บทที่ 3

วิธีการดำเนินงานวิจัย

วิทยานิพนธ์ฉบับนี้ เป็นการศึกษาเกี่ยวกับการค้นหาคำหลักในเสียงพูดแบบไม่จำกัดโดเมน โดยไม่ขึ้นตรงกับผู้พูด ซึ่งเป็นการเสนอการสร้างโมเดลเสียงของคำหลักที่ต้องการค้นหาโดยอัตโนมัติตามรายการคำหลักในโดเมนที่ต้องการ ซึ่งรายละเอียดในบทนี้ประกอบไปด้วย

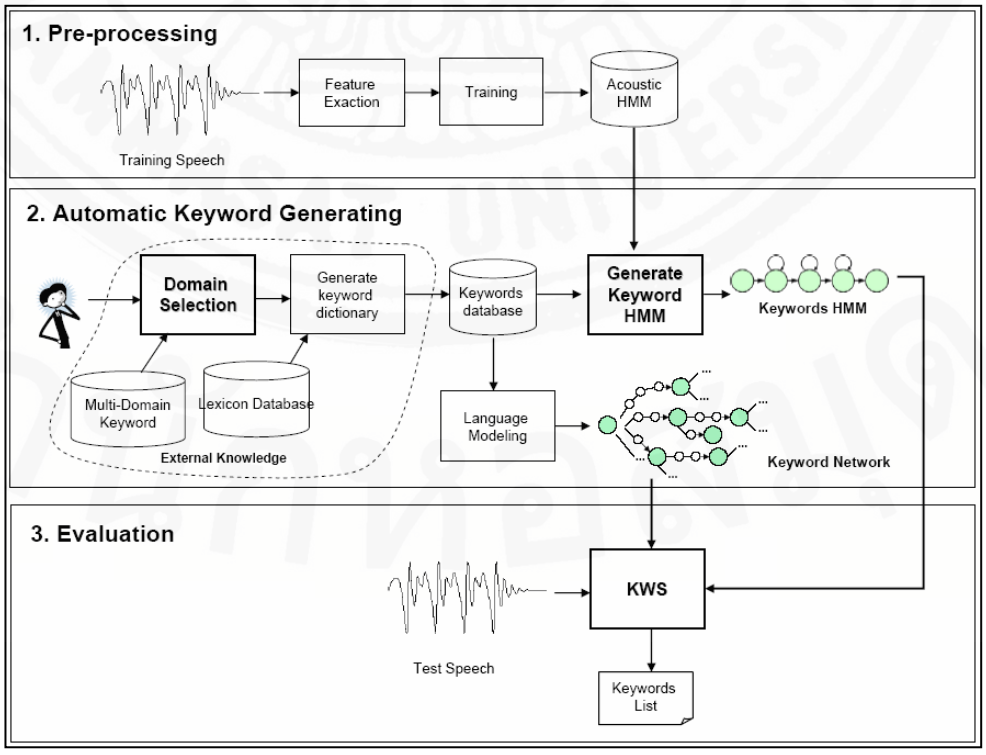
1. รายละเอียดในการค้นหาคำหลัก (Keyword Spotting)
2. การดำเนินการวิจัยในแต่ละขั้นตอน
3. เครื่องมือที่ใช้สำหรับทำวิทยานิพนธ์

สำหรับขั้นตอนในการวิจัยสามารถแบ่งออกเป็น 3 ขั้นตอนหลัก ดังภาพที่ 3.1 ได้ ดังนี้

1. ขั้นตอนการจัดเตรียมข้อมูล (Pre-processing)
2. ขั้นตอนการสร้างโมเดลคำหลักโดยอัตโนมัติ (Automatic Keyword Generating)
3. ขั้นตอนการประเมินประสิทธิภาพ (Evaluation)

ภาพที่ 3.1

ภาพรวมของการค้นหาคำหลักในเสียงพูดที่นำเสนอ



3.1 ขั้นตอนการจัดเตรียมข้อมูล (Pre-processing)

ขั้นตอนการจัดเตรียมข้อมูล มีจุดประสงค์เพื่อเตรียมแฟ้มข้อมูลต่าง ๆ ที่จำเป็นต้องใช้ก่อนที่จะทำการทดลองในขั้นตอนต่อไป ซึ่งการจัดเตรียมข้อมูลจะเป็นการขั้นตอนที่ทำเพียงครั้งเดียว โดยประกอบไปด้วยขั้นตอนย่อย ๆ ดังนี้

3.1.1 จัดเตรียมคลังข้อมูลเสียงเพื่อสร้างโมเดลเสียง

คลังข้อมูลเสียง (Speech corpus) คือ ชุดแฟ้มเสียงที่เป็นเสียงการพูดประโยคต่าง ๆ ซึ่งในวิทยานิพนธ์นี้ ได้ใช้คลังข้อมูลเสียงจาก 3 แห่งด้วยกันคือ

1. LOTUS speech corpus ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์ และคอมพิวเตอร์แห่งชาติ (NECTEC) ซึ่งเป็นคลังเสียงสำหรับพัฒนาระบบรู้จำเสียงพูดแบบต่อเนื่อง โดยเสียงที่บันทึกในคลังเสียงเป็นแบบเสียงจากอ่าน (reading speech) บทความข่าวสารทั่วไป ซึ่งเสียงที่บันทึกจะปราศจากเสียงรบกวน
2. Hotel Reservation corpus ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์ และคอมพิวเตอร์แห่งชาติ (NECTEC) แฟ้มเสียงที่บันทึกจะเป็นโดเมนของการจองห้องพักในโรงแรม ซึ่งเป็นแฟ้มเสียงที่บันทึกจะมีการพูดที่มีคำอุทาน เรียกว่า ”spontaneous speech” เช่น “**อืม** ไม่จองดีกว่าค่ะ ขอบคุณค่ะ” เป็นต้น
3. TU Course Registration corpus เป็นข้อมูลเสียงสำหรับโดเมนการลงทะเบียนนักศึกษา มหาวิทยาลัยธรรมศาสตร์ ได้มาจากการบันทึกเสียงการลงทะเบียนนักศึกษาระดับปริญญาตรี ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

นอกจากนี้ข้อมูลเสียงที่ใช้ในการฝึกฝนและทดลองงานวิจัย ยังประกอบไปด้วยเสียงพยัญชนะและสระในภาษาไทย โดยการบันทึกเสียงการออกเสียงพยัญชนะ และสระในภาษาไทย 62 หน่วยเสียง ตามตารางที่ 3.1 ถึงตารางที่ 3.2 ในการบันทึกเสียงนั้น จะทำการบันทึกเสียงด้วยโปรแกรม Sony Forge 8.0 โดยบันทึกการออกเสียงของผู้ชาย และผู้หญิง จำนวน 10 คน และให้แต่ละคนออกเสียงพยัญชนะ 44 ตัว และสระ 21 ตัว คนละ 3 ครั้ง ก่อนทำการสร้างโมเดลเสียง (Acoustic Model) โดยแต่ละแฟ้มเสียงจะมีค่าความถี่ที่ 16 KHz 16 Bit-dept MONO (ดังภาพที่ 3.2)

ตารางที่ 3.1

สัญลักษณ์แทนเสียงพยัญชนะต้นและพยัญชนะท้าย

ในภาษาไทย (26 หน่วยเสียง และ 12 หน่วยเสียง)

พยัญชนะ	หน่วยเสียง		พยัญชนะ	หน่วยเสียง	
	พยัญชนะต้น	พยัญชนะท้าย		พยัญชนะต้น	พยัญชนะท้าย
ก	k	k^	บ	b	b^
ข, ค, ฆ	kh	k^	ป	p	p^
ง	ng	ng^	ผ, พ, ภ	ph	p^
จ	c	t^	ฝ, ฟ	f	p^
ฉ, ฌ, ช	ch	t^	ม	m	m^
ซ, ศ, ษ, ส	s	t^	ร	r	n^
ญ, ย	j	j^	ล, ฬ	l	n^
ฎ, ด	d	t^	ว	w	w^
ฏ, ต	t	t^	ห, ฮ	h	-
ฐ, ฑ, ฒ,ถ, ท, ธ	th	t^	อ	z	-
ณ, น	n	n^	หน่วยเสียงทับศัพท์	br, bl, fr, fl, dr	f^, s^, h^,l^

ที่มา : ภัชริภา คชสำโรง และคณะ, 2005

ตารางที่ 3.2

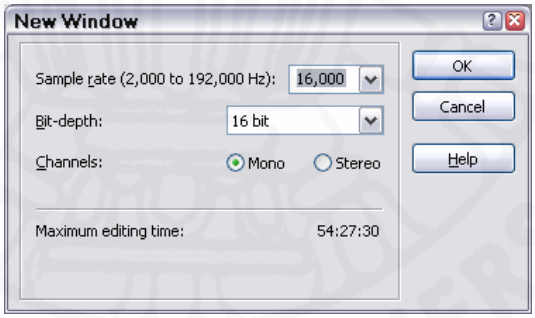
สัญลักษณ์แทนเสียงสระในภาษาไทย (24 หน่วยเสียง)

ตำแหน่งลิ้น / ความสูงของลิ้น	หน้า (สั้น/ยาว)	กลาง (สั้น/ยาว)	หลัง (สั้น/ยาว)
สูง	i, ii (อิ, อี)	v, vv (อึ, อือ)	u, uu (อุ, อู)
กลาง	e, ee (เอะ, เอ)	q, qq (เออะ, เออ)	o, oo (โอะ, โอ)
ต่ำ	x, xx (แอะ, แอ)	a, aa (อะ, อา)	@, @@ (เอาะ, ออ)
สระผสม	ia, iia (เอียะ,เอีย)	va, vva (เอือะ, เอือ)	ua, uua (อัวะ, อัว)

ที่มา : ภัชริภา คชสำโรง และคณะ, 2005

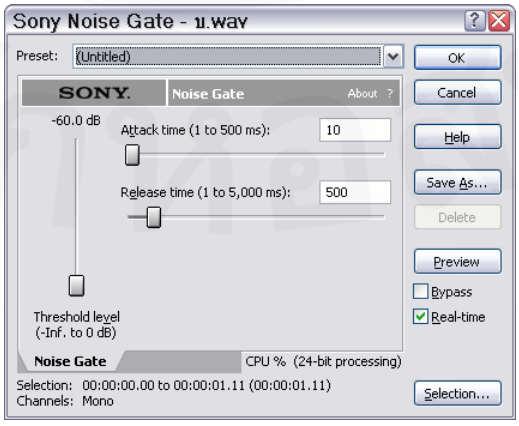
ภาพที่ 3.2

กำหนดค่าการบันทึกเสียง



ภาพที่ 3.3

การตัดเสียงแทรกที่ไม่ต้องการ

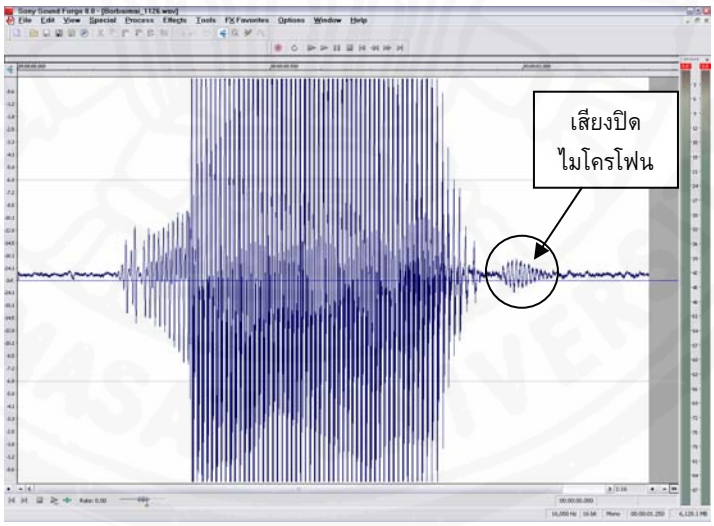


จากนั้นนำเสียงพยัญชนะและสระที่บันทึกมาปรับปรุงเพื่อให้คุณภาพเสียงที่ได้ดีขึ้นโดยใช้ โปรแกรม Sony Forge 8.0 ด้วยการใช้ Effect ที่ชื่อว่า Noise Gate (ดังภาพที่ 3.3) เพื่อตัดเสียงแทรกที่ไม่ต้องการ (นิพนธ์ กิติปภัสสิริ และเอกสิทธิ์ จิตรสภาพ, 2005) จากภาพที่ 3.3 ในแถบสไลด์ Threshold level เป็นการกำหนดว่าถ้าเสียงที่บันทึกไว้มีค่าต่ำถึงระดับ Threshold level ให้ตัดเสียงนั้นทิ้งไป สำหรับค่า Attack time คือการกำหนดเวลาที่จะให้มีการปรับเสียงเป็นปกติเมื่อเสียงมีระดับมากกว่า Threshold level ส่วน Release time คือการกำหนดเวลาที่จะปรับเสียงเป็น 0 เมื่อเสียงที่บันทึกมีระดับเสียงน้อยกว่า Threshold level ในการกำหนดค่าต่าง ๆ จำเป็นต้องทดลองปรับเปลี่ยนปรับค่าไปเรื่อย ๆ จนกว่าจะได้คุณภาพเสียงที่ดีขึ้น

เหตุผลในการปรับปรุงคุณภาพเสียงเนื่องจากเสียงที่บันทึกได้ อาจจะมีเสียงรบกวนอื่น เช่น เสียงหายใจของผู้บันทึกเสียง เสียงเปิด-ปิดไมโครโฟน เป็นต้น (ดังภาพที่ 3.4 และ ภาพที่ 3.5)

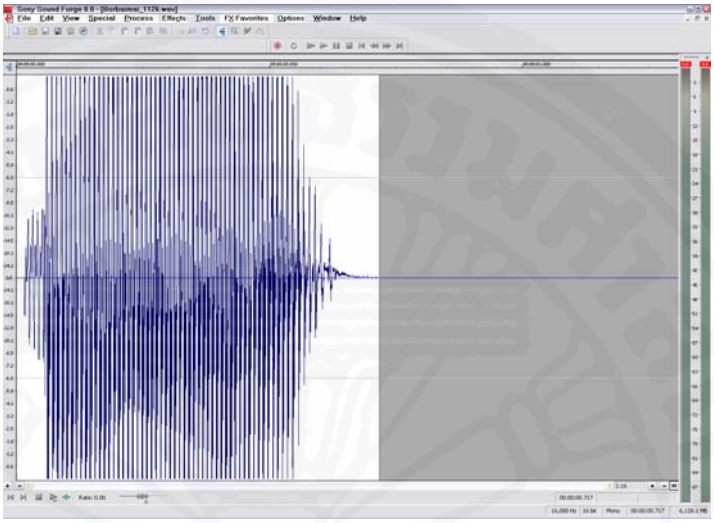
ภาพที่ 3.4

เสียงที่บันทึกได้ก่อนนำมาปรับปรุง



ภาพที่ 3.5

เสียงที่ได้หลังจากผ่านการแก้ไขโดยตัดเสียงรบกวนอื่น ๆ ออกไป



3.1.2 จัดเตรียมฐานข้อมูลพจนานุกรมคำศัพท์

การจัดเตรียมฐานข้อมูลพจนานุกรมคำศัพท์ จะนำคำศัพท์มาจากแฟ้มพจนานุกรม dic5k.txt ซึ่งเป็นแฟ้มข้อความ (Plain Text) ที่ประกอบด้วยคำศัพท์จำนวน 5,000 คำ ซึ่งแฟ้มพจนานุกรมนี้พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์ และคอมพิวเตอร์แห่งชาติ (NECTEC) เพื่อใช้แสดงคำศัพท์ใน LOTUS speech corpus สำหรับตัวอย่างคำศัพท์แสดงไว้ในภาพที่ 3.6

ภาพที่ 3.6

แสดงตัวอย่างคำศัพท์ใน แฟ้มพจนานุกรม dic5k.txt

```
ก. k-@@-z^-0|
กฎ k-o-t^-1|
กฎหมาย k-o-t^-1 m-aa-j^-4|
กฎหมายอาญา k-o-t^-1 m-aa-j^-4 z-aa-z^-0 j-aa-z^-0|
กฎเกณฑ์ k-o-t^-1 k-ee-n^-0|
กด k-o-t^-1|
กดขี่ k-o-t^-1 kh-ii-z^-1|
⋮
```

ในแต่ละบรรทัดของแฟ้ม dict5k.txt ประกอบด้วยคำศัพท์ และลำดับหน่วยเสียงของคำนั้น ๆ โดยจะมีการแบ่งแต่ละพยางค์ ด้วยเครื่องหมาย "|" พร้อมทั้งระบุระดับเสียงวรรณยุกต์ที่ท้ายพยางค์ทุกพยางค์ ในวิทยานิพนธ์นี้ ได้ประยุกต์ใช้พจนานุกรมตามแฟ้ม dict5k.txt ซึ่งมีการดัดแปลงให้เหมาะสมกับงานวิจัยที่นำเสนอ โดยตัดส่วนของวรรณยุกต์และตัดสัญลักษณ์พิเศษออก อาทิเช่น ยติภังค์ (Hyphen) ไปยาลน้อย ไม้ยมก ไปยาลใหญ่ จุลภาค (Comma) จุดคู่ (Colon) อัฒภาค (Semicolon) อัญประกาศเดี่ยว (Single quote) อัญประกาศคู่ (Double quote) ปรัศนีย์ (Question mark) อัศเจรีย์ (Exclamation mark) เครื่องหมายในการคำนวณ ระดับเสียงวรรณยุกต์และ วงเล็บ ออกไป ก่อนนำมาจัดเก็บในฐานข้อมูลพจนานุกรมคำศัพท์เพื่อใช้ในขั้นตอนต่อไป ดังภาพที่ 3.7

ภาพที่ 3.7

แสดงตัวอย่างการดัดแปลงคำศัพท์ในแฟ้มพจนานุกรม dic5k.txt

```
ก k@@
กฎ k o t^
กฎหมาย k o t^ m aa j^
กฎหมายอาญา k o t^ m aa j^ z aa j aa
กฎเกณฑ์ k o t^ k ee n^
กด k o t^
กดขี่ k o t^ kh ii
กดดัน k o t^ d a n^
:
:
```

นอกจากนี้ คำศัพท์ที่อยู่ในพจนานุกรมที่ดัดแปลงมานี้ ไม่ได้ครอบคลุมคำทุกคำที่มีอยู่ในแฟ้มเสียงของแต่ละโดเมนที่ใช้ในวิทยานิพนธ์ จึงจำเป็นต้องเพิ่มคำศัพท์ที่ไม่มีในพจนานุกรมแต่ปรากฏในแฟ้มเสียงลงในฐานข้อมูลพจนานุกรมคำศัพท์เพิ่มเติมด้วย

3.1.3 กำหนดคำหลักในแต่ละโดเมน

การกำหนดคำหลักในแต่ละโดเมน จะพิจารณาจากแฟ้มเสียงที่นำมาทดสอบว่าในแต่ละโดเมนมีคำใดอยู่บ้างและทำการสุ่มคำเหล่านั้นออกมาแล้วกำหนดเป็นคำหลัก จำนวน 10 และ 15 คำ ตามลำดับ ดังต่อไปนี้

1. โดเมนที่เกี่ยวกับการลงทะเบียนของนักศึกษา
   - คำหลัก 10 คำที่ใช้ทดสอบ ได้แก่ “ลง” “ทะเบียน” “ถอน” “วิชา” “เพิ่ม” “ดร็อบ” “บังคับ” “เลือก” “เสรี” และ “เฉพาะ”
   - คำหลัก 15 คำที่ใช้ทดสอบ ได้แก่ “ลง” “ทะเบียน” “ถอน” “วิชา” “เพิ่ม” “ดร็อบ” “บังคับ” “เลือก” “เสรี” “เฉพาะ” “หน่วยกิต” “เทอม” “หลักสูตร” “ภาค” และ “ศึกษา”
2. โดเมนที่เกี่ยวกับการจองห้องพักในโรงแรม
   - คำหลัก 10 คำที่ใช้ทดสอบ ได้แก่ “จอง” “ห้อง” “ราคา” “เตียง” “เดี่ยว” “คู่” “เปลี่ยน” และ ”ยกเลิก”
   - คำหลัก 15 คำที่ใช้ทดสอบ ได้แก่ “จอง” “ห้อง” “ราคา” “เตียง” “เดี่ยว” “คู่” “เปลี่ยน” ”ยกเลิก” “พัก” “คน” “วันที่” “ตกลง” “หนึ่ง” “สอง” และ “เดือน”
3. โดเมนที่เกี่ยวกับข่าวจาก LOTUS speech corpus
   - คำหลัก 10 คำที่ใช้ทดสอบ ได้แก่ “การเมือง” “ประชาชน” “พรรค” “สนใจ” “เปลี่ยนแปลง” “สำคัญ” “ประเทศ” “ระบบ” “ผู้แทน” และ “ปัจจัย”
   - คำหลัก 15 คำที่ใช้ทดสอบ ได้แก่ “การเมือง” “ประชาชน” “พรรค” “สนใจ” “เปลี่ยนแปลง” “สำคัญ” “ประเทศ” “ระบบ” “ผู้แทน” “ปัจจัย” “องค์กร” “เอกชน” “ปรับปรุง” “เสียง” และ “ปวงชน”

ในวิทยานิพนธ์ ได้จัดเก็บคำหลักในแต่ละโดเมนลงในแฟ้ม โดยมีรูปแบบ ดังนี้

- กำหนดให้บรรทัดแรกใช้คำว่า “<DOMAIN>” และตามด้วย ชื่อโดเมน เพื่อกำหนดว่าเป็นคำหลักในโดเมนใด
- กำหนดให้บรรทัดที่สองใช้คำว่า “<KEYWORD>” เพื่อกำหนดคำหลักในโดเมน ปิดท้ายคำหลักด้วย “|” ถ้ามีคำหลักอื่นอีกให้คั่นด้วย “,” แล้วกำหนดคำหลักคำต่อไป
- กำหนดให้บรรทัดที่สาม ใช้ “.” เพื่อแสดงจุดสิ้นสุดของการกำหนดคำหลักในโดเมน

ตัวอย่างของการกำหนดคำหลักในโดเมนทั้ง 3 โดเมน ดังภาพที่ 3.8

ภาพที่ 3.8

รูปแบบการนิยามโดเมนและคำหลักที่ต้องการค้นหา

<DOMAIN> โดเมนลงทะเบียนนักศึกษา
<KEYWORD> ลง|,ทะเบียน|,ถอน|,วิชา|,เพิ่ม|,ดร็อบ|,บังคับ|,เลือก |,เสรี|,เฉพาะ|,หน่วยกิต|,เทอม|,หลักสูตร|,ภาค|,ศึกษา|
.
<DOMAIN> โดเมนจองห้องพักในโรงแรม
<KEYWORD> จอง|, ห้อง|, ราคา|, เตียง|, เดี่ยว|, คู่|, เปลี่ยน|, ยกเลิก|,พัก |,คน|,วันที่|, ตกลง|,หนึ่ง|,สอง|,เดือน|
.
<DOMAIN> LOTUS SPEECH CORPUS
<KEYWORD> การเมือง|, ประชาชน|, พรรค|, สนใจ|, เปลี่ยนแปลง|, สำคัญ|, ประเทศ|, ระบบ|, ผู้แทน|,ปัจจัย|,องค์กร|, เอกชน|, ปรับปรุง|,เสียง|,ปวงชน|
.

3.1.4 **สร้างโมเดลเสียง** (Acoustic Modeling)

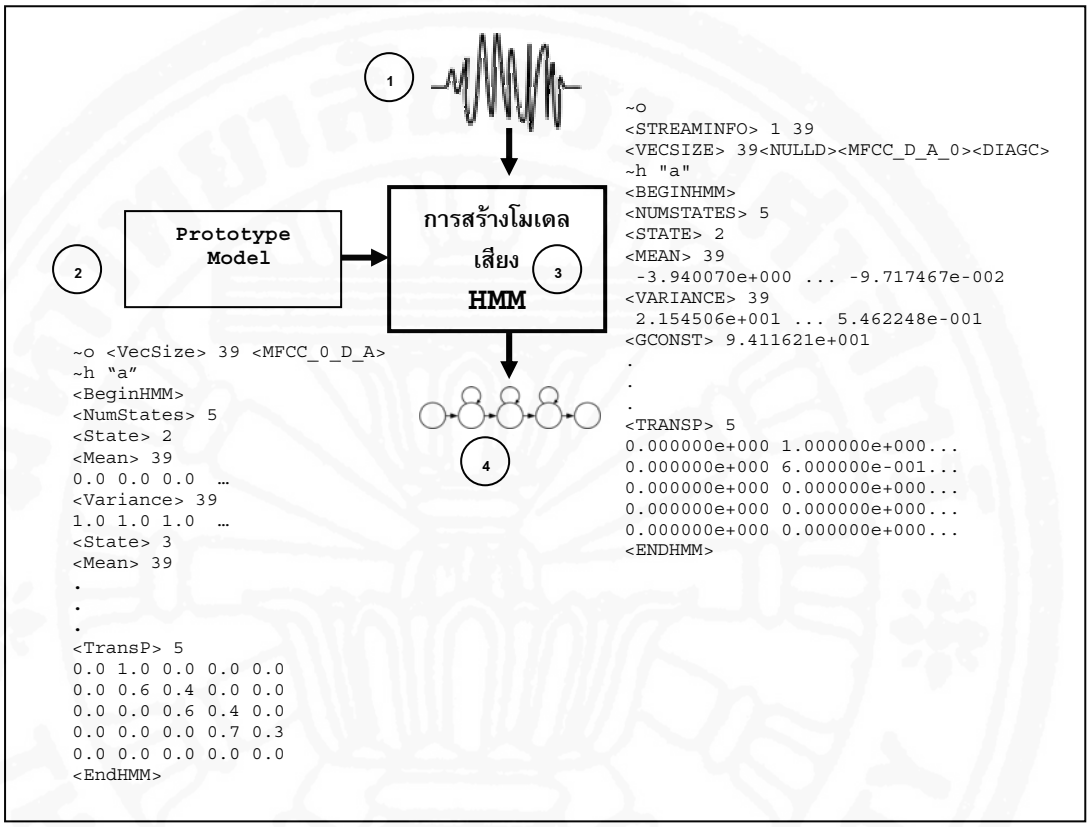
โมเดลเสียง (Acoustic Model) คือ แบบจำลองหน่วยเสียงที่มีหน้าที่คำนวณค่าความน่าจะเป็นของสัญญาณเสียงว่าเสียงที่เกิดเป็นหน่วยเสียงใด ในการสร้างโมเดลเสียง มีขั้นตอนในการทำดังนี้

1. นำแฟ้มเสียงที่ได้บันทึกเตรียมไว้มาทำการสกัดค่าลักษณะสำคัญ (Feature Extraction)
2. กำหนดต้นแบบโมเดลเสียงของหน่วยเสียงภาษาไทยทั้ง 74 หน่วยเสียง
3. สร้างโมเดลเสียง โดยใช้โปรแกรม HTK Toolkit เวอร์ชั่น 3.3
4. นำแฟ้มเสียงในข้อ 1. มาเป็นข้อมูลเข้า (input) สำหรับกำหนดค่าเริ่มต้นให้กับต้นแบบของหน่วยเสียงในข้อ 3.

เมื่อเสร็จขั้นตอนทั้ง 4 แล้ว จะได้โมเดลเสียงของหน่วยเสียงที่ต้องการ ตัวอย่างขั้นตอนแสดงดังภาพที่ 3.9

ภาพที่ 3.9

ขั้นตอนการสร้างโมเดลหน่วยเสียง



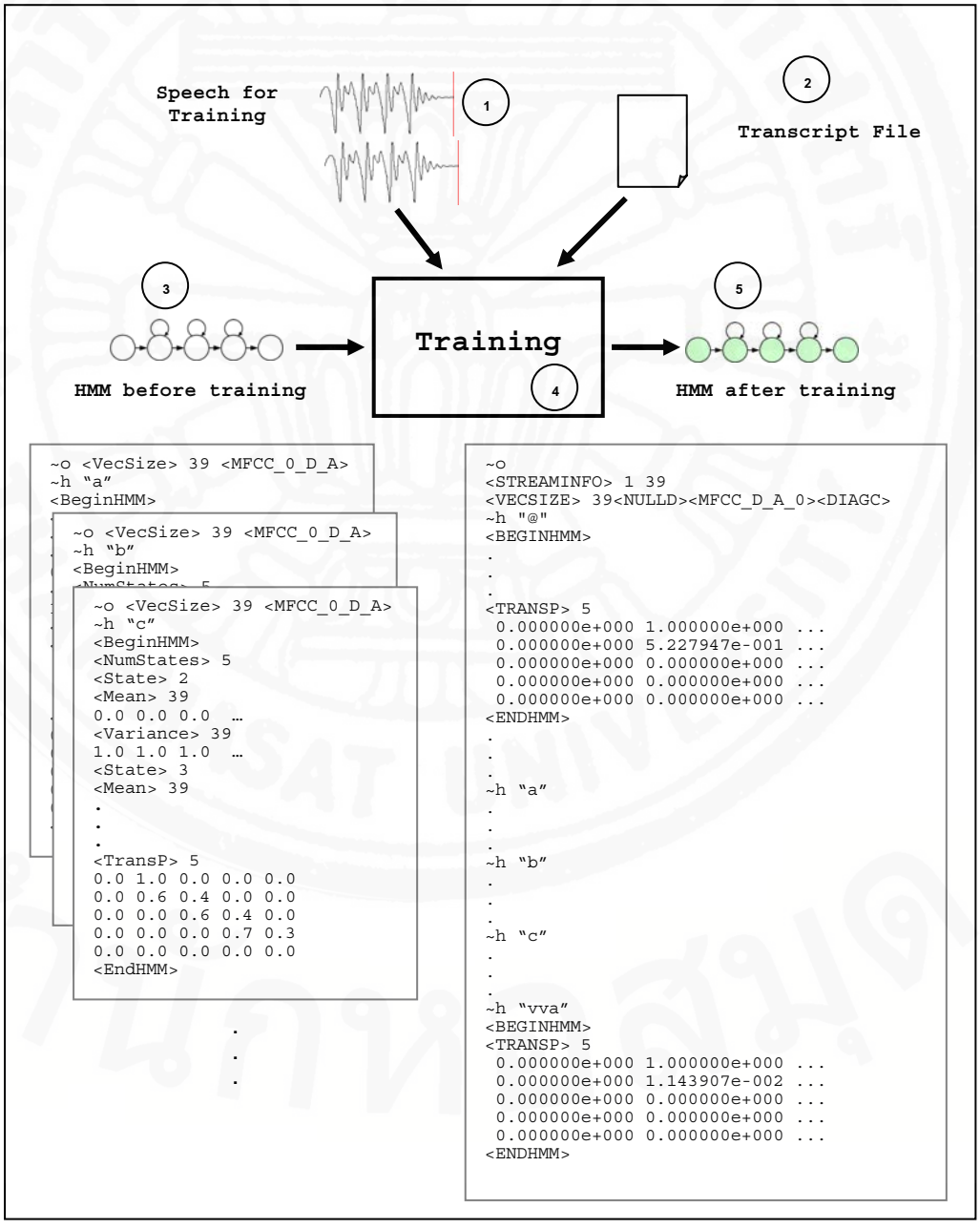
3.1.5 ขั้นตอนการฝึกฝน

หลังจากที่ได้โมเดลหน่วยเสียงในภาษาไทยทั้ง 74 หน่วยเสียงแล้ว ได้นำแฟ้มเสียงที่ได้จัดเตรียมไว้ พร้อมกับแฟ้มการออกเสียง (Transcription File) มาฝึกฝนให้กับโมเดลหน่วยเสียงที่สร้างจากขั้นตอน 3.1.4 เพื่อปรับค่าความน่าจะเป็นของโมเดลต้นแบบให้ใกล้เคียงกับแฟ้มเสียงที่ใช้ฝึกฝน ดังภาพที่ 3.10 โดยประกอบด้วยขั้นตอนแต่ละขั้นตอน ดังนี้

1. นำแฟ้มเสียงจากคลังข้อมูลเสียงที่ได้จัดเตรียมไว้ มาทำการสกัดค่าลักษณะสำคัญ (Feature Extraction)
2. กำหนดแฟ้มการออกเสียง โดยแสดงรายละเอียดว่าแฟ้มเสียงในข้อ 1. ประกอบด้วยคำใดบ้าง
3. นำโมเดลเสียงของหน่วยเสียงที่สร้างไว้ในขั้นตอน 3.1.4 มาทำการฝึกฝน

4. ทำการฝึกฝนโมเดลเสียงในข้อ 3. โดยนำแฟ้มใน ข้อ 1. ข้อ 2. มาเป็นข้อมูลเข้าเพื่อฝึกฝนโมเดลเสียงให้รู้จำหน่วยเสียงของคำแต่ละคำในแฟ้มเสียง
5. โมเดลเสียงที่ผ่านการฝึกฝนให้รู้จำหน่วยเสียง

ภาพที่ 3.10
ขั้นตอนการฝึกสอนให้ HMM รู้จำเสียง



ส่วนในภาพที่ 3.11 เป็นตัวอย่างแฟ้มการออกเสียง (Transcription File) ซึ่งจะมีการกำหนดว่าแฟ้มเสียงที่มาฝึกฝนประกอบด้วยคำอะไรบ้าง โดยกำหนดคำละ 1 บรรทัดเป็นลำดับต่อกันไปจนจบประโยค

ภาพที่ 3.11

ตัวอย่าง แฟ้มการออกเสียง (Transcription File)

```
#!MLF!#                          <-- Header File
"*/CCF001_Pa001_001.lab"         <-- Label file name
นาย
สง่า                             <-- Word sequence
สรรพศรี
.                                <-- End of label
"*/CCF001_Pa001_002.lab"
ดร.
ประสาร
อ่านเปรื่อง
.
"*/CCF001_Pa001_011.lab"
เพราะ
โฟโตสเฟียร์
สว่าง
ขึ้น
.
"*/CCF001_Pa001_019.lab"
เกณฑ์
การ
เลือก
ใช้
สมุนไพร
เพื่อ
การ
สาธารณสุขมูลฐาน
.
```

เมื่อฝึกฝนโมเดลเสียงของแต่ละโดเมนเรียบร้อย จะได้โมเดลหน่วยเสียงพื้นฐานที่นำมาสร้างโมเดลคำหลักในขั้นตอนต่อไป

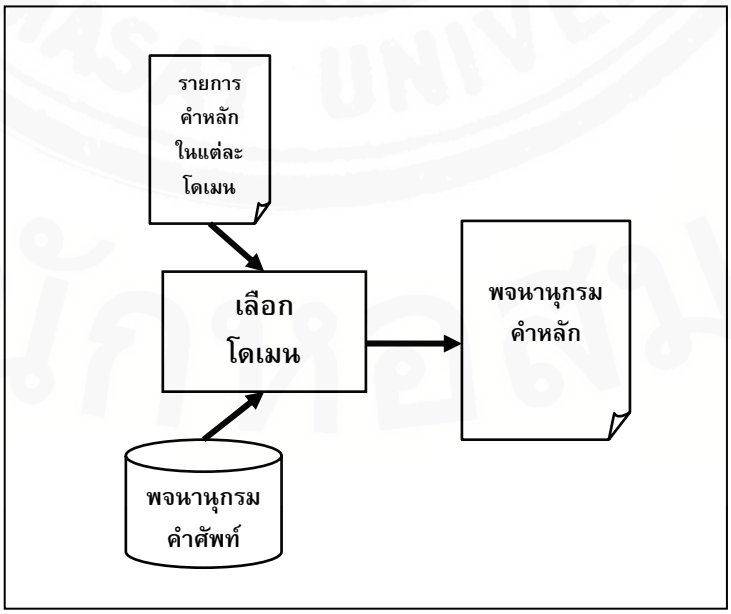
3.2 ขั้นตอนการสร้างโมเดลคำหลักโดยอัตโนมัติ (Automatic Keyword Generation)

ในขั้นตอนนี้ ผู้วิจัยได้พัฒนาโปรแกรมโดยใช้ภาษา C# ในชุดพัฒนาโปรแกรม Microsoft Visual Studio .NET 2003 และใช้ฐานข้อมูล Microsoft Access 2002 ซึ่งประกอบด้วยขั้นตอน ดังต่อไปนี้

3.2.1 เลือกโดเมนที่ต้องการค้นหาคำหลัก และสร้างพจนานุกรมคำหลัก

จุดประสงค์ของการเลือกโดเมนเพื่อแสดงรายการคำหลักในโดเมนว่าต้องการที่ค้นหาคำหลักใดในโดเมนที่เลือก ด้วยการอ่านแฟ้มรายการคำหลักในโดเมนที่สร้างไว้ในขั้นตอน 3.1.3 ดังภาพที่ 3.12 หลังจากเลือกคำหลักที่ต้องการได้แล้ว โปรแกรมจะทำการสร้างพจนานุกรมคำหลักตามคำหลักที่เลือก เพื่อนำไปใช้สร้างโมเดลคำหลักในขั้นตอนต่อไป ในการทดลองค้นหาคำหลักจะทำการทดลองหาคำหลักในโดเมนทั้ง 3 ทีละโดเมน

ภาพที่ 3.12
แสดงขั้นตอนในเลือกการโดเมน และสร้างพจนานุกรมคำหลัก



สำหรับการสร้างพจนานุกรมคำหลัก ใช้วิธีการสร้างโดยการค้นหารูปแบบการออกเสียงของคำหลักจากฐานข้อมูลพจนานุกรมคำศัพท์ที่จัดเตรียมไว้ในหัวข้อ 3.1.2 ซึ่งโปรแกรมจะสร้างพจนานุกรมคำหลักให้อัตโนมัติ ในภาพที่ 3.13 เป็นตัวอย่างรูปแบบพจนานุกรมคำหลักของโดเมนลงทะเบียนนักศึกษา

ภาพที่ 3.13
รูปแบบของพจนานุกรมคำหลัก

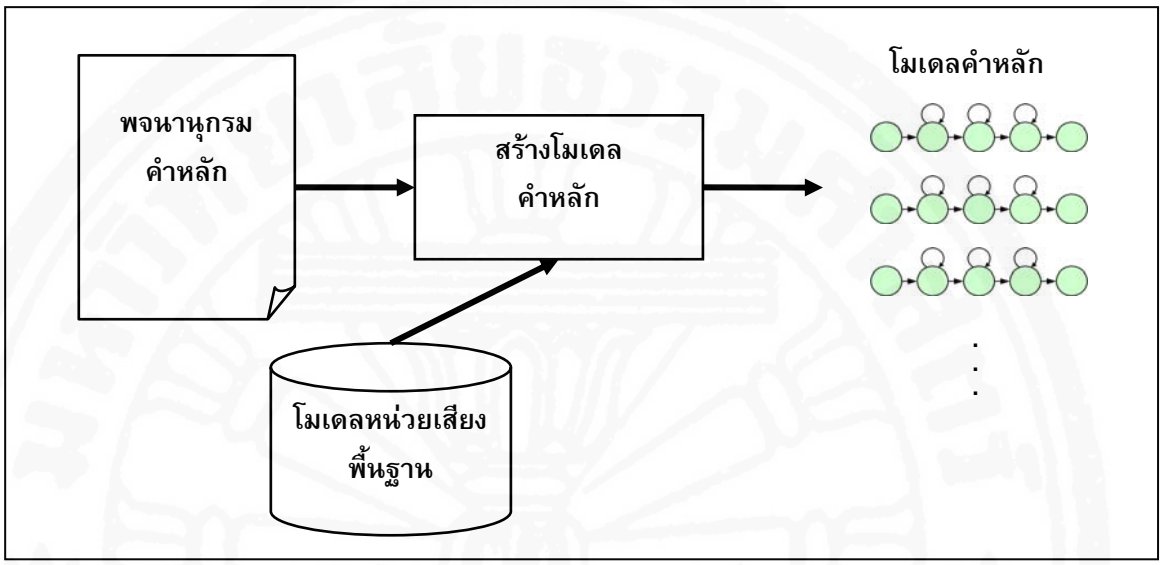
ลง	l o ng^ sp
ทะเบียน	th a b iia n^ sp
ถอน	th @@ n^ sp
วิชา	w i ch aa sp
เพิ่ม	ph qq m^ sp
ดร็อบ	dr @ p^ sp
บังคับ	b a ng^ kh a p^ sp
เลือก	l vva k^ sp
เสรี	s ee r ii sp
เฉพาะ	ch a ph @ sp
หน่วยกิต	n uua j^ k i t^ sp
เทอม	th qq m^ sp
หลักสูตร	l a k^ s uu t^ sp
ภาค	ph aa k^ sp
ศึกษา	s v k^ s aa sp

3.2.2 **สร้างโมเดลคำหลัก** (Keyword Model Generating)

ขั้นตอนต่อมาโปรแกรมจะทำการสร้างโมเดลคำหลักตามหน่วยเสียงที่กำหนดไว้ในพจนานุกรมคำหลักที่ได้จากขั้นตอน 3.2.1 ดังภาพที่ 3.14

ภาพที่ 3.14

สร้างโมเดลคำหลักจากพจนานุกรมคำหลัก



การสร้างโมเดลคำหลักนี้ จะสร้างโมเดลตามโมเดลหน่วยเสียงพื้นฐานที่ได้จัดเตรียมไว้ในขั้นตอน 3.1.5 โดยจะทำการสร้างโมเดลคำหลัก HMM แบบ Triphone model (Steve Yong, 2005, p. 175) ซึ่งจะประกอบด้วยสถานะ 4 รูปแบบดังนี้

- Ci+Vi เป็นสถานะเริ่มต้น (start state)
- Ci-Vi+Cf เป็นสถานะวนอยู่ที่สถานะเดิม หรือเคลื่อนไปข้างหน้า (emit state)
- Vi-Cf+Vf เป็นสถานะวนอยู่ที่สถานะเดิม หรือเคลื่อนไปข้างหน้า (emit state)
- Cf-Vf เป็นสถานะสุดท้าย (final state)

โดย

Ci หมายถึง เสียงของพยัญชนะต้น (initial consonant)

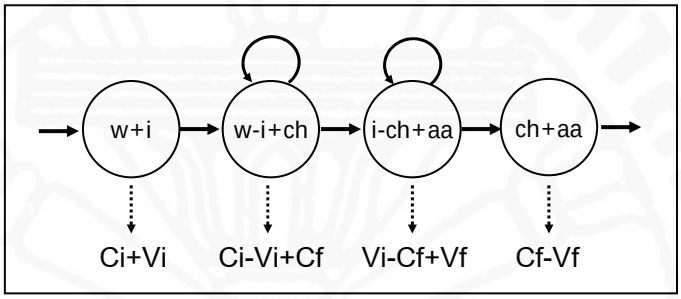
Vi หมายถึง เสียงของสระต้น (initial vowel)

Cf หมายถึง เสียงของพยัญชนะท้าย (final consonant)

Vf หมายถึง เสียงของสระท้าย (final vowel)

ตัวอย่างเช่น คำว่า “วิชา” ประกอบด้วยลำดับของหน่วยเสียง “w i ch aa” โปรแกรมจะทำการสร้างโมเดลคำหลักโดยอัตโนมัติ 4 สถานะ w+i w-i+ch i-ch+aa และ ch-aa ดังภาพที่ 3.15

ภาพที่ 3.15
โมเดลคำหลัก คำว่า “วิชา”



3.2.3 **สร้างโมเดลภาษา** (Language Model Generating)

ในวิทยานิพนธ์นี้ ได้สร้างโมเดลภาษาสำหรับคำหลักให้อัตโนมัติ ด้วยวิธีกำหนด Regular grammar โดยนำคำหลักจากพจนานุกรมคำหลักมาสร้าง เพื่อบอกว่ามีคำหลักใดบ้างอยู่ในภาษา ดังภาพที่ 3.16

ภาพที่ 3.16
ตัวอย่าง โมเดลภาษาของโดเมนลงทะเบียนนักศึกษา
ด้วยวิธีกำหนด Regular grammar

```
(silence (<ลง | ทะเบียน | ถอน| วิชา | เพิ่ม | ดร็อบ | บังคับ | เลือก
|เสรี | เฉพาะ | หน่วยกิต | เทอม | หลักสูตร | ภาค | ศึกษา>) silence)
```

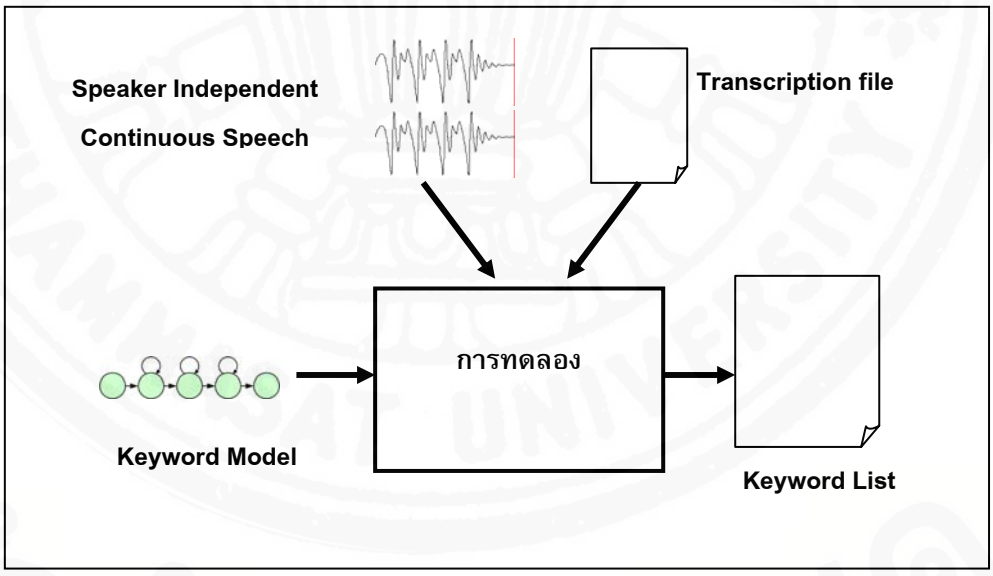
สำหรับการสร้างโมเดลภาษาด้วยวิธีนี้ โปรแกรมจะทำการสร้างหน่วยเสียง silence ซึ่งเป็นหน่วยเสียงพิเศษ หมายถึง เสียงเงียบ โดยจะกำหนดไว้ก่อนหน้าและหลังคำหลัก เพื่อบอกจุดเริ่มต้นเสียง และจุดสิ้นสุดเสียงของคำหลักแต่ละคำ

เมื่อได้โมเดลภาษาสำหรับคำหลักแล้ว โปรแกรมจะสร้าง network ของคำหลักเพื่อใช้ในการหาผลการทดลองในขั้นตอนต่อไป

3.3 **ขั้นตอนการประเมินประสิทธิภาพ** (Evaluation)

สำหรับขั้นตอนการทดลอง จะนำแฟ้มเสียงที่จัดเตรียมไว้สำหรับการทดลองในแต่ละโดเมนมาหาคำหลักทีละโดเมน ซึ่งแฟ้มเสียงดังกล่าวได้ผ่านการสกัดค่าลักษณะสำคัญแล้ว มาทำการค้นหาว่าแฟ้มเสียงนั้นมีเสียงของคำที่ตรงกับโมเดลคำหลักที่สร้างไว้หรือไม่ จากนั้นโปรแกรมจะบันทึกผลที่ค้นหาได้ เพื่อนำมาเปรียบเทียบกับแฟ้มออกเสียงว่ามีตำแหน่งของคำหลักที่ค้นหาได้ตรงกับคำหลักที่อยู่ในแฟ้มออกเสียงหรือไม่ ดังภาพที่ 3.17

ภาพที่ 3.17
ขั้นตอนการทดลอง



ผลลัพธ์ที่ได้หลังจากทำการค้นหาคำหลัก จะได้เป็นรายการคำหลักที่ค้นหาได้ โดยในแต่ละบรรทัดประกอบด้วย คำหลักที่ใช้ค้นหากับจำนวนของคำหลักนั้นที่สามารถค้นหาได้จากแฟ้มเสียง และบรรทัดสุดท้ายแสดงผลรวมของคำหลักทุกคำที่ค้นหาได้ ดังภาพที่ 3.18 ซึ่งค่าที่ได้จะนำไปใช้ในสูตรคำนวณในขั้นตอนต่อไป

ภาพที่ 3.18
ตัวอย่างรายการคำหลักที่ค้นหาได้ สำหรับคำหลัก 15 คำ
ที่กำหนดในไว้โดเมนลงทะเบียนนักศึกษา

วิชา = 71
ทะเบียน = 5
หน่วยกิต = 35
ลง = 5
ภาค = 3
ถอน = 4
ศึกษา = 20
บังคับ = 12
เทอม = 1
เพิ่ม = 1
เลือก = 8
เสรี = 2
ดร็อบ = 0
เฉพาะ = 0
หลักสูตร = 0
Summary : 167

สำหรับการวัดประสิทธิภาพการค้นหาคำหลักในวิทยานิพนธ์นี้ ใช้สูตรการวัดประสิทธิภาพในการสืบค้นข้อมูล (Information retrieval) 4 สูตร คือ

1. สูตร Precision เป็นสูตรที่ใช้วัดความถูกต้องในการค้นหาคำหลักจากจำนวนคำหลักทั้งหมดที่ปรากฏอยู่ในแฟ้มการออกเสียงที่ใช้ทดสอบ โดยจะนำผลรวมความถูกต้องที่โปรแกรมสามารถหาคำหลักได้ตรงกับคำหลักในแฟ้มออกเสียงมาหารกับผลรวมของคำหลักทั้งหมดที่ปรากฏอยู่ในแฟ้มการออกเสียง ดังแสดงในสมการที่ (3.1)
2. สูตร Recall เป็นสูตรที่ใช้วัดความถูกต้องในการค้นหาคำหลักจากจำนวนคำทั้งหมดที่ปรากฏอยู่ในแฟ้มการออกเสียงที่ใช้ทดสอบ โดยจะนำผลรวมความถูกต้องที่โปรแกรมสามารถหาคำหลักได้ตรงกับคำหลักในแฟ้มออกเสียงมาหารกับผลรวมของคำทั้งหมดที่ปรากฏอยู่ในแฟ้มการออกเสียง ดังแสดงในสมการที่ (3.2)

3. สูตร F-measure เป็นสูตรที่วัดสัดส่วนระหว่าง Precision กับ Recall ดังแสดงในสมการที่ (3.3)
4. สูตร Fall-out เป็นสูตรที่ใช้วัดความผิดพลาดในการค้นหาคำหลักจากจำนวนคำทั้งหมดที่ไม่ใช่คำหลัก โดยนำผลต่างระหว่างผลรวมของคำหลักที่อยู่แฟ้มเสียงที่ใช้ทดสอบทั้งหมด หักกับผลรวมของจำนวนคำหลักที่ระบบสามารถค้นหาได้ จากนั้นนำผลต่างที่ได้มาหารกับจำนวนคำที่ไม่ใช่คำหลักทั้งหมดที่อยู่ในแฟ้มเสียง ดังสมการที่ (3.4)

ซึ่งสามารถเขียนเป็นสูตรได้ดังภาพที่ 3.19

ภาพที่ 3.19
สูตรการวัดประสิทธิภาพของ
การค้นหาคำหลัก

$$\text{Precision} = \frac{|\text{No. of Hit keywords}|}{|\text{No. of all keywords in test speech}|} \quad (3.1)$$

$$\text{Recall} = \frac{|\text{No. of Hit keywords}|}{|\text{No. of all words in test speech}|} \quad (3.2)$$

$$\text{F - measure} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (3.3)$$

$$\text{Fall - Out} = \frac{|(\text{No.of all keyword intest speech} - \text{No.of hit keywords})|}{|(\text{No.of all wordsi n tests peech} - \text{No.of all keyword in test seepch})|} \quad (3.4)$$

- No. of Hit keywords = ผลรวมของคำหลักที่ค้นหาได้ตรงกับคำหลักในแฟ้มออกเสียง
- No. of all keywords in test speech = ผลรวมของคำหลักทั้งหมดที่ปรากฏอยู่ในแฟ้มการออกเสียง
- No. of all words in test speech = ผลรวมของคำทั้งหมดทั้งเป็นคำหลักและไม่ใช่คำหลักที่ปรากฏอยู่ในแฟ้มการออกเสียง

ตารางที่ 3.3 เป็นตัวอย่างผลการทดลองในการค้นหาคำหลัก 15 คำในโดเมนลงทะเบียนนักศึกษา โดยค่าในตารางสามารถอธิบายตามหมายเลขที่แสดงไว้ในเครื่องหมาย “[]” ได้ดังนี้

ตารางที่ 3.3
ตารางแสดงจำนวนคำหลัก 15 คำที่ค้นหาได้จาก
แฟ้มเสียงในโดเมนลงทะเบียนนักศึกษา

	คำหลัก [1]	**คำหลักที่อยู่ในแฟ้มเสียง [2]**	**คำหลักที่หาได้ [3]**
1	ลง	6	5
2	ทะเบียน	6	5
3	ถอน	4	4
4	วิชา	71	71
5	เพิ่ม	2	1
6	ดร็อบ	0	0
7	บังคับ	13	12
8	เลือก	8	8
9	เสรี	2	2
10	เฉพาะ	0	0
11	หน่วยกิต	35	35
12	เทอม	2	1
13	หลักสูตร	0	0
14	ภาค	3	3
15	ศึกษา	28	20
	รวม	**180 [4]**	**167 [5]**
		รวมจำนวนคำทั้งหมด	**1,241 [6]**

[1] หมายถึง คำหลักที่กำหนดไว้ในโดเมน

[2] หมายถึง จำนวนคำหลักทั้งหมดที่มีอยู่จริงในแฟ้มเสียงที่ใช้ทดสอบ

[3] หมายถึง จำนวนคำหลักที่ระบบสามารถค้นหาได้

[4] หมายถึง ผลรวมของคำหลักทั้งหมดที่มีปรากฏอยู่ในแฟ้มเสียงที่ใช้ทดสอบ

[5] หมายถึง ผลรวมของจำนวนคำหลักที่ระบบสามารถค้นหาได้

[6] หมายถึง ผลรวมของจำนวนคำทั้งหมดที่ปรากฏอยู่ในแฟ้มเสียงที่ใช้ทดสอบ

นำค่าที่ได้จากตารางที่ 3.3 มาคำนวณตามสูตรการวัดประสิทธิภาพของการค้นหาคำหลักทั้ง 4 สูตร ได้ผลลัพธ์ ดังนี้

1. Precision = 167/180 = 0.93
2. Recall = 167/1,241 = 0.14
3. F-Measure = 2(0.93 x 0.13) / (0.93+0.13) = 0.24
4. Fall-Out = (180-167)/(1,241-180) = 13/1,061 = 0.12

เมื่อได้ผลการคำนวณตามสูตรการวัดประสิทธิภาพของการค้นหาคำหลักทั้ง 4 สูตร แล้วจะทำการบันทึกผลลัพธ์ที่ได้จากการทดสอบค้นหาคำหลักในแต่ละโดเมน เพื่อนำมาวิเคราะห์ผลการทดลองในบทต่อไป