในการสร้างกฎ เพราะมีบางพยางค์ไม่เป็นไปตามกฎที่ตั้งไว้ ซึ่งงานวิจัยนี้ได้พัฒนาโดยใช้ภาษาพีแอล/I

ลักษณะของคำที่นำมาใช้ในการตัดพยางค์ ได้สร้างมาจากลักษณะไวยากรณ์ทางภาษาไทย โดยพิจารณาจากลักษณะของอักขระที่ปรากฏในพยางค์หรือคำ ซึ่งทำให้มีการการจัดหมวดหมู่ตัวอักขระภาษาไทย โดยแบ่งหมวดตามการนำไปใช้ สามารถแบ่งได้เป็น 5 กลุ่ม ดังต่อไปนี้

1. กลุ่มพยัญชนะ (Consonant)

- พยัญชนะที่อยู่หน้าพยางค์เสมอ
- พยัญชนะที่ส่วนใหญ่จะอยู่หน้าพยางค์
- พยัญชนะที่เป็นตัวสะกด
- พยัญชนะที่เป็นสระ
- อื่นๆ

2. กลุ่มสระ (Vowel)

- สระที่ไม่ต้องมีตัวสะกด
- สระที่จะอยู่หน้าพยางค์เสมอ
- สระที่ส่วนใหญ่จะมีตัวสะกดร่วมด้วย
- สระที่มีหรือ ไม่มีตัวสะกดร่วมด้วย

3. กลุ่มวรรณยุกต์ (Tone mark)

4. กลุ่มตัวเลข (Numeral)

5. กลุ่มอักขระพิเศษ (Special character)

ขั้นตอนการทำงานของวิธีการนี้จะตัดพยางค์จากขวามาช่วย โดยใช้กฎต่างๆ ที่สร้างมาจากลักษณะของตัวอักขระที่กล่าวไป และกฎต่างๆ ที่สร้างขึ้นมาจะจัดเก็บไว้ภายในรหัสต้นฉบับ (Source Code) ซึ่งทำให้การเพิ่มหรือแก้ไขกฎไม่สามารถทำได้สะดวก และผลความถูกต้องได้ไม่น้อยกว่า 85%

● งานของสุรินทร์ จรรยาพรพงษ์

สุรินทร์ จรรยาพรพงษ์ ได้ทำการวิจัยเกี่ยวกับการตัดคำภาษาไทยโดยใช้พยางค์ โดยกฎที่นำมาใช้นั้นได้มาจากหลักไวยากรณ์ภาษาไทย และได้ทำการวิเคราะห์ลักษณะต่างๆ ของพยางค์ภาษาไทย โดยลักษณะของกฎที่ได้นี้สามารถแบ่งได้เป็น 2 ชนิด คือ กฎการหาขอบเขตหน้า (Front boundary recognition rule) และ กฎการหาขอบเขตหลัง (Tail boundary recognition rule) และแต่ละกฎยังแบ่งออกเป็น 2 กลุ่มย่อยๆ คือแบ่งตามคุณลักษณะของตัวอักษรโดยกฎที่ได้เอามาจะจัดให้อยู่ในกลุ่มเอ (Group A) และแบ่งตามคุณสมบัติของรูปแบบการใช้สระแต่ละตัวซึ่งกฎที่ได้เอามาจะแบ่งให้อยู่ในกลุ่มบี (Group B)

2. ยุคการใช้พจนานุกรม

ในยุคนี้ถือได้ว่าเป็นยุคเริ่มแรกในการตัดคำ เนื่องจากยุคนี้เครื่องคอมพิวเตอร์ได้มีการพัฒนาขึ้นและมีหน่วยความจำมากขึ้น จากยุคที่แล้วมีการนำเอากฎเข้ามาช่วยในการแบ่งพยางค์ แต่สำหรับการแบ่ง

คำแล้วการใช้กฎอย่างเดียวไม่สามารถที่จะหาขอบเขตของคำได้ ทำให้ในขณะนี้ได้มีการคิดค้นหาวิธีการแบ่งคำโดยมีการนำเอาพจนานุกรมเข้ามาร่วมใช้กับกฎในการตัดคำด้วย โดยแนวคิดการตัดคำใช้พจนานุกรมแบบต่างๆ มีดังนี้

- ยีน ภู่วรรณ และวิวรรณ อิมอรณณ์

ในงานวิจัยนี้จะเป็นงานวิจัยการแบ่งพยางค์ด้วย ซึ่งถือได้ว่าเป็นงานวิจัยงานแรกของการตัดพยางค์ที่มีการนำพจนานุกรมเข้ามาใช้ โดยจะจัดเก็บพยางค์ต่างๆในพจนานุกรม และมีนำกฎไวยากรณ์ต่างๆ จำนวน 18 กฎ เข้ามาช่วยในกรณีที่ไม่มีพยางค์ในพจนานุกรม

หลักการทำงานของกระบวนการวิธีการตัดพยางค์ด้วยพจนานุกรมนี้อีกคือ จะทำการตรวจสอบสายอักขระ (String) ที่เข้ามาจากซ้ายไปขวากับพยางค์ที่ได้เก็บไว้ในพจนานุกรม ในกรณีที่ทำการตรวจสอบแล้วปรากฏว่าพบพยางค์มากกว่า 1 พยางค์ในพจนานุกรม ก็ให้ทำการเลือกพยางค์โดยเลือกพยางค์ที่ยาวที่สุด แล้วทำไปเรื่อยๆ จนจบสายอักขระ แต่ถ้าในกรณีที่เลือกยาวที่สุดไปแล้ว ทำให้เกิดพยางค์ที่ไม่ปรากฏในพจนานุกรมก็ยอมให้มีการย้อนรอย (Back Tracking) กลับไปเลือกพยางค์ที่ยาวรองลงมา แทน ซึ่งวิธีการนี้จะเป็นที่รู้จักกันในชื่อ การตัดคำ(พยางค์)แบบเลือกคำ(พยางค์)ที่ยาวที่สุด(Longest Matching)

จากงานวิจัยนี้ได้มีการเปรียบเทียบความเร็วในการแบ่งพยางค์ ซึ่งสรุปผลได้ว่าเมื่อนำพจนานุกรมเข้ามาใช้ในการแบ่งพยางค์จะสามารถตัดพยางค์ได้รวดเร็วกว่าการใช้กฎ โดยที่ถูกต้องของการตัดพยางค์นั้นสามารถตัดได้ถูกต้องมากกว่า 99 % แต่วิธีนี้ยังมีข้อเสียคือต้องเสียเนื้อที่ในการจัดเก็บพจนานุกรมในหน่วยความจำหลักประมาณ 50 กิโลไบต์

- งานของดวงแก้ว สวามิภักดิ์

งานวิจัยชิ้นนี้คือ การสร้างซอฟต์แวร์วิเคราะห์ไวยากรณ์ไทยภายใต้ระบบยูนิคซ์ เป็นงานวิจัยด้านการตัดคำภาษาไทย โดยใช้ไวยากรณ์ที่สร้างขึ้นเอง และมีการนำพจนานุกรมเข้ามาใช้ประกอบรวมด้วย โดยสาเหตุที่นำทั้งกฎไวยากรณ์และพจนานุกรมเข้ามาช่วยในการตัดคำนั้นก็เพื่อที่จะแก้ไขปัญหาการตัดคำโดยใช้พจนานุกรมเพียงอย่างเดียว ซึ่งไม่สามารถตัดคำได้ถูกต้องในกรณีที่คำนั้นไม่มีอยู่ในพจนานุกรม

งานวิจัยการตัดคำนี้ได้ทำภายใต้ระบบปฏิบัติการยูนิคซ์ และได้มีการโปรแกรมเล็กซ์ (Lex) เข้ามาช่วยในการตัดคำ โดยมีการสร้างกฎต่างๆ ให้อยู่ในรูปแบบนิพจน์ที่มีกฎเกณฑ์ (Regular Expression) โดยกฎที่สร้างขึ้นมาประกอบไปด้วย 43 กฎ ซึ่งกฎที่ได้มานี้จะไม่มีการรวมตัวสะกดเข้าไปในกฎด้วยยกเว้นบางกรณี เนื่องจากลักษณะโปรแกรมเล็กซ์ จะพยายามสร้างกลุ่มตัวตัวอักษร (Token) ที่มีขนาดที่ยาวที่สุดก่อน ดังนั้นถ้ามีการนำการนำกฎที่มีตัวสะกดเข้ามาใช้ จะเป็นสาเหตุให้มีการรวมเอาตัวอักษรตัวหน้าของคำถัดไปมาเป็นตัวสะกดได้ ซึ่งเมื่อได้ผ่านการวิเคราะห์ด้วยกฎแล้ว ขั้นตอนต่อไปก็จะมี

รวมกลุ่มตัวอักษรเข้าด้วยกัน โดยทำการตรวจสอบจากพจนานุกรม ส่วนโครงสร้างของพจนานุกรมที่นำมาใช้ในที่นี้คือฐานข้อมูล (Relational DBMS) ซึ่งใช้ค่าเป็นครรชนี(Index) และไฟล์ครรชนีได้พัฒนาขึ้นโดยใช้โครงสร้างข้อมูลแบบบีทรี(B-Tree)

- สัมพันธ์ ะรึนรมณ์

ในงานวิจัยนี้เป็นงานวิจัยการแบ่งคำด้วยพจนานุกรม โดยเป้าหมายของงานวิจัยนี้จะเน้นเรื่องประสิทธิภาพในด้านความเร็วของขั้นตอนวิธีในการตัดคำ และลดขนาดของพจนานุกรม เนื่องจากเมื่อนำพจนานุกรมเข้าใช้ในการตัดคำแล้วจะทำให้ความถูกต้องในการตัดคำเพิ่มขึ้นมากกว่าการใช้กฎ ดังนั้นงานวิจัยนี้จึงไม่เน้นการเพิ่มประสิทธิภาพในด้านความถูกต้องมากนัก งานวิจัยนี้ยังได้ทำการจัดเก็บคำลงในพจนานุกรมแทนพยางค์ ขั้นตอนการทำงานยังใช้วิธีการเลือกคำที่ยาวที่สุดเหมือนเดิม ตัวอย่างการตัดคำโดยเลือกคำที่ยาวที่สุดดังตาราง

ตารางที่ 2.1 ตัดคำแบบเลือกคำที่ยาวที่สุด

ประโยค	คำที่ได้	คำที่ถูกเลือก
โคนมนอนบนกองหญ้า	โค, โคน	โคน
มนอนบนกองหญ้า	-	(ย้อนรอย)
โคนมนอนบนกองหญ้า	โค, โคน	โค(เลือกคำรองลงมา)
นมนอนบนกองหญ้า	นม	นม
นอนบนกองหญ้า	นอน	นอน
บนกองหญ้า	บน	บน
กองหญ้า	กอง	กอง
หญ้า	หญ้า	หญ้า

จากตาราง จะแสดงการตัดคำแบบเลือกคำที่ยาวที่สุด โดยประโยคที่นำมาตัดคือ “โคนมนอนบนหญ้า” สามารถตัดคำได้เป็น “โค นม นอน บน หญ้า” ส่วนโครงสร้างของพจนานุกรมที่ได้นำมาใช้ในงานวิจัยนี้คือ โครงสร้างข้อมูลแบบทรี(Trie) ซึ่งการนำโครงสร้างทรีเข้ามาใช้สามารถลดขนาดของพจนานุกรมได้

- วิรัช ศรีเลิศล้ำวานิช

สำหรับงานวิจัยชิ้นนี้ ได้มีการพัฒนาการตัดคำโดยเรียกว่า “การตัดคำโดยเลือกแบบเหมือนมากที่สุด (Maximal Matching)” ซึ่งขั้นตอนวิธีนี้จะสามารถแก้ไขความบกพร่องของการตัดคำแบบเลือกคำยาว

ที่สุดได้ โดยจุดบกพร่องที่กล่าวนี้คือขั้นตอนวิธีการตัดคำแบบเลือกคำยาวที่สุดจะเลือกคำที่ยาวเกินไป ตั้งแต่ครั้งแรก ทำให้ข้อความที่ตามมาเกิดข้อผิดพลาดได้ ตัวอย่างเช่น ประโยค “ไปห้ามเหลื” จะตัดได้ เป็น “ไป หาม เหลื” โดยถูกต้องเป็น “ไป หาม เหลื”

หลักการของการตัดคำโดยเลือกแบบเหมือนมากที่สุดคือ ขั้นตอนแรกคือจะทำการคำที่เป็นไปได้ ทุกๆ แบบก่อน แล้วหลังจากนั้นก่อนให้ประโยคที่มีจำนวนคำน้อยที่สุด ตัวอย่างเช่น “ไปห้ามเหลื” สามารถตัดได้เป็น “ไป หาม เหลื” กับ “ไป หามเหลื” ซึ่งเมื่อพิจารณาจำนวนคำแล้ว วิธีการนี้จะเลือก ประโยค “ไป หาม เหลื” ซึ่งเป็นประโยคที่ถูกตัด สำหรับกรณีที่ได้อาจมีจำนวนคำเท่ากันก็ให้นำการตัดคำ แบบเลือกคำยาวที่สุดเข้ามาช่วยพิจารณา ตัวอย่างเช่น “ฉั น ั น ั ง ตาก ลม” สามารถตัดได้ 2 แบบคือ “ฉั น ั น ั ง ตาก ลม” และ “ฉั น ั น ั ง ตา ลม” ซึ่งจะมีจำนวนคำเท่ากัน 2 ประโยค แต่เมื่อใช้การตัดคำแบบ เลือกคำยาวที่สุดเข้ามาพิจารณา ประโยคที่ได้คือ “ฉั น ั น ั ง ตาก ลม”

สรุปวิธีการนี้สามารถช่วยแก้ไขข้อบกพร่องของการตัดคำแบบเลือกคำที่ยาวที่สุดได้เพราะว่าการ เลือกคำที่ยาวที่สุดเมื่อเจอข้อความที่กำวมกวม โดยไม่มีการพิจารณาถึงข้อความถัดไป ซึ่งลักษณะ เหมือนการใช้ขั้นตอนวิธีแบบโลภ (Greedy Algorithm) ที่พิจารณาเฉพาะบริเวณใกล้ๆ เท่านั้น แต่วิธีการ ตัดคำโดยเลือกแบบเหมือนมากที่สุดจะเป็นการใช้ขั้นตอนวิธีแบบโลภโดยพิจารณาข้อความทั้งหมด แทน แต่อย่างไรก็ตามเนื่องจากวิธีการนี้ใช้เฉพาะพจนานุกรมในการตัดคำเท่านั้น ดังนั้นการตัดคำนี้ยัง ไม่สามารถที่จะตัดคำได้ถูกต้องทั้งหมด

2. ยุคการใช้คลังข้อความ

จากการพัฒนาการตัดคำในยุคที่ผ่านมาเราใช้เพียงกฎ หรือพจนานุกรมในการแบ่งคำเท่านั้น ทำให้การตัดคำในยุคก่อนไม่สามารถที่จะตัดคำได้ถูกต้องทั้งหมด และในยุคนี้เครื่องคอมพิวเตอร์มีประ สติธิภาพมากยิ่งขึ้น มีหน่วยความจำมากขึ้นเป็นจำนวนมาก และได้มีการพัฒนาคลังข้อความ (Corpus) ขึ้นจำนวนมาก ทำให้ในยุคนี้ได้มีการตัดคำขึ้นมาใหม่ โดยนอกเหนือการใช้กฎ พจนานุกรมแล้วยังมีการ นำความรู้ต่างๆ จากคลังข้อความเข้ามาประยุกต์ใช้ด้วย ตัวอย่างความรู้ที่ได้จากคลังข้อความเช่น คำสถิติ การใช้คำภายในคลังข้อความและลักษณะไวยากรณ์ที่ใช้ในคลังข้อความ เป็นต้น

● อัศนีย์ ก่อตระกูลและคณะ

ในงานนี้ทำการวิจัยเรื่อง “A Statistical Approach to Thai Word Filtering” เนื่องจากปัญหาของ การวิเคราะห์หน่วยคำ (Morphological Analysis) สำหรับภาษาไทยนั้นจะมีปัญหาต่างๆ ดังนี้คือ ปัญหา ความกำกวม ปัญหาความกำกวมของการกำหนดหน้าที่คำ (Part-of-Speech Tagging Ambiguity) และ ปัญหาการสะกดคำผิด โดยปัญหาต่างๆ เหล่านี้จะทำให้เกิดผลลัพธ์ที่ตัดคำและกำหนดหน้าที่ของคำแล้ว ออกมาหลายๆ รูปแบบ ซึ่งงานวิจัยนี้จะทำการลดผลลัพธ์ที่ไม่เหมาะสมออกไป เพื่อที่สามารถจะทำให้ พาสเซอร์(Parser) ทำงานได้รวดเร็วขึ้น

ในงานวิจัยนี้จะนำเรื่องสถิติเข้ามาใช้แก้ปัญหาคัดคำและการกำหนดหน้าที่ของคำโดยมีการนำเรื่องโมเดลไตรแกรม เข้ามาช่วยในการแก้ปัญหาคัดคำ

- **สุรพันธ์ เมฆนาวินและคณะ**

จากการตัดคำที่ผ่านมาจะมีการนำเอาพจนานุกรมเข้ามาใช้ในการตัดคำเพียงอย่างเดียวเท่านั้น และการนำวิทยาการศึกษาลำบาก (Heuristics) ต่างๆ เข้ามาช่วยแก้ปัญหาคำกำกวมที่เกิดขึ้นนั้น ไม่สามารถที่จะแก้ปัญหาคำกำกวมได้ทั้งหมด ดังนั้นได้มีการพัฒนาการตัดคำ โดยมีการนำวิธีทางสถิติเข้ามาช่วยในการแก้ไขปัญหาคำกำกวม ซึ่งวิธีการทางสถิติที่นำมาใช้คือการใช้ค่าสถิติที่เกิดจากลำดับของหน้าที่คำ หรืออาจกล่าวได้ว่าเป็นการนำส่วนหนึ่งของไวยากรณ์ มาใช้ในการแก้ไขปัญหาคำกำกวมการตัดคำ โดยให้หน้าที่คำแบบไตรแกรมโมเดล (Surapant Meknavin, Paisarn Charoenpornsawat and Boonserm Kijirikul, 1997) คือการตัดคำโดยมีการนำเอาค่าสถิติ ซึ่งพิจารณาจากความต่อเนื่องของหน้าที่คำ ส่วนวิธีการเลือกแบบการตัดคำที่ดีที่สุดนั้นทำได้โดยหาประโยคที่มีความน่าจะเป็นมากที่สุด

- **อศินีย์ ก่อตระกูลและคณะ**

ในยุคนี้คอมพิวเตอร์มีความรวดเร็วในการประมวลผลมากขึ้น และมีหน่วยความจำขนาดใหญ่ ทำให้ปัญหาที่สำคัญของยุคนี้ไม่ใช่เรื่องความเร็ว หรือการประหยัดเนื้อที่หน่วยความจำ แต่ปัญหาที่สำคัญในยุคนี้คือความถูกต้องของการตัดคำ เพราะในระบบต่างๆ มีความต้องการที่จะได้การตัดคำที่มีประสิทธิภาพที่สุด ซึ่งในยุคนี้ได้มีการคิดค้นวิธีการที่จะนำมาใช้ในการแก้ปัญหาคำศัพท์ที่ไม่ปรากฏในพจนานุกรมขึ้น

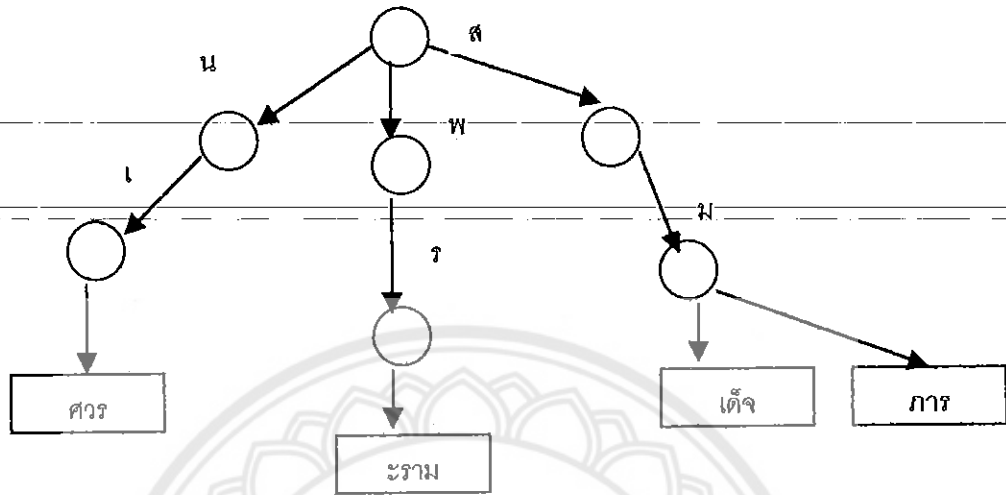
ในงานวิจัยได้ทำแก้ปัญหาเรื่องคำศัพท์ที่ไม่ปรากฏในพจนานุกรม โดยไม่ได้แค่หาขอบเขตของคำเท่านั้น แต่ยังสามารถที่จะบอกถึงหน้าที่คำและแสดงถึงลักษณะทางความหมาย (Semantic Attribute) และยังสามารถที่แก้ไขคำในกรณีที่เกิดการสะกดผิดด้วย ซึ่งในงานวิจัยนี้มีการนำวิธีการทางสถิติ (Statistical Model) และมีการนำกฎต่างๆ เข้ามาช่วยในการพิจารณาด้วย

2.7 โครงสร้างข้อมูลแบบทรี(Trie)

เนื่องจากโครงสร้างข้อมูลแบบทรีมีประสิทธิภาพในแง่ความเร็วในการทำงาน และยังใช้เนื้อที่หน่วยความจำในการจัดเก็บคำศัพท์ในพจนานุกรมน้อย

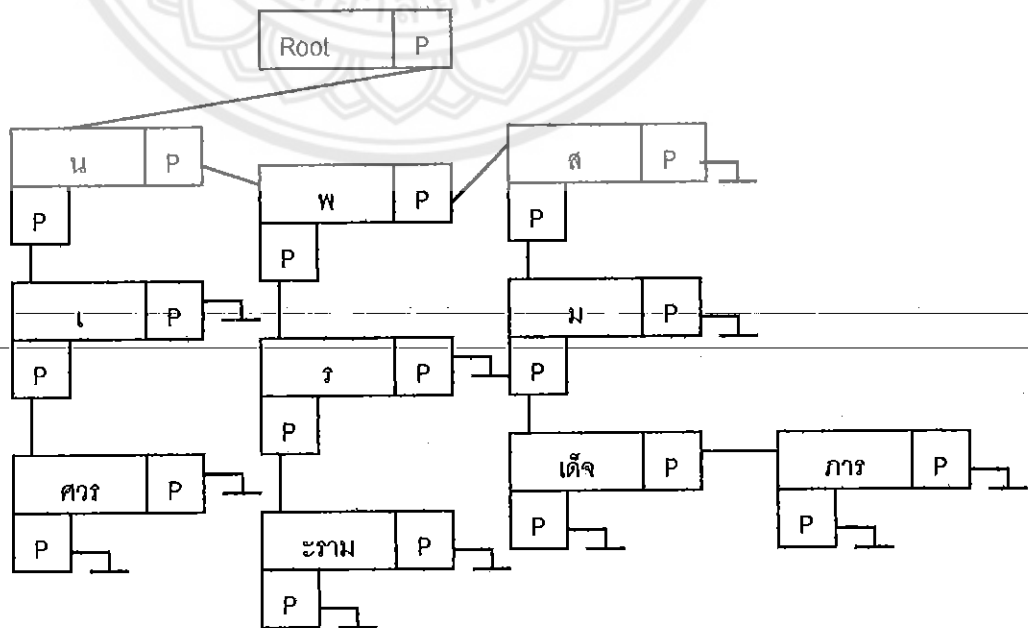
โครงสร้างข้อมูลแบบทรี (Corman, Leiserson and Rivest, 1990: Frankes and Baeza-Yates, 1992) จะมีลักษณะคล้ายกับโครงสร้างข้อมูลแบบต้นไม้ แต่ต่างกันตรงที่การเก็บข้อมูล เพราะโครงสร้างแบบทรีจะเก็บข้อมูลเป็นตัวอักษร แต่โครงสร้างแบบต้นไม้จะเก็บเป็นคำ โดยที่โครงสร้างข้อมูลแบบทรีจะกึ่ง(Edge)ในการจัดเก็บตัวอักษรและมีโหนดซึ่งในหลายๆ โหนดจะเก็บพ้อยเตอร์ที่ชี้ไปยังโหนดที่เก็บตัวอักษรตัวต่อไป

โครงสร้างข้อมูลแบบทรีจะมีการเก็บข้อมูลเป็นโหนด โดยที่แต่ละโหนดจะประกอบไปด้วยพ้อยเตอร์ที่ชี้ไปยังโหนดถัดไป ตามตัวอักษรที่เราเก็บ ลักษณะของการเก็บข้อมูลในโครงสร้างข้อมูลแบบทรีที่ใช้ ของคำว่า “นเรศวร พระราม สมเด็จ สมภาร” คือ



รูปที่ 2.3 การเก็บคำ “นเรศวร พระราม สมเด็จ สมภาร” ไว้ในโครงสร้างข้อมูลแบบทรี

โดยการเก็บที่ใช้ในโครงการครั้งนี้ได้มีการเก็บตัวอักษรตัวแรกไว้ 1 โหนด ตัวที่สองอีก 1 โหนด จากนั้นโหนดต่อไปเก็บตัวอักษรทั้งหมดที่เหลือ โดยโหนดแรกจะเป็นโหนดเริ่มต้นเรียกว่า โหนดพ่อแม่ (Parent) โดยที่โหนดลงมาเรียกว่า โหนดลูก (Children) ซึ่งสืบทอดมาจากโหนดบน ส่วนโหนดที่อยู่ในระดับเดียวกันเรียกว่า โหนดพี่น้อง (Siblings) ซึ่งสามารถเขียนในรูปแบบอาร์เรย์ได้ดังนี้



รูปที่ 2.4 โครงสร้างทรีในรูปแบบอาร์เรย์

ลักษณะการค้นหาคำจะหาตัวอักษรตัวแรกในโหนดที่อยู่ในระดับเดียวกัน จากนั้นจะหาตัวอักษรตัวที่สองในโหนดลูกของโหนดพ่อแม่ที่พบตัวอักษรตัวแรกหากไม่พบก็คือไม่มีค่านั้นในพจนานุกรม หากเมื่อพบคำที่สองแล้วจะค้นหากลุ่มอักษรที่ต้องการ หากพบแล้วสรุปว่าค่านั้นมีในพจนานุกรม

1. ประสิทธิภาพด้านความเร็ว

โครงสร้างข้อมูลที่เหมาะสมที่นำมาใช้ในการตัดคำต้องมีความรวดเร็วในการทำงานซึ่งเวลาที่ใช้ในการค้นหาคำศัพท์ในพจนานุกรมจะขึ้นกับจำนวนครั้งในการค้นหาคำศัพท์ทั้งหมด และเวลาที่ใช้ในการค้นหาแต่ละครั้ง ส่วนความเร็วที่ใช้ในโครงสร้างข้อมูลแบบทรีจะไม่ขึ้นกับจำนวนคำ แต่จะขึ้นกับความยาวของคีย์ที่ใช้ในการสืบค้น โดยที่โครงสร้างข้อมูลแบบอื่นจะขึ้นกับจำนวนคำที่เก็บในพจนานุกรมด้วย

2. ประสิทธิภาพในการใช้หน่วยความจำ

โครงสร้างข้อมูลแบบทรีนี้ยังมีการจองเนื้อที่เป็นจำนวนมาก ซึ่งเนื้อที่ส่วนใหญ่จะไม่ได้ถูกใช้งาน ดังนั้นได้มีผู้พัฒนา ทรีแบบใหม่ขึ้นมา ซึ่งช่วยลดเนื้อที่ในหน่วยความจำลง โดยโครงสร้างแบบใหม่มีชื่อว่า โครงสร้างทรีแบบสามแถวลำดับ (Triple-Array Trie) ซึ่งพัฒนาโดยจอห์นสัน โดยที่วิธีการนี้จะมีการใช้แถวลำดับ 3 ชุด ในการเก็บคำ ต่อมาได้มีการพัฒนาให้มีการใช้หน่วยความจำน้อยไปอีก วิธีการที่พัฒนาขึ้นมาใหม่นี้ได้พัฒนาโดยอะเอโอเอ ซึ่งมีชื่อว่า โครงสร้างทรีแบบแถวลำดับคู่ (Double-Array Trie) ซึ่งวิธีการนี้ใช้แถวลำดับเพียง 2 ชุดเท่านั้น

บทที่ 3

วิธีการดำเนินงาน

3.1 วิธีการศึกษาค้นคว้าวิจัย

การวิจัยที่ใช้ในโครงการนี้ จะใช้การวิจัยแบบค้นคว้าจากเอกสาร จากงานวิทยานิพนธ์ และค้นคว้ามาจากอินเทอร์เน็ต จะเน้นการค้นคว้าจากเอกสารเป็นหลัก ส่วนจากงานวิทยานิพนธ์และจากอินเทอร์เน็ตนั้น จะใช้เพื่อประกอบข้อมูลที่ได้จากกรค้นคว้าจากเอกสาร

3.2 ศึกษาลักษณะของข้อมูล การเลือกข้อมูลและเหตุผลในการเลือกข้อมูล

ลักษณะข้อมูลที่เราค้นคว้านั้นจะมีลักษณะเป็นข้อมูลที่เกี่ยวข้องกับความรู้ที่ใช้ในโครงการที่จะทำ ข้อมูลส่วนใหญ่ที่ค้นคว้าจะค้นคว้ามาจากหนังสือ นอกจากนี้เรายังมีการค้นคว้าข้อมูลจากอินเทอร์เน็ตเพิ่มขึ้นอีกด้วย การเลือกข้อมูลนั้นจะมีการเลือกข้อมูลที่สอดคล้องกับเรื่องที่จะทำ ซึ่งการค้นข้อมูลนั้น โดยก่อนที่จะเริ่มหาข้อมูลเราจะมีการปรึกษาหารือกันก่อน เพื่อแบ่งการค้นคว้านั้นไป จากนั้นก็เอาข้อมูลที่ได้มารวมกัน และจะทำการวิเคราะห์หาส่วนสำคัญออกมา

ในการวิเคราะห์นั้นจะดูว่าข้อมูลที่เราหามานั้น ตรงกับความต้องการ และมีเนื้อหาสอดคล้องกับโครงการของเราหรือไม่

เครื่องมือและวิธีการดำเนินการวิจัย

สำหรับเครื่องมือที่ใช้ในการทำวิจัยประกอบด้วย

1. Microsoft Visual C++ 6.0
2. Microsoft Access 97
3. Adobe PhotoShop 6.0
4. Microsoft Word 97
5. Microsoft Windows 98 Operating System

ในส่วนของการดำเนินการวิจัย

1. ใช้ทฤษฎีของ Natural Language Processing เข้ามาช่วยในการแยกประโยคคำภาษาไทย
2. มีการเก็บฐานความรู้(Knowledge Base) เพื่อนำมาตอบผู้ใช้งาน
3. ใช้การเก็บพจนานุกรมภาษาไทย โครงสร้างแบบ Trie เพื่อช่วยในการวิเคราะห์ประโยค
4. ใช้เทคโนโลยี Microsoft Foundation Class ใน Microsoft Visual C++ ช่วยในการเขียนโปรแกรม

ส่วนของ Software แบ่งเป็นส่วนได้

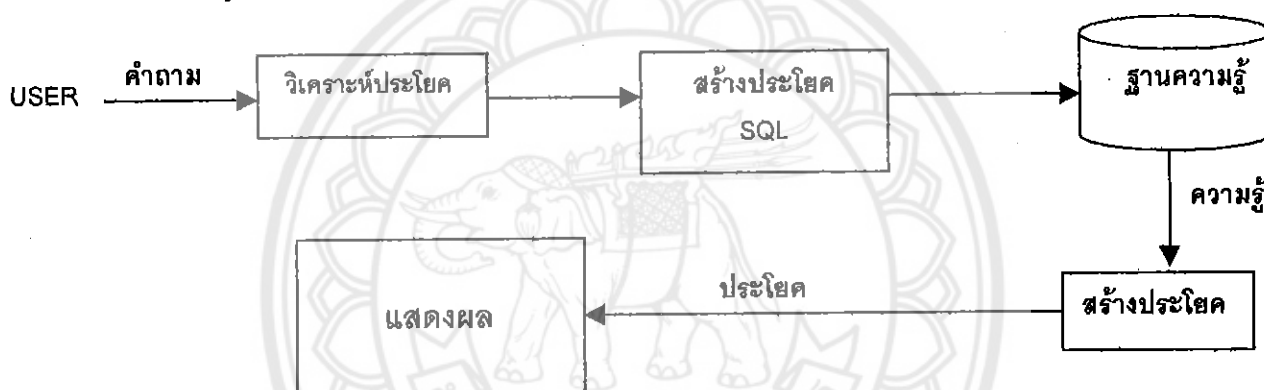
1. เกี่ยวกับการวิเคราะห์ประโยค โดยที่เราจะมีการเก็บข้อมูลพจนานุกรมภาษาไทย โครงสร้างแบบ Trie แล้วเขียนโปรแกรมในรูปแบบ Microsoft Foundation Class (MFC) ใน Microsoft Visual C++ เพื่อวิเคราะห์ประโยค

2. ส่วนที่ใช้เชื่อมกับผู้ใช้จะเขียนด้วย Microsoft Visual C++ จะเขียนในรูปของ Dialog เพื่อให้ผู้ใช้สามารถรอกคำถามไปได้และมีลูกเล่นต่าง ๆ

3. ส่วนติดต่อกับฐานความรู้ จะเขียนคำสั่งในรูปของ SQL เพื่อไปเรียกข้อมูลจากฐานความรู้ที่เก็บโดยใช้ Microsoft Access

ลักษณะของ Software

จากรูปที่ 3.1 จะแสดงการทำงานของโปรแกรมที่จัดทำขึ้นมา



รูปที่ 3.1 รูปแบบการทำงาน โดยสังเขป

จะเน้น Input และ Output จึงเป็นดังนี้

Input:

1. ประโยคคำถามเกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา

Output:

1. ประโยคคำตอบเกี่ยวกับประวัติศาสตร์สมัยกรุงศรีอยุธยาที่อยู่ในรูปข้อความ

ซอฟต์แวร์ที่พัฒนาจึงแบ่งเป็น

1. ส่วนของการตัดคำ ใช้การเขียนโปรแกรมแบบ Microsoft Foundation Class ใน Microsoft Visual C++
2. ส่วนเชื่อมต่อกับผู้ใช้
3. ส่วนของการเชื่อมต่อกับฐานความรู้ ใช้การหาข้อมูลแบบ Query เพื่อค้นหาข้อมูลมาใช้ตอบ
4. ส่วนของการเก็บฐานความรู้

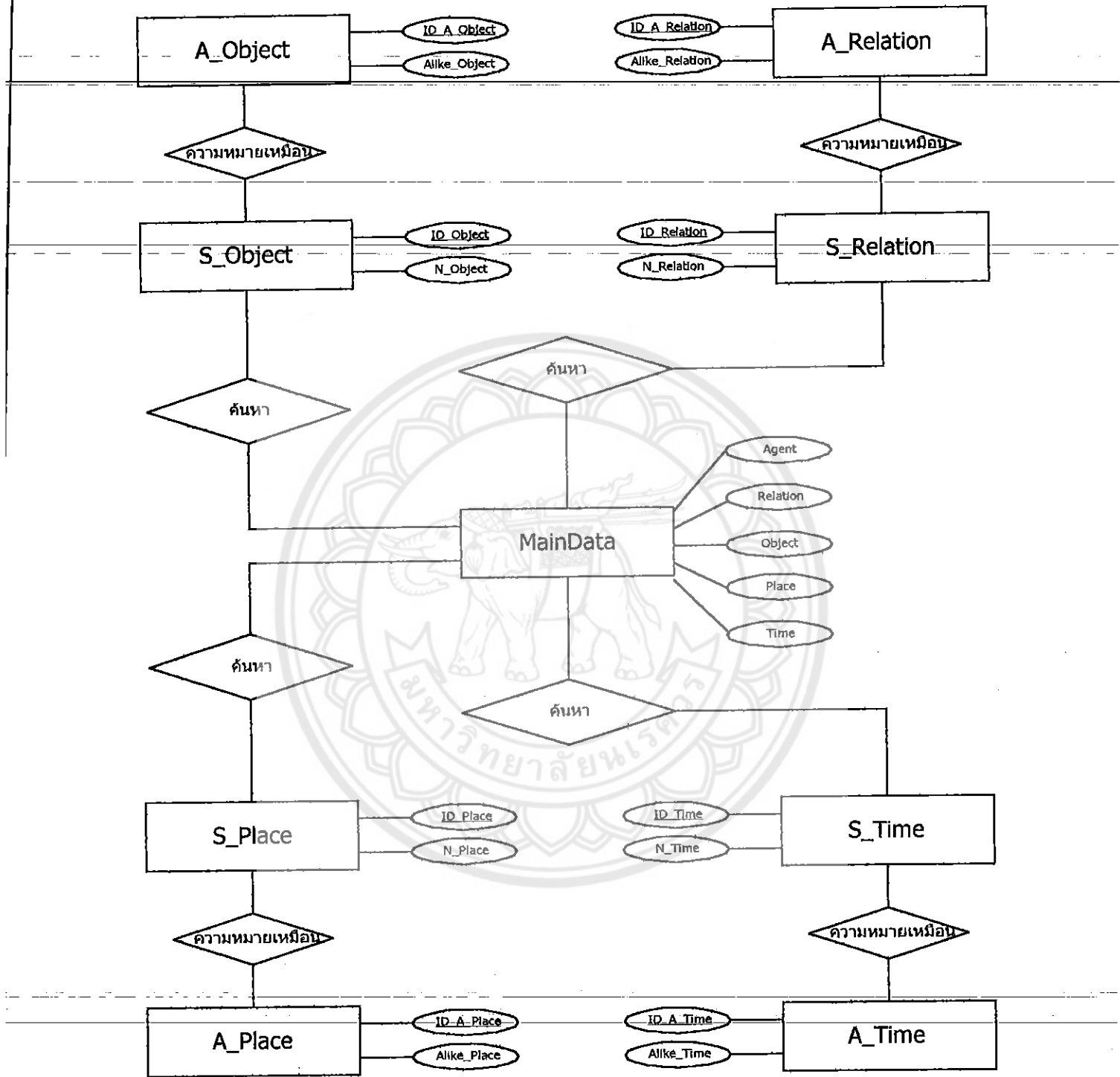
3.3 การออกแบบฐานข้อมูล

เมื่อกำหนดรูปแบบของโปรแกรมต่างๆ ได้แล้ว ก็จะทำการออกแบบฐานข้อมูลที่ใช้ในโครงการ ซึ่งจะใช้เป็นฐานความรู้ในการตอบคำถามที่ผู้ใช้ป้อนเข้ามา ฐานข้อมูลจะเป็นข้อมูลที่นำมาจากหนังสือเรื่อง เส้นทางประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา การเก็บฐานข้อมูลประกอบด้วย ฐานข้อมูลหลักในส่วนฐานข้อมูลหลักนั้นจะเป็นการเก็บความรู้ทั้งหมด เราจะแบ่ง Field ของฐานข้อมูลเป็น 6 field ประกอบด้วย

1. Agent ใช้เก็บในส่วนของผู้กระทำ
2. Relation ใช้เก็บความสัมพันธ์ของ Agent และ Object
3. Object ใช้เก็บกรรมหรือตัวถูกกระทำ ซึ่งเราแบ่งเป็น Object และ Object2
4. Place ใช้เก็บสถานที่ ที่เกี่ยวข้องกับ Agent และ Object นั้นๆ
5. Time ใช้เก็บเวลาของการกระทำของ Agent และ Object นั้นๆ

และนอกจากนี้ยังมีการเก็บฐานข้อมูลย่อยซึ่งการเก็บฐานข้อมูลย่อยเป็นการเก็บความรู้ที่มีความสัมพันธ์กันหลายคำ เช่นคำว่าพ่อ จะมีคำที่เหมือนกัน เช่นคำว่า พระราชบิดา พระบิดา เป็นต้น ในฐานข้อมูลหลักจะเก็บเป็นคำที่สมบูรณ์ แล้วจากนั้นจึงมาเชื่อมต่อกับฐานข้อมูลย่อยเพื่อหาคำศัพท์ที่มีความหมายที่เหมือนกัน ในส่วนของฐานข้อมูลย่อยนั้นเราจะเก็บ เป็น 8 ฐานข้อมูลด้วยกันตามลักษณะของความหมายของคำที่มีหน้าที่ต่างๆ ในประโยค ได้แก่ S_Object (ผู้กระทำและผู้ถูกกระทำ), S_Relation (กริยา), S_Place (สถานที่) และ S_Time (เวลา) และอีก 4 ฐานข้อมูลซึ่งเป็นฐานข้อมูลของคำเหมือน ได้แก่ A_Object (คำเหมือนของผู้กระทำและผู้ถูกกระทำ), A_Relation (คำเหมือนของกริยา), A_Place (คำเหมือนของสถานที่) และ A_Time (คำเหมือนของเวลา)

3.3.1 ER Diagram ของฐานข้อมูลที่ใช้ในโครงการงาน



รูปที่ 3.2 ER Diagram ของฐานข้อมูลที่ใช้ในการทำโครงการงาน

4400602

QA

76.9.D3

01525

2544C.2

จากรูปที่ 3.2 เราสามารถเขียนแยกเป็นตารางได้ดังนี้

MainData

Agent	Relation	Object	Place	Time
-------	----------	--------	-------	------

รายละเอียดฐานข้อมูลหลัก

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
Agent	Agent	Text	50	ผู้กระทำ
Relation	Relation	Text	50	กริยา
Object	Object	Text	50	ผู้ถูกกระทำ
Place	Place	Text	50	สถานที่
Time	Time	Text	50	เวลา

A_Object

ID_A_Object	Alike_Object
-------------	--------------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_A_Object	Idintity Alike Object	Number	-	หมายเลขกำกับคำ
Alike_Object	Alike Object	Text	50	คำเหมือนผู้ กระทำและผู้ถูก กระทำ

A_Relation

ID_A_Relation	Alike_Relation
---------------	----------------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Collumn	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_A_Relation	Idintity Alike Relation	Number	-	หมายเลขกำกับคำ
Alike_Relation	Alike Relation	Text	50	คำเหมือนกริยา

A_Place

ID_A_Place	Alike_Place
------------	-------------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Collumn	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_A_Place	Idintity Alike Place	Number	-	หมายเลขกำกับคำ
Alike_Place	Alike Place	Text	50	คำเหมือนสถานที่

A_Time

ID_A_Time	Alike_Time
-----------	------------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Collumn	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_A_Time	Idintity Alike Time	Number	-	หมายเลขกำกับคำ
Alike_Time	Alike Time	Text	50	คำเหมือนเวลา

S_Object

ID_Object	N_Object
-----------	----------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_Object	Idintity Object	Number	-	หมายเลขกำกับคำ
N_Object	Name Object	Text	50	ชื่อผู้กระทำ

S_Relation

ID_Relation	N_Relation
-------------	------------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_Relation	Idintity Relation	Number	-	หมายเลขกำกับคำ
N_Relation	Name Relation	Text	50	ชื่อกริยา

S_Object

ID_Place	N_Place
----------	---------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_Place	Idintity Place	Number	-	หมายเลขกำกับคำ
N_Place	Name Place	Text	50	ชื่อสถานที่

S_Time

ID_Time	N_Time
---------	--------

รายละเอียดฐานข้อมูลคำเหมือนผู้กระทำและผู้ถูกกระทำ

ชื่อ Column	ชื่อภาษาอังกฤษ	ประเภทข้อมูล	ขนาด	ความหมาย
ID_Time	Idintity Time	Number	-	หมายเลขกำกับคำ

N_Time	Name Time	Text	50	ชื่อเวลา
--------	-----------	------	----	----------

ตัวอย่างของฐานข้อมูลแสดงได้ดังรูปที่ 3.3 ข้างล่างนี้

Agent	Relation	Object	Place	Time
พระเจ้าหงสาวดี	สวรรคต			พ.ศ. 2124
พระเจ้าหงสาวดีมูรงนอง	สวรรคต			พ.ศ. 2124
สมเด็จพระนเรศวร	หลังน้ำ		เมืองเมฆ	พ.ศ. 2127
สมเด็จพระนเรศวร	ประกาศ	อิสรภาพ	เมืองเมฆ	พ.ศ. 2127
สมเด็จพระนเรศวร	ทำสงคราม	พระเจ้าเชียงใหม่	ที่บ้านสระบก	พ.ศ. 2128
สมเด็จพระนเรศวรและสมเด็จพระสุริเย		พระเจ้าเชียงใหม่	ที่บ้านสระบก	พ.ศ. 2128
พระเจ้าหงสาวดีนันทบุเรง	ยกทัพไปตี	กรุงศรีอยุธยา	กรุงศรีอยุธยา	พ.ศ. 2129
สมเด็จพระนเรศวร	ครองราชย์			พ.ศ. 2133
สมเด็จพระนเรศวร	ประกาศ	เมืองพรวนบุรี	เมืองพรวนบุรี	พ.ศ. 2134
สมเด็จพระนเรศวร	สร้าง	พระตำหนัก	ที่ตำบลโตนดหลวง	พ.ศ. 2134
สมเด็จพระนเรศวร	ยกทัพไปตี	เมืองทวายและตะนาวศรี	เมืองทวายและตะนาวศรี	พ.ศ. 2135
สมเด็จพระนเรศวร	ประหาร	พระยาละแวก	เมืองละแวก	พ.ศ. 2135
พระสุพรรณภักดี	สิ้นพระชนม์			พ.ศ. 2135
สงครามยุทธหัตถี	เกิด		ที่ตำบลหนองค้ำพ่าย	พ.ศ. 2135
พระสุวรรณหงส์	สิ้นพระชนม์			พ.ศ. 2135
สมเด็จพระนเรศวร	ยกทัพไปตี	เมืองกันทูลา	เมืองกันทูลา	พ.ศ. 2136
เมืองเมาะตะมะ	เมืองขึ้น	กรุงศรีอยุธยา		พ.ศ. 2137
สมเด็จพระนเรศวร	ยกทัพไปตี	กรุงหงสาวดี	กรุงหงสาวดี	พ.ศ. 2138-2142

รูปที่ 3.3 ตัวอย่างการเก็บฐานข้อมูลหลัก

3.4 เขียนโปรแกรมระบบฐานความรู้

หลังจากออกแบบฐานข้อมูลและ สร้างฐานข้อมูลเสร็จแล้วก็จะทำการเขียนโปรแกรม ระบบฐานความรู้โดยออกแบบ 3 ส่วนคือ โปรแกรมในส่วนของการตัดคำ โปรแกรมในส่วนของการกำหนดหน้าที่ของคำ และ โปรแกรมในส่วนของการติดต่อกับฐานข้อมูลเพื่อนำข้อมูลในฐานข้อมูลมาสร้างประโยคคำตอบ โดยโปรแกรมทั้ง 3 ส่วนจะมีการทดลองเขียนขึ้นก่อน หลังจากนั้นก็จะนำมารวมกันเป็นโปรแกรมที่สำเร็จเป็นโปรแกรมทั้งหมด เป็นโปรแกรมที่สามารถตอบคำถามเกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยาได้ โดยใช้ภาษาธรรมชาติ

3.5 ทำการติดตั้งและทดลองโปรแกรม

ในด้านการติดตั้งและทดลองการใช้งานของโปรแกรมนั้น เราจะทำการเมื่อเราสร้างโปรแกรมเสร็จแล้ว ในด้านการทดสอบโปรแกรมนั้นเราจะทดสอบโดยการให้ผู้ใช้งานทั้งหมด 50 คนมาลองตั้งคำถามคนละ 10 คำถามและลอง ไล่คำถามเหล่านั้นไปและสังเกตผลพร้อมหาข้อบกพร่องของโปรแกรมและแก้ไข ปรับปรุงให้โปรแกรมสามารถใช้งานได้ครอบคลุมกับความรู้ให้มากที่สุด

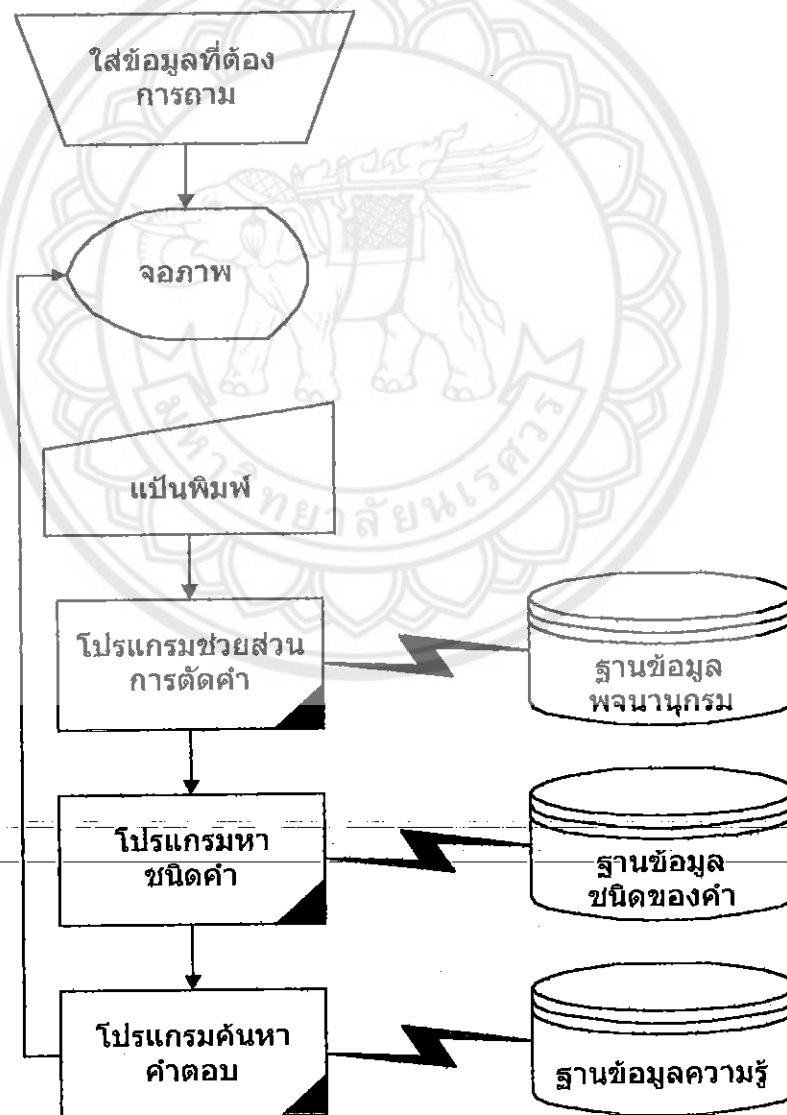
บทที่ 4

การออกแบบโปรแกรม

4.1 ส่วนประกอบของโปรแกรม

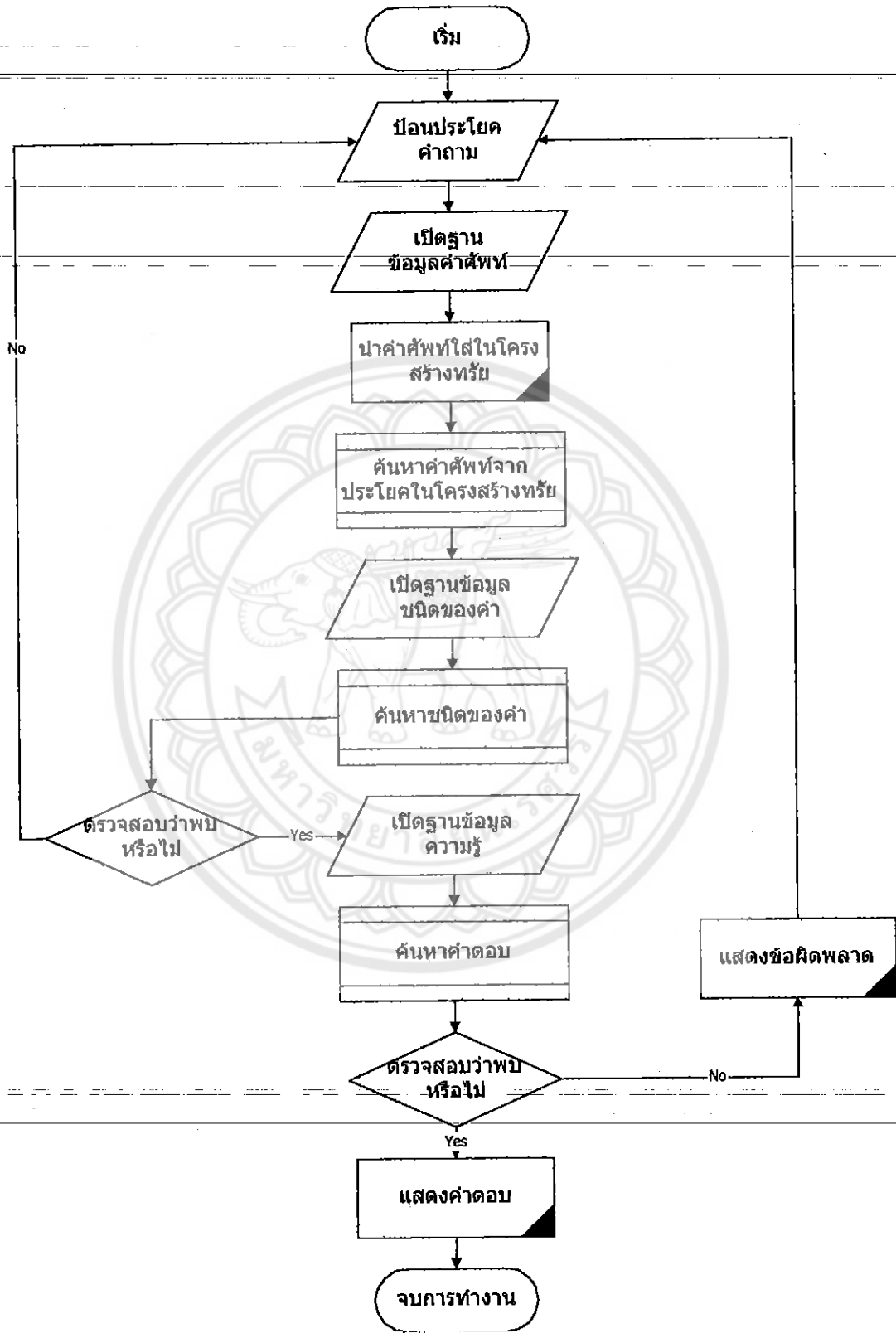
ในการพัฒนาโปรแกรมนั้น จะแบ่งการพัฒนาโปรแกรมออกเป็นส่วนหลัก ๆ 3 ส่วนในการพัฒนาโปรแกรมคือ ส่วนที่เกี่ยวกับการตัดคำ ส่วนที่หาส่วนประกอบของประโยค และส่วนที่ดึงคำตอบของประโยค จากนั้นจึงนำโปรแกรมทั้ง 3 ส่วนมาประกอบกันเป็นโปรแกรมหลักขึ้นมา

4.2 ผังระบบงานแสดงการทำงานของโปรแกรม



รูปที่ 4.1 แพลนผังการทำงานของโปรแกรม

4.3 ฟังก์ชันโปรแกรมแสดงการทำงานของโปรแกรม



รูปที่ 4.2 ฟังก์ชันโปรแกรมแสดงการทำงานของโปรแกรม

4.4 ตัวอย่างการใช้งานโปรแกรม

The screenshot shows a window titled "Untitled - SearchDao" with a menu bar (File, Edit, View, Help) and a toolbar. The main area contains a search form with the following elements:

- A label "กรุณาใส่ประโยคคำถามที่นี่" (Please enter the question sentence here).
- A text input field containing "ศรีปราชญ์เป็นกวีสมัยใด" (Which era of Sri Prachuan was a poet?).
- A "Search" button.
- A "Result" label above an empty text input field.
- Fields for "Agent", "Relation", "Object", "Place", "Time", and "Tagit" (likely a typo for Tag), each with an empty input field.

รูปที่ 4.3 ฟอรัมให้กรอกคำถาม

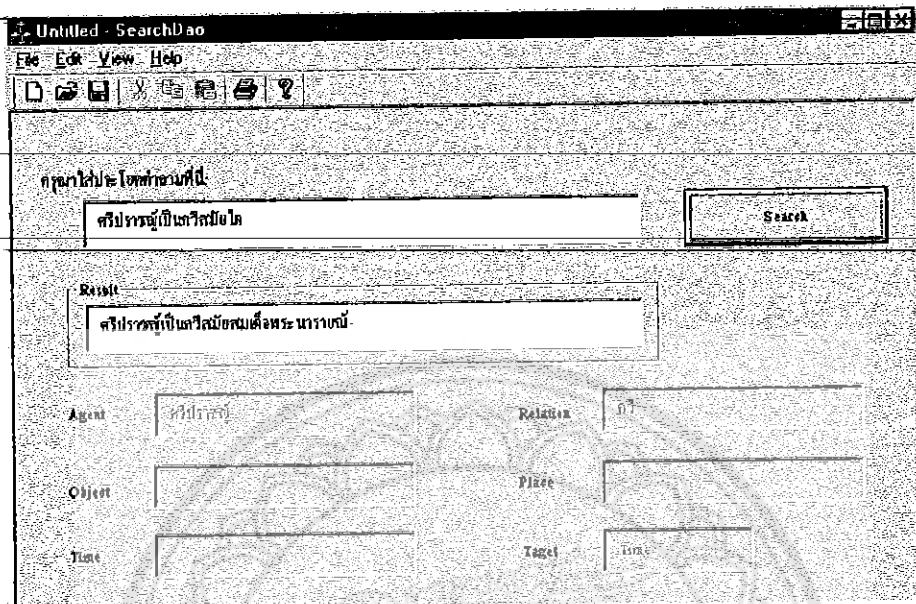
จากรูปที่ 4.3 การกรอกคำถามลงไปในฟอร์มนี้ผู้ใช้กรอกคำถามเป็นภาษาธรรมชาติจากตัวอย่างผู้ใช้กรอกคำถาม ด้วยประโยค "ศรีปราชญ์เป็นกวีสมัยใด" ลงในช่องว่างที่มีประโยค "กรุณาใส่ประโยคคำถามที่นี่" เมื่อผู้ใช้กดปุ่ม Search หรือคีย์ Enter จะได้ผลออกมาดังรูปที่ 4.4

The screenshot shows the same SearchDao window as in Figure 4.3, but with the following changes:

- The "Result" field now contains the text "ศรีปราชญ์เป็นกวีสมัยใด" (Which era of Sri Prachuan was a poet?).
- A small dialog box titled "SearchDao" is open over the result field, containing the same text "ศรีปราชญ์เป็นกวีสมัยใด" and an "OK" button.
- The "Agent", "Relation", "Object", "Place", "Time", and "Tagit" fields remain empty.

รูปที่ 4.4 การตัดคำประโยคภาษาไทย

ในรูปที่ 4.4 โปรแกรมจะแสดงการตัดคำของประโยคที่เราพิมพ์ใส่เข้าไปโดย โปรแกรมจะแสดงคำเป็นคำและจะคั่นด้วยเครื่องหมาย # ระหว่างแต่ละคำ และเมื่อผู้ใช้กดปุ่ม OK หรือปุ่ม Enter แล้วนั้น โปรแกรมก็จะแสดงคำตอบออกมาที่ช่อง Result แสดงได้ดังรูปที่ 4.5



รูปที่ 4.5 โปรแกรมแสดงคำตอบออกมา

ในรูปที่ 4.5 นั้นจะได้คำตอบออกมาเป็น “ศรีปราชญ์เป็นกวีสมัยสมเด็จพระนารายณ์” ซึ่งเป็นคำตอบโดยคำตอบที่ตอบออกมานั้นสังเกตได้ว่าเป็นภาษาธรรมชาติ และตรงกับคำถามที่ถามเข้าไป

บทที่ 5

สรุปและวิเคราะห์ผล

5.1 สรุป

โครงการ ระบบฐานความรู้เกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา เป็นโครงการที่จัดทำขึ้น โดยมีวัตถุประสงค์หลักที่จะศึกษาการกระบวนการภาษารวมชาติ และศึกษากระบวนการสร้างประโยคภาษาไทย ซึ่งผลที่ได้รับจากกรทำโครงการชิ้นนี้คือโปรแกรมที่สามารถตอบคำถามเกี่ยวกับประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา โดยได้สร้างฐานข้อมูลภาษาไทย มีลักษณะการจัดเก็บจากการแตกประโยคที่ต้องการเป็นคำหรือกลุ่มคำตามหน้าที่ของคำไปเก็บในแต่ละคอลัมน์ พร้อมกันนี้ได้ทำการทดสอบโปรแกรมจากผู้ใช้งาน 50 คน คนละ 10 คำถาม สามารถตอบคำถามได้ ประมาณ 54.6 เปอร์เซ็นต์ คิดจากคำถาม ที่ตอบได้ทั้งหมด 273 คำถามจากคำถามทั้งหมด 500 คำถาม และสรุปมาดังตาราง

ตารางที่ 5.1 ตัวอย่างประโยคคำถามที่ได้สุ่มมาจากผู้ใช้

คำถาม	ตอบ		หมายเหตุ
	ได้	ไม่ได้	
1.เจ้าฟ้าศรีเสาวภาคย์เป็น โอรสของใคร	×		
2.พระเจ้าอยู่หัวองค์ไหนสมรสกับใคร	×		
3.น้ำพระเนตรพระพุทธชินราชปรากฏเมื่อใด		×	ไม่มีในฐานข้อมูล
4.ท้าวติโลกราชเป็นกษัตริย์เมืองใด	×		
5.กรุงศรีอยุธยาเสียกรุงครั้งแรกเมื่อใด	×		
6.ใครเป็นผู้ขับไล่ทัพท้าวติโลกราชที่เข้ายึดเมืองกำแพงเพชร		×	ประโยคคำถามมีตัวกระทำ 2 ตัว ซึ่งในฐานข้อมูลมีตัวเดียว
7.โอรสของพระนเรศวรมีใครบ้างที่ได้ครองราชย์		×	ไม่มีในฐานข้อมูล
8.ใครเป็นกษัตริย์องค์ที่ 2 ของกรุงศรีอยุธยา	×		
9.คลองโคกขามอยู่แขวงเมืองใด	×		
10.กษัตริย์องค์สุดท้ายที่ครองกรุงศรีอยุธยาพระนามว่าอะไร		×	ไม่มีในฐานข้อมูล

5.2 ข้อเสนอแนะ

โปรแกรมนี้ยังไม่สามารถครอบคลุมการทำงานของกระบวนการภาษาธรรมชาติจำเป็นต้องมีการพัฒนาเพิ่มเติมในส่วนของ

1. พัฒนาโปรแกรมให้สามารถตอบคำถามด้านเหตุผลได้
 2. การตัดคำ กล่าวคือยังตัดคำในรูปแบบคำยาวทำให้ได้ความหมายที่ไม่ถูกต้อง
 3. ระบบฐานความรู้ ต้องมีการเก็บความรู้ที่มีความละเอียดมากๆ ครอบคลุมเนื้อหาทั้งหมดตามที่เราต้องการ
 4. เรื่องของการวิเคราะห์ประโยคภาษาไทย หาวิธีการที่จะวิเคราะห์ประโยคภาษาไทยได้ ครอบคลุมทุกรูปแบบของประโยคภาษาไทย
- คณะผู้จัดทำหวังเป็นอย่างยิ่งว่าจะมีผู้ให้นำโปรแกรมนี้ไปศึกษา และพัฒนาต่อไปให้ได้ โปรแกรมที่มีความสมบูรณ์ยิ่งขึ้น



เอกสารอ้างอิง

- [1] - Ishizaki, S. et al. (1990). **Concept Dictionary As Interlingua**. Bul. Electrotech. Lab. Vol. 54 No.4, 432-495.
- [2] Winograd, T. (1992) **Understand Natural Language**, Academic Press.
- [3] Wendy Sarrett, 1998 **Visual C++ Database Programmint Tutorial**, Wrax Press.
- [4] นิรุช อำนวยศิลปี่, 2542 คู่มือการเขียนโปรแกรม Microsoft Visual C++ 6.0, บริษัท ส.เอเชียเพรส จำกัด
- [5] ยุทธนา ลีลาวัฒน์กุล, **Advance Visual C++ Version 6.0**, บริษัท ชัคเซส มีเดีย จำกัด
- [6] สิริวัฒน์ ไชยชนะ, 2543 **เส้นทางประวัติศาสตร์ไทยสมัยกรุงศรีอยุธยา**, ห้างหุ้นส่วนจำกัด วี เจ ปรีนติ้ง
- [7] ไพศาล เจริญพรสวัสดิ์, 2541 **การตัดคำภาษาไทยโดยใช้คุณลักษณะ**, วิทยานิพนธ์ ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย



ประวัติผู้เขียนโครงการ

ชื่อ	นายคมศักดิ์ สุขเกษม
วัน / เดือน / ปีเกิด	7 มิถุนายน 2522
ภูมิลำเนา	238/1 หมู่2 ต.บ้านแก่ง อ.ตรอน จ.อุตรดิตถ์ 53140
ประวัติการศึกษา	ระดับประถมศึกษา โรงเรียนชุมชนบ้านแก่ง(ไกรสรพงษ์สงเคราะห์) จ.อุตรดิตถ์ ระดับมัธยมศึกษา โรงเรียนอุตรดิตถ์
ปัจจุบัน	กำลังศึกษาอยู่ ณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร จ.พิษณุโลก หลักสูตรวิศวกรรมคอมพิวเตอร์
ชื่อ	นายวิทวัส อริยะสุนทร
วัน / เดือน / ปีเกิด	18 กันยายน 2522
ภูมิลำเนา	80/2 หมู่5 ต.ตาจืด อ.บรรพตพิสัย จ.นครสวรรค์ 60180
ประวัติการศึกษา	ระดับประถมศึกษา โรงเรียนลาซาลโชติรวินนครสวรรค์ ระดับมัธยมศึกษา โรงเรียนลาซาลโชติรวินนครสวรรค์
ปัจจุบัน	กำลังศึกษาอยู่ ณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร จ.พิษณุโลก หลักสูตรวิศวกรรมคอมพิวเตอร์

ชื่อ	นายสุรชาติ โอปิ่น
วัน / เดือน / ปีเกิด	21 มิถุนายน 2523
ภูมิลำเนา	239 หมู่ 7 ต.บึงสามพัน อ.บึงสามพัน จ.เพชรบูรณ์ 67160
ประวัติการศึกษา	ระดับประถมศึกษา โรงเรียนสมอท้อศึกษา จ.เพชรบูรณ์ ระดับมัธยมศึกษา โรงเรียนบึงสามพันวิทยาคม จ.เพชรบูรณ์
ปัจจุบัน	กำลังศึกษาอยู่ ณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร จ.พิษณุโลก หลักสูตรวิศวกรรมคอมพิวเตอร์

