

รายงาน

เรื่อง

การแยกแยะหน่วยคำออกจากย่อหน้า
เพื่อการแปลภาษาด้วยคอมพิวเตอร์

โดย

นายสมศักดิ์ จันวัน

สำนักวิจัยและบริการคอมพิวเตอร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปี 2540

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
สารบัญ	2
บทที่ 1 บทนำ	3
1.1 ความเป็นมาของงานวิจัย	3
1.2 จุดประสงค์และขอบเขตของงานวิจัย	3
บทที่ 2 ความรู้พื้นฐานเกี่ยวกับชนิดของคำและ ไวยากรณ์ที่ใช้ในการวิเคราะห์	6
2.1 บทนำ	6
2.2 การแบ่งประเภทของคำในภาษาไทย	6
2.3 โครงสร้างของประโยคภาษาไทย	6
บทที่ 3 การจัดเก็บฐานข้อมูลคำศัพท์	11
3.1 .1 บทนำ	11
3.2 การจัดเก็บข้อมูลลงพจนานุกรม	11
3.3 ชนิดของคำศัพท์	13
บทที่ 4 การแยกแยะหน่วยคำออกจากประโยคและการตรวจสอบ ไวยากรณ์	16
4.1 บทนำ	16
4.2 การแยกหน่วยคำออกจากประโยค	16
บทที่ 5 การทดลอง	25
5.1 บทนำ	25
5.2 การทดลอง	25
5.3 สรุปผลการทดลอง	30

RCH

GA

76.76

T83

๗๖๗๖ ค-๒

เลขหมู่.....

เลขทะเบียน..... 64411

วัน,เดือน,ปี 1 1 ก.ย. 2549

b. 109 86212
i.....

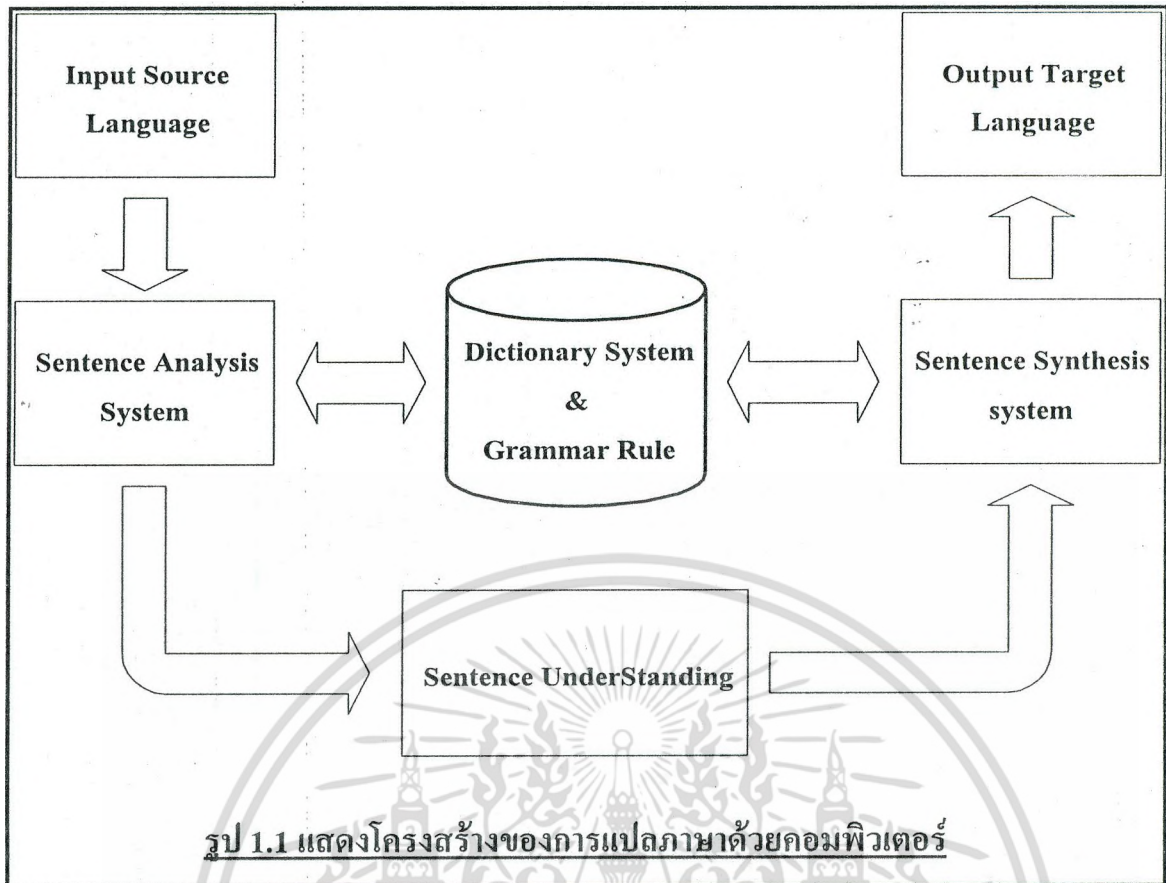
บทที่ 1 บทนำ

1.1 ความเป็นมาของงานวิจัย

ภาษาเป็นสื่อกลางที่ช่วยให้มนุษย์สามารถติดต่อสื่อสารกันได้ เพื่อถ่ายทอดความคิดของบุคคลผู้นั้นไปยังบุคคลที่ต้องการติดต่อสื่อสารกัน และเนื่องด้วยมนุษย์มีอยู่จำนวนมากมายกระจ่ายไปทุกมุมโลกของโลก ดังนั้นจึงเป็นการยากที่มนุษย์แต่ละคนที่ใช้ภาษาเดียวกัน ดังนั้นภาษาที่ใช้ในแต่ละแห่งจึงเป็นภาษาที่มีเอกลักษณ์เป็นของตนเอง ทั้งชนิดที่เป็นภาษาพูดและภาษาเขียน เมื่อเทคโนโลยีได้มีการพัฒนามากยิ่งขึ้นก็ทำให้การติดต่อสื่อสารกันระหว่างบุคคลต่างเชื้อสายต่างเผ่าพันธุ์ก็มีมากยิ่งขึ้นทั้งการแลกเปลี่ยนวัฒนธรรม วิชาการ การเดินทางท่องเที่ยว การแลกเปลี่ยนมนการความคิด การค้าขาย และอื่นๆ อีกมากมาย แต่อุปสรรคอย่างหนึ่งที่ทำให้การติดต่อสื่อสารกันนั้นไม่มีประสิทธิภาพ นั่นก็คือ ภาษา ซึ่งถ้าต้องการจะติดต่อสื่อสารกันก็ต้องใช้บุคคล ๑ คนทำหน้าที่ในการแปลภาษาจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่ง จึงจะทำให้การติดต่อสื่อสารบรรลุผลสำเร็จ ดังนั้นการแปลภาษาจึงเป็นกุญแจสำคัญอันหนึ่งซึ่งช่วยให้การติดต่อสื่อสารเกิดประโยชน์มากที่สุด ดังนั้นจึงได้มีการพัฒนาให้ Computer สามารถช่วยเป็นตัวกลางในการแปลภาษา

การแปลภาษาด้วยระบบคอมพิวเตอร์นั้นได้มีการคิดค้นและทดลองกันมาเป็นเวลาหลายปี โดยมีวัตถุประสงค์หลักเพื่อช่วยงานในด้านการแปลภาษาที่มีปริมาณมาก และเหตุผลอีกประการหนึ่งก็คือ บุคลากรที่ทำหน้าที่นี้เพื่อมีค่าใช้จ่ายสูงและมีขีดจำกัดในกรณีที่มีปริมาณงานจำนวนมาก โดยหลักของการแปลจะเกี่ยวข้องกับการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ร่วมกับวิทยาการทางด้าน Software (Computer Software Technology) และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistic) ซึ่งเป็นการจำลองการทำงานด้านการแปลของมนุษย์ โดยมีหลักการที่ว่ามนุษย์เรานั้นมีหลักการคิดอย่างไรในการเข้าใจภาษาแล้วจะนำเอาหลักการเดียวกันนั้นไปแปลเป็นภาษา Computer เพื่อให้ Computer นั้นมีความสามารถเช่นเดียวกับมนุษย์

การแปลภาษาของคอมพิวเตอร์นั้นสามารถทำงานได้ตลอดเวลาซึ่งมนุษย์นั้นไม่สามารถทำได้ดังนั้นถ้าได้มีการพัฒนาให้ Computer สามารถทำได้ดังจุดประสงค์ที่ตั้งไว้ก็จะเป็นการช่วยในการแปลภาษาได้ เป็นจำนวนมากได้ โดยมีหลักการทำงานเบื้องต้นดังรูป 1.1

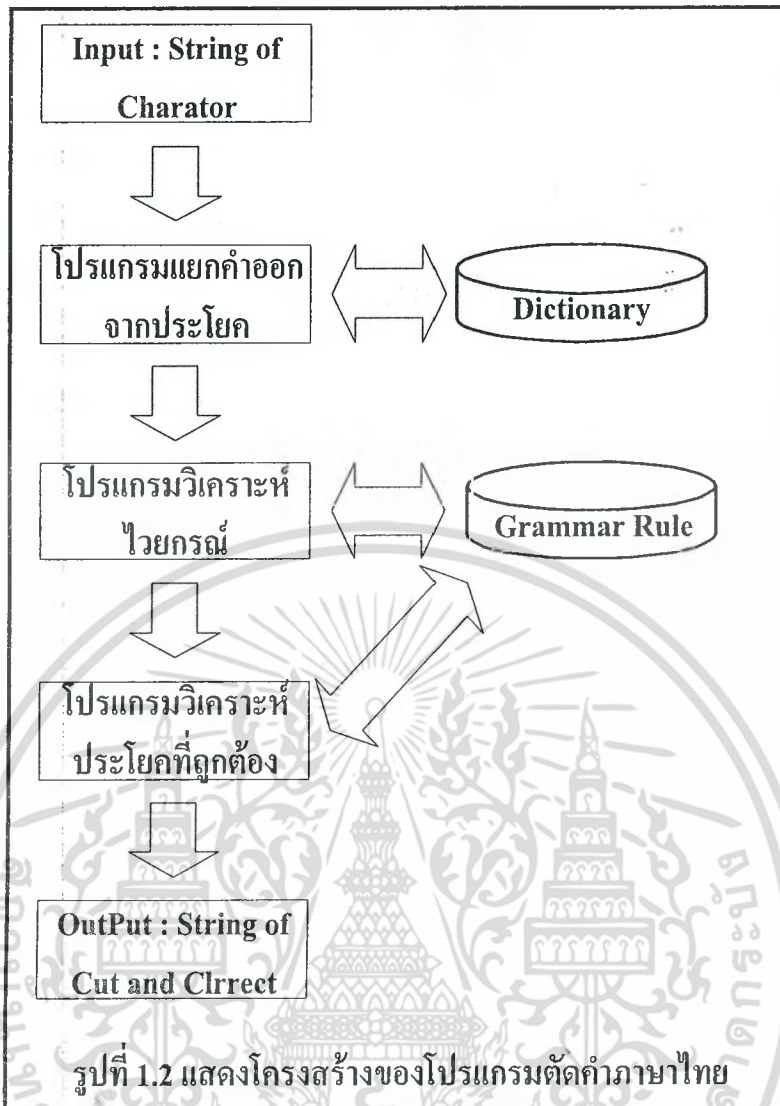


1.2 จุดประสงค์และขอบเขตของงานวิจัย

งานวิจัยชิ้นนี้จะมีหลักการทำงานคล้ายกับ โครงสร้างของการแปลภาษาด้วยคอมพิวเตอร์ ซึ่งมีหลักการงานหลักๆ ที่คล้ายกันมากโดยหลักการทำงานของ โปรแกรมตัดคำด้วยคอมพิวเตอร์ จะสามารถแยกออกเป็นส่วนๆ ได้ดังต่อไปนี้

1. ส่วนโปรแกรมแยกคำ
2. ส่วนโปรแกรมวิเคราะห์ไวยากรณ์
3. ส่วนโปรแกรมพิจารณาข้อความที่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



หลักการดำเนินงานเบื้องต้นมีดังต่อไปนี้

1. ส่วน โปรแกรมแยกคำจากประโยค โดยในส่วนนี้จะนำเอาประโยคที่รับเข้ามาทำการแยกออกเป็นคำๆ โดยใช้ข้อมูลอ้างอิงภายใน Dictionary โดยการเปรียบเทียบคำที่อยู่ในประโยคกับคำที่อยู่ใน Dictionary ว่าเหมือนกันหรือไม่ ถ้าเหมือนกันก็จะแยกคำนั้นออกมา
2. ส่วน โปรแกรมวิเคราะห์ไวยากรณ์ โดยในส่วนนี้จะนำเอาคำที่แยกได้จากส่วนโปรแกรมแยกคำออกจากประโยคโดยนำเอาคำที่ได้มาทำการตรวจสอบกับกฎไวยากรณ์ที่ได้ทำการตั้งไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ส่วนโปรแกรมวิเคราะห์ประโยคที่ถูกต้อง โดยในส่วนนี้จะนำเอาประโยคที่ได้ทำการตรวจสอบไวยากรณ์แล้วว่ามี ความถูกต้องตามหลักไวยากรณ์แต่เนื่องจากอาจมีประโยคถูกต้องตามหลักไวยากรณ์มากกว่า 1 ประโยคคั้งนั้นจึงต้องมาพิจารณาว่าประโยคใดถูกต้องมากที่สุด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ความรู้พื้นฐานเกี่ยวกับชนิดของคำและไวยากรณ์ที่ใช้ในการวิเคราะห์

2.1 บทนำ

ความรู้พื้นฐานเกี่ยวกับชนิดของคำและไวยากรณ์ที่ใช้ในการวิเคราะห์ที่ใช้ในงานวิจัยชิ้นนี้รวบรวมมาจากงานวิจัยด้านการวิเคราะห์โครงสร้างของภาษาไทย และเอกสารประกอบที่เกี่ยวข้องกับหลักภาษาไทยซึ่งจำเป็นมากในการศึกษาเกี่ยวกับภาษาไทย โดยจะเน้นในด้านการวิเคราะห์โครงสร้างภาษาไทยด้วยระบบ Software Computer ร่วมกับทฤษฎีทางด้านภาษาศาสตร์คอมพิวเตอร์และการประมวลผลภาษาธรรมชาติ ดังนั้นในงานวิจัยชิ้นนี้จึงได้แบ่งประเภทของคำและโครงสร้างทางไวยากรณ์ภาษาตามหลักทฤษฎีของการประมวลผลภาษาธรรมชาติ

2.2 การแบ่งประเภทของคำในภาษาไทย

ในการแบ่งประเภทของคำจะยึดถือการแบ่งคำตามหน้าที่ของคำเป็นหลัก โดยจะเน้นการวิเคราะห์ภาษาไทยตามหลักการประมวลผลภาษาธรรมชาติ ดังนั้นจึงมีบางส่วนที่แตกต่างไปจากการแบ่งประเภทของคำตามหลักภาษาศาสตร์ภาษาไทยไปบ้าง โดยการแบ่งประเภทของคำจะแบ่งได้ดังนี้

2.2.1 คำนาม คำนามเป็นคำที่ใช้เรียก คน สัตว์ สิ่งของ

2.2.2 คำสรรพนาม คำสรรพนามเป็นคำที่ใช้แทนคำนามและอาจแทนด้วยคำบ่งชี้ที่มีความหมายรวมถึงคำนามที่ละไว้ด้วย

2.2.3 คำนามเฉพาะ คำนามเฉพาะเป็นคำที่เป็นชื่อเฉพาะมีความหมายหรือไม่มี ความหมายก็ได้ซึ่งบางครั้งอาจมีความหมายในตัว เช่น สมศักดิ์ หมายถึง ชื่อ คน เป็นต้น

2.2.4 คำวิเศษณ์ คำวิเศษณ์เป็นคำขยายกริยาประเภทหนึ่งที่ใช้บอกว่าเหตุการณ์หรือสิ่ง ที่เกิดขึ้นนั้นเกิดขึ้นเมื่อไร ที่ไหน หรืออย่างไร

2.2.5 คำกริยา คำกริยาคือคำที่แสดงกริยาอาการหรือสภาพของประธานในประโยคและบาง ครั้งยังทำหน้าที่เป็นคำหลักของประโยคด้วย

2.2.6 คำกริยาช่วย คำกริยาช่วยเป็นคำที่เกิดร่วมกับคำกริยาเพื่อแสดง กาล มาลา วาก ลักษณ์การ รวมทั้งเจตนาของประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.7 คำคุณศัพท์

2.2.8 คำลักษณะนาม คำลักษณะนามเป็นคำที่ทำหน้าที่ประกอบนามอื่นเพื่อแสดงรูป
ลักษณะ ขนาด หรือประมาณของนามนั้นให้ชัดเจนยิ่งขึ้น

2.2.9 คำกำหนด

2.2.10 คำของ

2.2.11 คำเชื่อม คำเชื่อมเป็นคำที่ทำหน้าที่เชื่อมประโยคหนึ่งเข้าอีกประโยคหนึ่ง หรือเชื่อม
ประโยคเข้ากับวลี เป็นต้น

2.2.12 คำนับ

2.2.13 คำปฏิเสธ

2.2.14 คำลงท้าย

2.2.15 คำบุพบท คำบุพบทคือคำที่นำหน้าคำนามหรือคำสรรพนามเพื่อบอกตำแหน่ง
สถานที่ และความสัมพันธ์ระหว่างคำกริยาและคำนามหรือคำสรรพนามที่ตามมา

2.2.16 คำแสดงคำถาม



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 โครงสร้างของประโยคภาษาไทย

ประโยคภาษาไทยประกอบไปด้วยหน่วยคำที่ถูกนำมาเรียงโดยให้มีความสัมพันธ์กัน โดยพื้นฐานของภาษาไทยจะพบว่าคำประกอบขึ้นด้วยกลุ่มตัวอักษร และคำเหล่านี้จะผสมกันเป็นกลุ่มของวลี กลุ่มของวลีจะมีความสัมพันธ์กันเป็นโครงสร้างของประโยคภาษาไทย ดังนั้นการทำคำและวลีมาเขียนเรียงติดกันให้เป็นประโยคภาษาไทย จึงต้องมีไวยากรณ์เป็นหลักในการกำหนดตำแหน่งของคำและวลีที่จะนำมาเรียงกัน ข้อกำหนดนี้เรียกว่า “กฎไวยากรณ์” และขบวนการใช้กฎไวยากรณ์พิจารณาประโยคนี้เรียกว่า “การวิเคราะห์โครงสร้างประโยคภาษาไทย”

2.3.1 กฎไวยากรณ์ของนามวลีภาษาไทย

นามวลีในประโยคภาษาไทยเป็นวลีที่มีคำนามเป็นแกนหลักของกลุ่มคำ โดยจะมีโครงสร้างเป็นคำนาม คำสรรพนาม และ คำนามเฉพาะ ดังนี้

2.2.1.1 โครงสร้างนามวลีภาษาไทยที่ 1 ได้แก่ นามวลีที่ประกอบด้วยคำนาม

นามวลี → คำนาม + (คำขยายนาม)

2.2.1.2 โครงสร้างนามวลีภาษาไทยที่ 1 ได้แก่ นามวลีที่ประกอบด้วยคำสรรพนาม

นามวลี → คำสรรพนาม + (คำขยายนาม)

2.2.1.3 โครงสร้างนามวลีภาษาไทยที่ 1 ได้แก่ นามวลีที่ประกอบด้วยคำนามเฉพาะ

นามวลี → คำนามเฉพาะ

หมายเหตุ คำขยายนามจะหรือไม่มีก็ได้

2.3.2 กฎไวยากรณ์ของกริยาวลีภาษาไทย

กริยาวลีในประโยคภาษาไทยเป็นวลีที่มีคำกริยาเป็นแกนหลักของกลุ่มคำ โดยจะมีโครงสร้างเป็นคำกริยาเพียงคำเดียว หรือคำกริยาประกอบกับส่วนขยาย ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2.1 โครงสร้างกริยาวลีภาษาไทยแบบที่ 1 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา คำนาม และ บุรพบทวลี

กริยาวลี → คำกริยา + คำนาม + บุรพบทวลี

2.3.2.1 โครงสร้างกริยาวลีภาษาไทยแบบที่ 2 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา คำนาม และ คำนาม

กริยาวลี → คำกริยา + คำนาม + คำนาม

2.3.2.1 โครงสร้างกริยาวลีภาษาไทยแบบที่ 3 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา และ คำนาม

กริยาวลี → คำกริยา + คำนาม

2.3.2.4 โครงสร้างกริยาวลีภาษาไทยแบบที่ 4 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา คำนาม และ คำวิเศษณ์

กริยาวลี → คำกริยา + คำนาม + คำวิเศษณ์

2.3.2.5 โครงสร้างกริยาวลีภาษาไทยแบบที่ 5 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา คำนาม และ คำช่วยกริยา

กริยาวลี → คำกริยา + คำนาม + คำช่วยกริยา

2.3.2.6 โครงสร้างกริยาวลีภาษาไทยแบบที่ 6 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยา คำบุรพบท และ คำนาม

กริยาวลี → คำกริยา + คำบุรพบท + คำนาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2.1 โครงสร้างกริยาวลีภาษาไทยแบบที่ 7 ได้แก่ กริยาวลีที่ประกอบด้วย คำกริยาเพียงอย่างเดียว



2.3.3 กฎไวยากรณ์ของนุรพบทวลีภาษาไทย

นุรพบทวลีในประโยคภาษาไทยเป็นวลีที่มีคำนุรพบทเป็นแกนหลักของกลุ่มคำ โดยจะมีโครงสร้างเป็นคำนุรพบท และ คำนาม

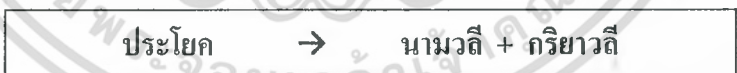
2.3.3.1 โครงสร้างนุรพบทวลีภาษาไทยที่ 1 ได้แก่ นุรพบทวลีที่ประกอบด้วยคำนุรพบท และ คำนาม



2.3.4 กฎไวยากรณ์ประโยคภาษาไทย

โครงสร้างของประโยคภาษาไทยประกอบด้วยนามวลีภาษาไทย และกริยาวลีภาษาไทยซึ่งงานวิจัยนี้ได้แบ่งโครงสร้างของประโยคภาษาไทยออกได้ 1 ประเภทดังนี้

2.3.4.1 โครงสร้างของประโยคภาษาไทยที่ 1 ได้แก่ ประโยคที่ประกอบด้วยนามวลีภาษาไทยและกริยาวลีภาษาไทย



บทที่ 3

การจัดเก็บฐานข้อมูลคำศัพท์

3.1 คำนำ

ในขั้นตอนการจัดเก็บข้อมูลลงในพจนานุกรมนั้น เราจำเป็นต้องจัดเก็บให้มีความเป็นระเบียบ ง่ายแก่การค้นหา และที่สำคัญเนื้อที่ที่ใช้ในการจัดเก็บต้องมีการจัดเก็บที่มีประสิทธิภาพที่ดี เพื่อเป็นการประหยัดเนื้อที่ในการจัดเก็บ ดังนั้นในบทนี้จะกล่าวถึงการจัดเก็บข้อมูลลงในพจนานุกรม และถ้าเราทราบถึงวิธีการจัดเก็บเราก็สามารถที่จะทำการเพิ่ม ลบ หรือแก้ไขก็ไม่ใช้เรื่องยากอีกต่อไป ซึ่งการจัดเก็บก็มีอยู่หลายวิธี แต่ละวิธีก็มีข้อดีและข้อเสียต่างกันออกไป ซึ่งในงานวิจัยนี้จะใช้วิธีการจัดเก็บแบบ เรียงลำดับ (Sequential) โดยก่อนที่จะทำการจัดเก็บจะมีการทำการเรียงลำดับข้อมูลก่อน โดยจะเรียงจากมากไปหาน้อยตามลำดับ โดยจะอธิบายรายละเอียดต่อไปในหัวข้อต่อไป

3.2 การจัดเก็บข้อมูลลงพจนานุกรม

ในการจัดเก็บข้อมูลลงพจนานุกรมนั้นในข้อมูล 1 ชุดที่จะทำการจัดเก็บจะประกอบไปด้วยส่วนประกอบที่สำคัญ 2 ส่วน ได้แก่

1. คำศัพท์
2. ประเภทของคำศัพท์

โดยการจัดเก็บข้อมูลลงในพจนานุกรมนั้นจะทำการจัดเก็บข้อมูลเป็นชุด โดยแต่ละชุดจะแยกออกจากกันโดยใช้ ASCII Code 0Dh และ 0Ah ตามลำดับ และจะทำการแบ่งพจนานุกรมออกเป็นพจนานุกรมย่อยๆ หลายพจนานุกรม โดยจะพิจารณาคำศัพท์ตามข้อกำหนดดังต่อไปนี้

1. จะต้องหา Set ของข้อมูลที่จะทำการบันทึก แล้วทำการแยก SubSet ดังกล่าวออกเป็นกลุ่มๆ
2. ในแต่ละ SubSet จะพิจารณาตัวอักษรตัวแรกสุดของคำศัพท์
3. นำเอา SubSet ที่แยกออกเป็นกลุ่มแล้วนำมาสร้างเป็นพจนานุกรมย่อยๆ ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นเราจึงสามารถแยกคำศัพท์ออกได้ตามตัวอักษรตัวแรกที่สามารถปรากฏต้นคำศัพท์ได้
ซึ่งจะแบ่งได้ออกได้ดังต่อไปนี้

พยัญชนะ

ก ข ค ฅ ง จ ฉ ช ซ ฌ ญ
ฎ ฐ ฒ ณ ด ต ถ ท ธ น บ
ป ผ ฝ พ ฟ ภ ม ย ร ฤ ล
ว ศ ส ห อ ฮ

สระ

เ แ ใ โ ใ

โดยแต่ละตัวอักษรก็จะมีพจนานุกรมเป็นของตัวเองโดยมีรายละเอียดดังต่อไปนี้

อักษร	เพิ่มข้อมูล	อักษร	เพิ่มข้อมูล	อักษร	เพิ่มข้อมูล
ก	Kai.dic	ช	Tong.dic	ศ	Sala.dic
ข	Kao.dic	ฅ	Hnoo.dic	ส	Tiger.dic
ค	Kind.dic	บ	Baimai.dic	ห	Heep.dic
ฅ	*.dic	ป	Pla.dic	อ	Ang.dic
ง	Ngoo.dic	ผ	Ph.dic	ฮ	*.dic
จ	Jan.dic	ฝ	*.dic	เ	A.dic
ฉ	Ching.dic	พ	Pan.dic	แ	Air.dic
ช	Chang.dic	ฟ	Fun.dic	โ	O.dic
ซ	So.dic	ภ	Sumphao.dic	ใ	Maun.dic
ฌ	Ying.dic	ม	Man.dic	ใ	Malai.dic
ญ	Than.dic	ย	Yuk.dic		
ฎ	Dek.dic	ร	Ruer.dic		
ฐ	Tao.dic	ฤ	*.dic		
ฒ	Toong.dic	ล	Ling.dic		
ณ	Tahan.dic	ว	Hvan.dic		

ตารางที่ 3.1 แสดงความสัมพันธ์ระหว่างตัวอักษรและเพิ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับใช้ในการศึกษาเท่านั้น เมื่ออนุญาตให้ผู้อื่นใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ชนิดของคำศัพท์

ชนิดของคำคือการจำแนกหมวดหมู่ของคำให้อยู่เป็นกลุ่ม โดยมีวัตถุประสงค์เพื่อจะทำให้ทราบว่าคำศัพท์คำนั้นทำหน้าที่เป็นอะไรในประโยค ซึ่งหน้าที่ของคำศัพท์นั้นได้ทำการอธิบายไปแล้วในบทที่ 2 แต่เนื่องจากคำศัพท์บางคำทำหน้าที่ได้หลายหน้าที่ดังนั้นในการจัดเก็บข้อมูลลงในพจนานุกรมอย่างได้ให้มีประสิทธิภาพที่ดีนั้นก็เป็นส่วนประกอบที่สำคัญส่วนหนึ่งที่ว่าได้ดังที่กำลังจะกล่าวถึงดังต่อไปนี้

เนื่องจากการจำเก็บหน้าที่ของคำศัพท์ในงานวิจัยชิ้นนี้ต้องการจัดเก็บให้มีประสิทธิภาพมากที่สุดที่สามารถทำได้ดังนั้นจึงได้นำหน้าการของเลขฐานมาประยุกต์ใช้งานกล่าวคือหน้าที่แต่ละหน้าที่ก็จะมีเลขหลักประจำตัวของหน้าที่นั้นๆ ถ้าคำศัพท์คำใดมีหน้าที่มากกว่า 1 หน้าที่ก็จะนำเอาเลขประจำหน้าที่ที่คำศัพท์คำนั้นสามารถเป็นไปได้ โดยได้กำหนดเลขประจำหลักไว้ดังต่อไปนี้

หน้าที่ของคำศัพท์	เลขประจำหลักฐานสอง	เลขประจำหลักฐานสิบหก
คำนาม	2^0	1
คำสรรพนาม	2^1	2
คำนามเฉพาะ	2^2	4
คำวิเศษณ์	2^3	8
คำกริยา	2^4	10
คำกริยาช่วย	2^5	20
คำคุณศัพท์	2^6	40
คำลักษณนาม	2^7	80
คำกำหนด	2^8	100
คำของ	2^9	200
คำเชื่อม	2^{10}	400
คำนับ	2^{11}	800
คำปฏิเสธ	2^{12}	1000
คำลงท้าย	2^{13}	2000
คำบุพบท	2^{14}	4000
คำแสดงคำถาม	2^{15}	8000

ตารางที่ 3.2 แสดงความสัมพันธ์ระหว่างหน้าที่ของคำกับค่าประจำหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก้อน	→	คำนาม และ คำลักษณนาม
	→	$1 + 128 = 129$
กัน	→	คำสรรพนาม , คำวิเศษณ์ และ คำกริยา
	→	$2 + 8 + 16 = 26$
คลุม	→	คำกริยา และ คำวิเศษณ์
	→	$16 + 8 = 24$
กัน	→	คำนาม , คำกริยา และ คำลักษณนาม
	→	$1 + 16 + 128 = 145$
มัน	→	คำนาม , คำสรรพนาม และ คำกริยา
	→	$1 + 2 + 16 = 19$
ฉบับ	→	คำนาม และ คำลักษณนาม
	→	$1 + 128 = 129$
ส่วน	→	คำนาม และ คำเชื่อม
	→	$1 + 1024 = 1025$

รูปที่ 3.1 ตัวอย่างหน้าที่ของคำศัพท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทราบถึงวิธีการจัดเก็บหน้าที่ของคำแล้ว ต่อไปก็จะเป็นขั้นตอนในการตรวจสอบว่าคำศัพท์คำนั้นมีหน้าที่ที่ต้องการหรือไม่ ซึ่งวิธีหนึ่งในการตรวจสอบก็คือ การนำหน้าที่ของคำศัพท์ที่ต้องการมาทำการ And กับหน้าที่ของคำที่ต้องการตรวจสอบถ้าได้ค่าเท่ากับหน้าที่ของคำก็แสดงว่าคำศัพท์คำนั้นทำหน้าที่ตามที่เราทำการทดสอบเช่น

ก่อน หน้าที่ของคำ 129 ต้องการทดสอบว่าเป็นคำลักษณะนามหรือไม่

ขั้นตอนที่ 1 ทำการแปลงเลขหน้าที่ของคำเป็นเลขฐานสอง

$$129 = 10000001$$

ขั้นตอนที่ 2 ทำการแปลงหน้าที่ของคำที่ต้องการทดสอบ

$$128 = 10000000$$

ขั้นตอนที่ 3 นำเอาหน้าที่ของคำศัพท์กับหน้าที่ของคำที่ต้องการทดสอบนำมา And กัน

$$\begin{array}{r} 10000001 \\ 10000000 \\ \hline 10000000 \end{array} ; \text{And}$$

ขั้นตอนที่ 4 พิจารณาว่าผลลัพธ์ที่ได้มีค่าเท่ากับหน้าที่ของคำที่ทำการทดสอบหรือไม่ถ้าเท่าแสดงว่าคำศัพท์คำนั้นทำหน้าที่ตามที่เราทดสอบ

สิ่งที่กล่าวมาข้างต้นซึ่งไม่ใช่เรื่องยากแต่ถ้าต้องมานั่งทำด้วยมือคงจะไม่สะดวกมากนัก ดังนั้นเพื่อให้ง่ายแก่การเพิ่มคำศัพท์ ลบคำศัพท์ และ แก้ไขคำศัพท์ จึงได้เขียนโปรแกรมที่ช่วยให้การขูดข้อมูล เปลี่ยนแปลงแก้ไขข้อมูล และ เพิ่มข้อมูล ให้ง่ายขึ้น โดยมีการใช้งานดังนี้

บทที่ 4

การแยกแยะหน่วยคำออกจากประโยคและการตรวจสอบไวยากรณ์

4.1 บทนำ

ประโยคในภาษาไทยเป็นโครงสร้างที่แตกต่างจากภาษาอื่นๆ กล่าวคือในประโยคจะประกอบด้วยคำย่อยๆหลายคำโดยแต่ละคำจะเขียนติดกัน ดังนั้นจึงเป็นการยากที่จะวิเคราะห์ว่าคำใดมีการแยกคำออกจากกันอย่างไร เช่นประโยค “เรือโคลงเพราะโคลง” จะสามารถแยกได้

“เรือ_โคลง_เพราะ_โคลง”

“เรือ_โคลง_เพราะ_โค_ลง”

“เรือ_โค_ลง_เพราะ_โคลง”

“เรือ_โค_ลง_เพราะ_โค_ลง”

แต่คำที่ถูกต้องจะมีอยู่เพียงคำเดียวเท่านั้น ได้แก่คำ “เรือ_โคลง_เพราะ_โค_ลง” ซึ่งการที่จะวิเคราะห์ประโยคที่แยกได้นั้น ประโยคใดถูกต้องก็ต้องอาศัยความรู้ทางด้านกฎไวยากรณ์ของภาษาเป็นตัวที่ทำหน้าที่วิเคราะห์ประโยค

ดังนั้นงานวิจัยชิ้นนี้จึงได้ทำการแบ่งขั้นตอนในการวิเคราะห์ออกเป็นสองส่วนด้วยกันดังนี้

1. การแยกคำย่อยออกจากประโยค
2. การวิเคราะห์และตรวจสอบไวยากรณ์

การทำงานของทั้งสองส่วนจะทำงานสัมพันธ์กันตั้งแต่ต้นจนจบ ดังรูปที่ 4.1

4.2 การแยกหน่วยคำออกจากประโยค

ในการแยกหน่วยคำออกจากประโยคนั้นมีการทำงานได้หลายวิธีแต่ละวิธีก็จะมีข้อดีข้อเสียต่างกันออกไป ซึ่งงานวิจัยชิ้นนี้จะแบ่งการทำงานในส่วนของการแยกหน่วยคำออกจากประโยคโดยมีหลักการทำงานดังต่อไปนี้

1. รับประโยคที่ต้องการตรวจสอบ
2. วิเคราะห์ว่าจะต้องเปิดพจนานุกรมตัวใดโดยการนำเอาตัวอักษรหรือสระตัวแรกของประโยคมาเป็นตัวพิจารณา
3. เมื่อทราบว่าจะต้องเปิดพจนานุกรมตัวใดแล้วก็ทำการเปิดพจนานุกรม
4. นำประโยคที่ต้องการตรวจสอบเปรียบเทียบกับข้อมูลในพจนานุกรม
5. ถ้าไม่พบคำศัพท์ที่เหมือนกันให้เพิ่มตัวชี้ที่ชี้ไปยังข้อมูลในพจนานุกรมชิ้น 1 แล้วกลับไปทำขั้นตอนที่ 4 ใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. ถ้าพบก็จะทำการตัดประโยคที่ค้นพบโดยจะตัดความยาวเท่ากับความยาวของคำศัพท์
7. นำประโยคที่เหลือจากการตัดประโยคกลับไปทำขั้นตอนที่ 2 ใหม่จนกระทั่งหมดประโยค
8. เมื่อตัดหมดประโยคแล้ว ก็จะเก็บคำศัพท์ที่สามารถตัดออกได้เป็นคำศัพท์ย่อยที่เป็นไปได้ในประโยค แต่คำในประโยคบางประโยคสามารถตัดได้มากกว่า 1 คำ ดังนั้นจึงต้องทำการตัดประโยคใหม่จนกระทั่งไม่สามารถตัดได้โดยการเพิ่มตัวชี้ข้อมูลในพจนานุกรมขึ้น 1
9. เมื่อได้จนไม่สามารถได้ได้แล้วก็จะได้ประโยคที่เป็นไปได้ทั้งหมด

ตัวอย่างการแยกคำย่อยออกจากประโยคหลัก

- ขั้นตอนที่ 1 รับข้อความอินพุตเข้าสู่โปรแกรมแยกหน่วยคำ
“เรือ โคลงเพราะ โคลง”
- ขั้นตอนที่ 2 หาตัวอักษรหรือสระตัวแรกของประโยค
“ุ”
- ขั้นตอนที่ 3 ทำการเปิดพจนานุกรมสระ “ุ”
- ขั้นตอนที่ 4 นำประโยคที่ต้องการแยกคำกับข้อมูลในพจนานุกรม
เรือโคลงเพราะโคลง เปรียบเทียบกับ เหลว
- ขั้นตอนที่ 5 ถ้าตัวชี้ข้อมูลไม่ได้ชี้ไปที่คำศัพท์คำว่า เรือ ให้เพิ่มค่าตัวชี้ข้อมูลในพจนานุกรมขึ้น 1 แล้วกลับไปทำขั้นตอนที่ 4 ใหม่ จนกระทั่งตัวชี้ชี้ไปที่คำว่า เรือ
- ขั้นตอนที่ 6 ตัดประโยคที่ต้องการแยกคำศัพท์ออกโดยการเปรียบเทียบความยาวของคำศัพท์กับความยาวของประโยค
- | | | |
|--------------------|---------|----|
| เรือโคลงเพราะ โคลง | ความยาว | 17 |
| เรือ | ความยาว | 4 |
- ดังนั้นเมื่อตัดแล้วจะเหลือ โคลงเพราะ โคลง
- ขั้นตอนที่ 7 นำประโยคที่เหลือไปทำขั้นตอนที่ 2 ใหม่
- ขั้นตอนที่ 2 หาตัวอักษรหรือสระตัวแรกของประโยค
“โ”
- ขั้นตอนที่ 3 ทำการเปิดพจนานุกรมสระ “โ”

เอกสารนี้เป็นลิขสิทธิ์ของสำนักหอสมุดกลางพระจอมเกล้าลาดกระบัง หากท่านนำข้อมูลไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลง เปรียบเทียบกับ ลาก

ขั้นตอนที่5 ถ้าตัวชี้ข้อมูลไม่ได้ชี้ไปที่คำศัพท์คำว่า ลง เพิ่มค่าตัวชี้ข้อมูลในพจนานุกรมขั้น
1 แล้วกลับไปทำขั้นตอนที่4ใหม่ จนกระทั่งตัวชี้ชี้ไปที่คำว่า ลง

ขั้นตอนที่6 ตัดประโยคที่ต้องการแยกคำศัพท์ออก โดยการเปรียบเทียบความยาวของคำศัพท์
กับความยาวของประโยค

ลง ความยาว 2

ลง ความยาว 2

ดังนั้นเมื่อตัดแล้วก็จะหมดข้อมูลทำการตัด

ขั้นตอนที่7 เนื่องจากประโยคที่ใช้ทำการหมดลงแล้วดังนั้นจึงเก็บคำศัพท์ที่สามารถตัดออก
ได้เป็นชุดคำศัพท์ชุดที่ 2

เรือ_โคลง_เพราะ_โค_ลง เป็นคำศัพท์ที่สามารถแยกออกได้ชุดที่ 2

ขั้นตอนที่8 ทำการเลื่อนตัวชี้ข้อมูลในพจนานุกรมขั้น 1 แล้วกลับไปทำขั้นตอนที่ 2 ใหม่
โดยประโยคที่ต้องการตัดจะใช้ตัวเดิม

ขั้นตอนที่9 เนื่องจากตัวชี้ข้อมูลในพจนานุกรมชี้คำศัพท์ที่เปรียบเทียบแล้วไม่พบคำศัพท์
ที่เหมือนกันอีกดังนั้นจึงต้องลดระดับของประโยคที่ต้องการตัดลงอีก 1 ชั้น
ซึ่งจะได้คำศัพท์ เพราะโคลง แล้วกลับไปทำขั้นตอนที่ 2 ใหม่ แต่เนื่อง
จากตัวชี้ข้อมูลไม่สามารถหาคำที่เหมือนได้จึงต้องลดชั้นระดับลงอีก 1
ระดับ ซึ่งจะได้คำศัพท์ โคลงเพราะโคลง แล้วจึงกลับไปทำขั้นตอนที่2 ใหม่

ขั้นตอนที่2 หาตัวอักษรหรือสระตัวแรกของประโยค
“โ”

ขั้นตอนที่3 ทำการเปิดพจนานุกรมสระ “โ”

ขั้นตอนที่4 นำประโยคที่ต้องการแยกคำกับข้อมูลในพจนานุกรม

โคลงเพราะโคลง เปรียบเทียบกับ โค

ขั้นตอนที่5 ถ้าตัวชี้ข้อมูลไม่ได้ชี้ไปที่คำศัพท์คำว่า โค เพิ่มค่าตัวชี้ข้อมูลในพจนานุกรมขั้น
1 แล้วกลับไปทำขั้นตอนที่4ใหม่ จนกระทั่งตัวชี้ชี้ไปที่คำว่า โค

ขั้นตอนที่ 4 นำประโยคที่ต้องการแยกคำกับข้อมูลในพจนานุกรม

โคลง เปรียบเทียบกับ โค

ขั้นตอนที่ 5 ถ้าตัวชี้ข้อมูลไม่ได้ชี้ไปที่คำศัพท์คำว่า โค เพิ่มค่าตัวชี้ข้อมูลในพจนานุกรมขึ้น
1 แล้วกลับไปทำขั้นตอนที่ 4 ใหม่ จนกระทั่งตัวชี้ชี้ไปที่คำว่า โค

ขั้นตอนที่ 6 ตัดประโยคที่ต้องการแยกคำศัพท์ออกโดยการเปรียบเทียบความยาวของคำศัพท์
กับความยาวของประโยค

โคลง	ความยาว	4
------	---------	---

โค	ความยาว	2
----	---------	---

ดังนั้นเมื่อตัดแล้วจะเหลือ ลง

ขั้นตอนที่ 7 นำประโยคที่เหลือไปทำขั้นตอนที่ 2 ใหม่

ขั้นตอนที่ 2 หาตัวอักษรหรือสระตัวแรกของประโยค

“ล”

ขั้นตอนที่ 3 ทำการเปิดพจนานุกรมสระ “ล”

ขั้นตอนที่ 4 นำประโยคที่ต้องการแยกคำกับข้อมูลในพจนานุกรม

ลง เปรียบเทียบกับ ลาก

ขั้นตอนที่ 5 ถ้าตัวชี้ข้อมูลไม่ได้ชี้ไปที่คำศัพท์คำว่า ลง เพิ่มค่าตัวชี้ข้อมูลในพจนานุกรมขึ้น
1 แล้วกลับไปทำขั้นตอนที่ 4 ใหม่ จนกระทั่งตัวชี้ชี้ไปที่คำว่า ลง

ขั้นตอนที่ 6 ตัดประโยคที่ต้องการแยกคำศัพท์ออกโดยการเปรียบเทียบความยาวของคำศัพท์
กับความยาวของประโยค

ลง	ความยาว	2
----	---------	---

ลง	ความยาว	2
----	---------	---

ดังนั้นเมื่อตัดแล้วก็จะหมดข้อมูลทำการตัด

ขั้นตอนที่ 7 เนื่องจากประโยคที่ใช้ทำการหมดลงแล้วดังนั้นจึงเก็บคำศัพท์ที่สามารถตัดออก
ได้เป็นชุดคำศัพท์ชุดที่ 4

เรือ_โค_ลง_เพราะ_โค_ลง เป็นคำศัพท์ที่สามารถแยกออกได้ชุดที่ 4

ขั้นตอนที่ 8 ทำการเลื่อนตัวชี้ข้อมูลในพจนานุกรมขึ้น 1 แล้วกลับไปทำขั้นตอนที่ 2 ใหม่
โดยประโยคที่ต้องการตัดจะใช้ตัวเดิมแต่เนื่องจากข้อมูลที่ใช้ในการพิจารณา

ลดลงเป็นลำดับขั้นแล้วก็ยังไม่มีความสามารถนำมาแยกประ โยคได้อีกคั้ง
นั้นจึงจบการทำงาน

ดังนั้นเมื่อทำงานตามขั้นตอนต่างๆจนจบการทำงานจะได้เขตของประ โยคที่
แยกออกได้คั้งต่างที่ 4.1

เขตคำตอบ / หน่วยคำ	1	2	3	4	5	6
1	เรือ	โคลง	เพราะ	โคลง		
2	เรือ	โคลง	เพราะ	โค	ลง	
3	เรือ	โค	ลง	เพราะ	โคลง	
4	เรือ	โค	ลง	เพราะ	โค	ลง

ตารางที่ 4.1 แสดงผลลัพธ์ที่ได้จากตัวอย่างการแยกหน่วยคำออกจากประ โยค



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5 ผลการทดลอง

5.1 บทนำ

หลังจากที่ได้วิธีการทำงานข้างต้นแล้วนั้นก็ได้นำ วิธีการดังกล่าวมาพัฒนาเป็น โปรแกรมคอมพิวเตอร์ทั้งในส่วนของการแยกแยะหน่วยคำออกจากประโยคและย่อหน้า และในส่วนของการแก้ไขเพิ่มเติมคำศัพท์ในพจนานุกรม

5.2 การทดลอง

โดยพัฒนาโปรแกรม ด้วยภาษา Visual Basic ซึ่งมีการทำงานได้ดังนี้ การทำงานของโปรแกรมจะสามารถทำงานได้โดยการ Run โปรแกรม Cutword.exe จะปรากฏหน้าต่างจอคอมพิวเตอร์ดังรูปที่ 5.1



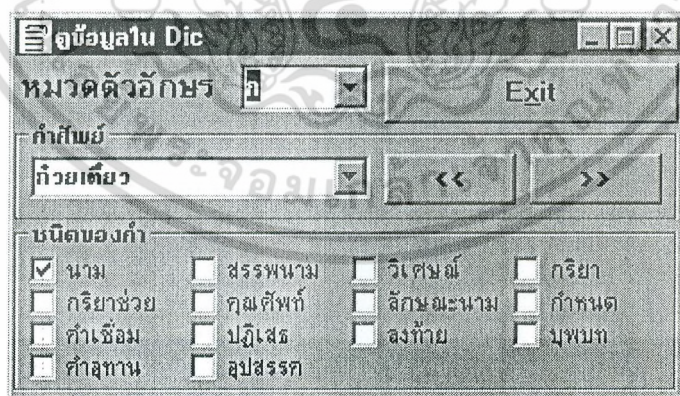
รูปที่ 5.1 แสดงถึงหน้าจอหลักของโปรแกรม

รายการ Dic จะประกอบไปด้วยรายการย่อย Open Dic กับ Modify Dic ดังรูปที่ 5.2



รูปที่ 5.2 แสดงถึงหน้าจอรายการย่อยของรายการ Dic ของโปรแกรม

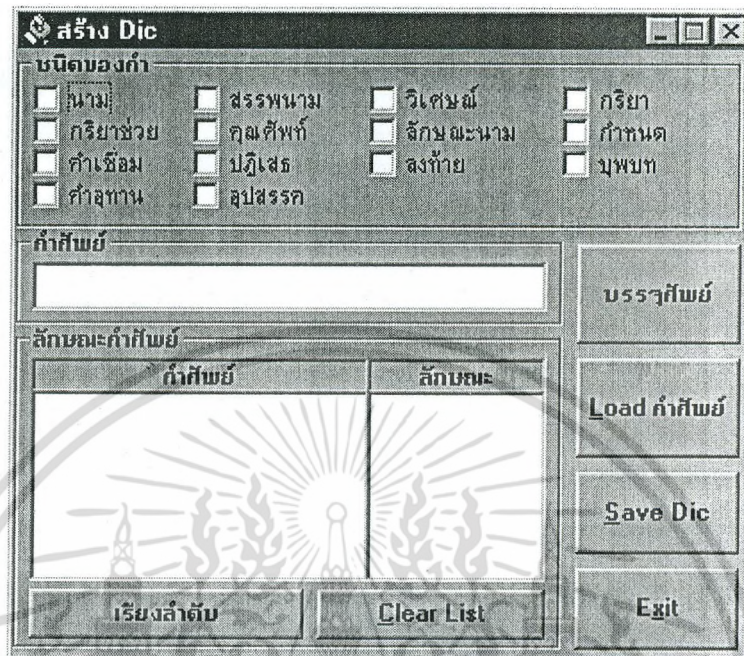
รายการ Open Dic เป็นรายการสำหรับดูคำศัพท์ใหม่ใน Dictionary โดยมีหน้าจอดังรูปที่ 5.3



รูปที่ 5.3 แสดงถึงหน้าจอแสดงรายการคำศัพท์ใน Dictionary

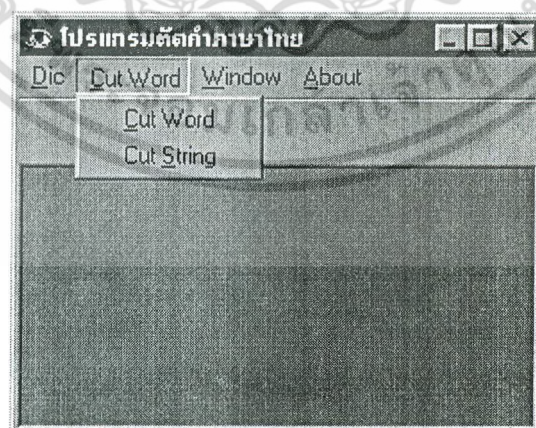
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายการ **Modify Dic** เป็นรายการสำหรับแก้ไขที่มีอยู่หรือสร้างคำศัพท์ใหม่ใน Dictionary โดยมีหน้าจอ ดังรูปที่ 5.4



รูปที่ 5.4 แสดงถึงหน้าจอการแก้ไขเพิ่มคำศัพท์ใน Dictionary

รายการ **Cut word** เป็นรายการที่ประมวลผลประโยคและย่อหน้าโดยแบ่งออกเป็น 2 รายการดังรูปที่ 5.5



รูปที่ 5.5 แสดงถึงหน้าจอรายการย่อยในรายการ **Cut Word**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

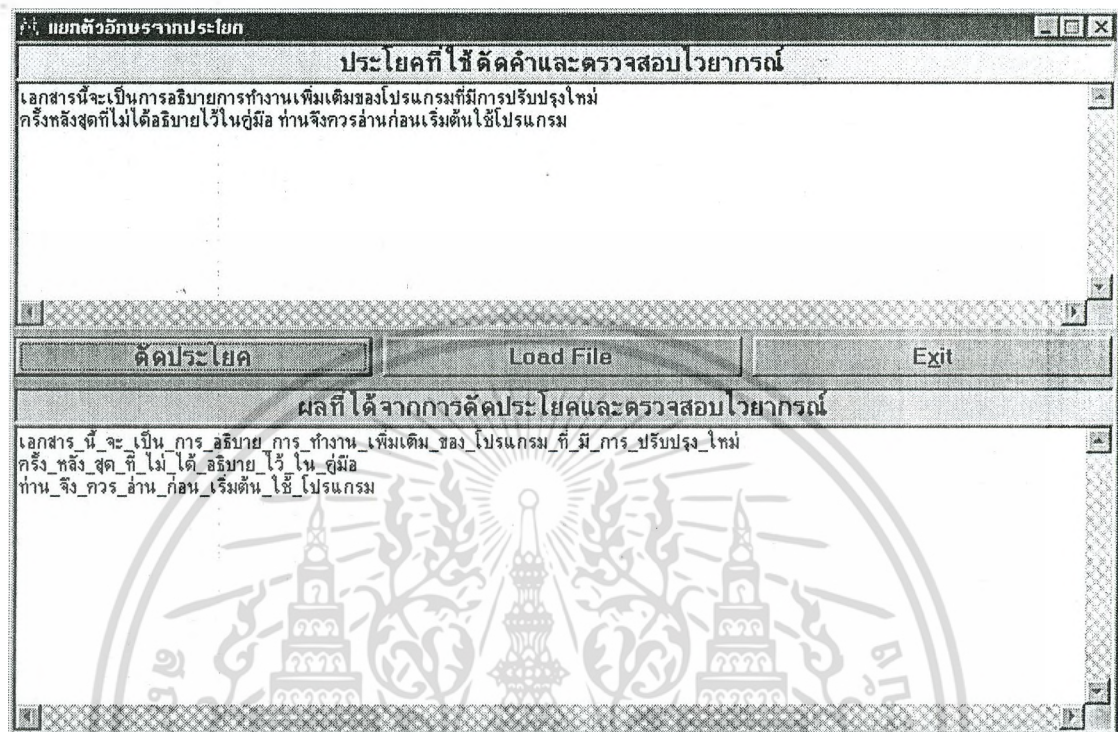
รายการย่อย **Cut Word** จะเป็นรายการที่แสดงถึงผลลัพธ์ของการแยกแยะหน่วยคำและการตรวจสอบทางไวยากรณ์ในประโยคที่ป้อนหรือเปิดจาก File ที่เตรียมไว้ โดยอาศัยวิธีการข้างต้นเป็นการประเมินผล ซึ่งมีหน้าจอดังรูปที่ 5.6

ทดสอบตัดคำและตรวจไวยากรณ์		
ประโยคที่ใช้ตัดคำ : ทลานมารกรามปู่		
ผลลัพธ์ที่ได้จากการหน่วยคำ (คำศัพท์)	ผลลัพธ์ที่ได้จากการแยกหน่วยคำ (ชนิดคำ)	
ทลาน_มาร_อก_ราม_ปู่ ทลาน_มา_รอก_ราม_ปู่ ทลาน_มา_รอ_กราม_ปู่	1_1_1_20_1 1_18_1_20_1 1_18_8_8_1	
ผลลัพธ์ที่ได้จากการวิเคราะห์ไวยากรณ์ได้แก่		
ทลาน_มา_รอ_กราม_ปู่		
แยกหน่วยคำย่อย	ตรวจสอบไวยากรณ์	ยกเลิกโปรแกรม

รูปที่ 5.6 แสดงถึงการทำงานของรายการย่อย Cut word

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายการ **Cut String** จะเป็นรายการที่แสดงถึงผลลัพธ์ของการแยกแยะหน่วยคำในย่อหน้าที่ป้อนหรือเปิดจาก File ที่เตรียมไว้ ซึ่งมีหน้าจอ ดังรูปที่ 5.7



รูปที่ 5.7 แสดงถึงการทำงานของรายการย่อย Cut String

การเก็บคำศัพท์จะมีการเก็บคำศัพท์ในรูปแบบของ Text File เนื่องจากความต้องการที่จะให้มีการแก้ไขเพิ่มเติม ได้ง่ายในขณะทดลอง โดยมีรูปแบบการเก็บดังตัวอย่างต่อไปนี้

คำช	1
ก๊อกน้ำ	1
ก๊อก	41
ก๊ิก	41
ก้ำกึ่ง	8
ก้ำวเดิน	8
ก้ำวหน้า	8
ก้ำวร้าย	8

โดยคำแรกเป็นคำศัพท์ ส่วนตัวเอกเป็นการบอกชนิดของคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 สรุปผลการทดลอง

จากการทดลองป้อนประโยคและย่อหน้าเพื่อให้มีการแยกแยะหน่วยคำ วิธีการนี้ให้ความถูกต้องค่อนข้างสูงมาก อาจจะมีผลผิดพลาดบ้างจาก ในDictionary ไม่มีคำศัพท์ที่อยู่หรือประโยคที่ป้อนเป็นประโยคที่ใช้ในการพูดทำให้การตีค่าทางไวยากรณ์ผิดไป ซึ่งสามารถที่จะพัฒนาให้มีความถูกต้องเพิ่มขึ้นโดยการเพิ่มคำศัพท์และขยายโครงสร้างทางไวยากรณ์ขึ้น วิธีการนี้สามารถนำไปใช้ในการแยกแยะหน่วยคำออกจากประโยคก่อนการแปลภาษาด้วยคอมพิวเตอร์ได้อย่างมีประสิทธิภาพขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้