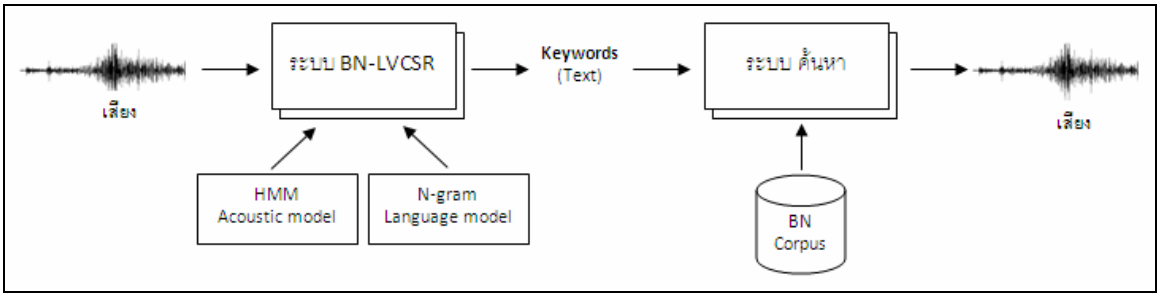
# บทที่ 3

## วิธีการดำเนินงานวิจัย

งานวิทยานิพนธ์นี้เป็นการนำเสนอระบบการสืบค้นข้อมูลด้วยเสียงสำหรับข่าวภาษาไทย ซึ่งเป็นระบบที่ใช้เป็นการรู้จำเสียงพูดต่อเนื่องที่ใช้คำศัพท์จำนวนมาก (Large vocabulary continuous speech recognition: LVCSR) ในระบบ LVCSR ที่นำเสนอนี้เป็นข้อมูลของข่าวจึงขอแทนระบบนี้ว่า Broadcast News - Large vocabulary continuous speech recognition: BN-LVCSR มีแบบจำลองหน่วยเสียง (Acoustic model) และแบบจำลองภาษา (Language model) เป็นส่วนประกอบของระบบ ข้อมูลเข้าของระบบที่นำเสนอเป็นข้อมูลเสียง เมื่อผ่านระบบ BN-LVCSR จะได้ข้อความ ซึ่งก็คือคำสำคัญ (keyword) นำคำสำคัญไปค้นหาข้อมูลในข่าวที่ถอดความแล้ว (BN transcription) ผลลัพธ์ที่ดีคือเสียงที่มีคำสำคัญนั้นเป็นส่วนประกอบในเสียง

ภาพที่ 3.1
โครงสร้างของระบบสืบค้นข้อมูลด้วยเสียงสำหรับข่าวภาษาไทย



## ขอบเขตของงานวิจัย

### 1. ข้อมูลที่ใช้ในการทำวิจัย

#### 1.1 ข้อมูลเสียงทดสอบที่ใช้ในงานวิจัย

ชุดข้อมูลเสียงการทดสอบ (Test set data) เป็นเสียงพูดต่อเนื่องแบบอ่าน โดยทำการอัดเสียง 16 บิต 16 กิโลเฮิร์ต โมโน และเก็บไฟล์ในรูปแบบ ของ PCM wav คำที่อัดเสียงนั้นต้องเป็นคำที่มีอยู่ในคลังข้อมูลข่าวภาษาไทย (Thai Broadcast News Corpus)

#### 1.2 โมเดลทางกายภาพของเสียง (Acoustic Model)

อาศัยแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model: HMM) 1 แบบจำลองต่อหน่วยเสียง 1 หน่วย (Phoneme-based) ไม่มีการใช้ข้อมูลเกี่ยวกับเสียงวรรณยุกต์ในการรู้จำ

#### 1.3 โมเดลทางภาษา (Language Model)

อาศัยแบบจำลองเอ็นแกรม (N-gram Model) ในการคำนวณความน่าจะเป็นในการเกิดประโยค (Language probability) โดย N=3

#### 1.4 โดเมนของงานวิจัย

โดเมนของเสียงจะเป็นข่าว โดยใช้คลังข้อมูลเสียงข่าวชื่อ Broadcast News ซึ่งอัดเสียงมาจากรายการข่าวภาษาไทย มีข้อมูลเสียงประมาณ 17 ชั่วโมง และข้อมูลที่อยู่ในรูปการถอดความคำพูดจากข่าว 35 ชั่วโมง 13,045 ไฟล์

#### 1.5 คลังข้อมูลเสียงที่ใช้ในการฝึกฝนและทดสอบ

ข้อมูลในงานวิจัยนี้ใช้ฐานข้อมูลเสียงขนาดใหญ่สำหรับระบบรู้จำเสียงพูดต่อเนื่องภาษาไทย (LOTUS Corpus) ข้อมูลอยู่ในรูปแบบการพูดโดยอ่านจากหนังสือพิมพ์ มีข้อมูลเสียงจำนวน 50 ชั่วโมง และข้อมูลที่เป็นข้อความ 50 ชั่วโมง 22,841 ไฟล์

#### 1.6 ลักษณะของผู้พูดเสียงในการทดสอบ

ลักษณะของผู้พูดเสียงในการทดสอบเป็น ผู้พูดที่เจาะจง (Speaker dependent) โดยเป็นเสียงพูดผู้ชาย 1 คน

## 2. เครื่องมือที่ใช้ในการทำวิจัย

### 2.1 เครื่องมือซอฟต์แวร์ (Software Tools)

2.1.1 Hidden Markov Toolkit (HTK) version 3.4 ดาว์นโหลด ได้ที่ http://htk.eng.cam.ac.uk/ HTK คือ เครื่องมือสำหรับ HMM (Hidden Markov Model) เพื่อประมวลผลเกี่ยวกับการสร้างและรู้จำเสียง

2.1.2 Carnegie Mellon Statistical Language Modeling Toolkit (CMU SLM) version 2.05 ดาว์นโหลดได้ที่ http://www.speech.cs.cmu.edu/SLM/toolkit.html CMU SLM คือเครื่องมือที่ใช้ในการเตรียม N-gram Language model

2.1.3 JULIUS version 4.1.1 ดาว์นโหลด ได้ที่ http://julius.sourceforge.jp/en_index.php JULIUS คือ เครื่องมือที่ใช้ทำ Speech Decoder แบบ Two-pass search

### 2.2 เครื่องมือฮาร์ดแวร์ (Hardware Tools)

2.2.1 ระบบปฏิบัติการ Ubuntu (Linux) version 8.10 ขึ้นไป

2.2.2 หน่วยความจำ 256 Mb ขึ้นไป

2.2.3 ติดตั้ง JavaSDK 1.5 ขึ้นไป

2.2.4 PHP เวอร์ชั่น 5

2.2.5 MySQL เวอร์ชั่น 5

## สถาปัตยกรรมของระบบ

### 1. แนวคิดและระบบโดยรวม

โครงสร้างระบบการสืบค้นข้อมูลด้วยเสียงสำหรับข่าวภาษาไทยที่นำเสนอดังภาพที่ 3.1 โดยระบบจะแบ่งการทำงานออกเป็น 2 ขั้นตอน ดังนี้

1. ขั้นตอนการแปลงไฟล์เสียงเป็นคำสำคัญโดยในส่วนนี้จะทำการสร้างระบบ BN-LVCSR เพื่อทำการรู้จำเสียง
2. ขั้นตอนการค้นหาโดยใช้คำสำคัญ จากผลลัพธ์ที่ได้จากระบบ BN-LVCSR จะเป็นลำดับคำที่มีความน่าจะเป็นสูงสุดซึ่งก็คือคำสำคัญของระบบ โดยจะใช้ในการค้นหาในคลังข้อมูลข่าว

## ขั้นตอนการทดลอง

### 1. ขั้นตอนการสร้างระบบ BN-LVCSR

#### 1.1 กำหนดค่าลักษณะสำคัญ

ค่าลักษณะสำคัญที่ใช้มีจำนวน 25 ค่าต่อสัญญาณเสียง 1 เฟรม ประกอบด้วยสัมประสิทธิ์เซปสตรัมบนสเกลแบบเมล (Mel-scale cepstral coefficient: MFCC) จำนวน 12 ค่า ค่าผลต่าง (Delta) ของ MFCC 12 ค่า และค่าผลต่างของค่าพลังงาน (Delta energy) อีก 1 ค่า

#### 1.2 กำหนดแบบจำลองหน่วยเสียง

สำหรับการฝึกฝนแบบจำลองหน่วยเสียงนั้น จะใช้เครื่องมือฮิดเดนมาร์คอฟ (Hidden Markov Toolkit: HTK) ซึ่งเป็นเครื่องมือสำหรับแบบจำลองฮิดเดนมาร์คอฟเพื่อประมวลผลเกี่ยวกับการสร้างและรู้จำเสียง การสร้างแบบจำลองหน่วยเสียงมีขั้นตอน ดังนี้

1.2.1 การแปลงไฟล์เสียงอยู่ในรูปแบบ .wav ให้เป็นไฟล์ค่าลักษณะสำคัญอยู่ในรูปแบบ .mfc ด้วยคำสั่ง HCopy ใน HTK จะนำไฟล์ค่าลักษณะสำคัญมาใช้ในการฝึกฝนแบบจำลองหน่วยเสียงต่อไป

1.2.2 เตรียมไฟล์ที่กำกับว่าในแต่ละไฟล์เสียงที่ใช้ในการฝึกฝนประกอบด้วยหน่วยเสียงใดบ้าง ในแต่ละไฟล์เสียงจะมีหน่วยเสียง “sil” ปิดหัวปิดท้าย ซึ่งก็คือเสียงเงียบ (Silence) เพื่อกำหนดการเริ่มต้นและสิ้นสุดของเสียง ตัวอย่างของไฟล์ที่กำกับหน่วยเสียงแสดงในภาพที่ 3.2

ภาพที่ 3.2
ตัวอย่างของไฟล์ที่กำกับด้วยหน่วยเสียง

```
#!MLF!#
"*/UOF001_Pa001_001.lab"
sil
n
aa
j^
s
a
ng
aa
s
a
p^
ph
a
s
ii
sil
.
"*/UOF001_Pa001_002.lab"
sil
d
@
k^
t
qq
pr
a
s
aa
n^
z
aa
n^
pr
vva
ng^
sil
.
```

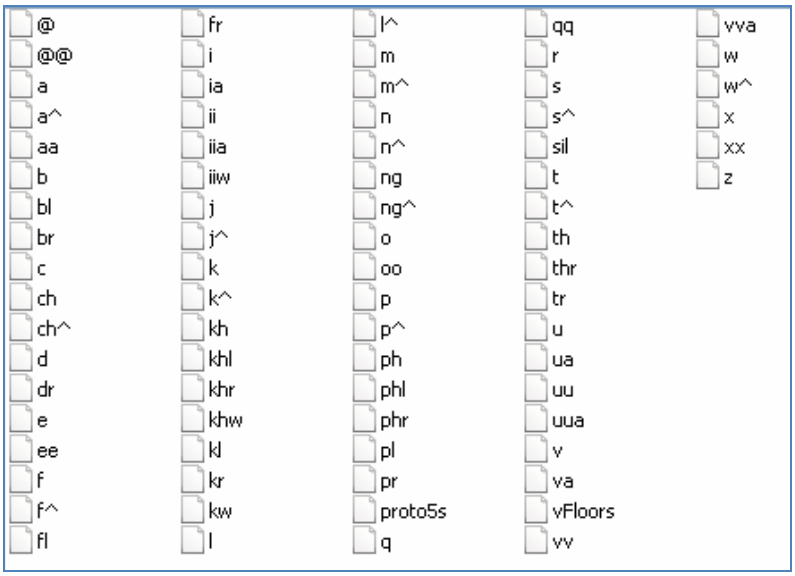
1.2.3 สร้างหน่วยเสียงเดี่ยว (Monophone) โดยเริ่มต้นนั้นจะคำนวณค่าเฉลี่ยของแบบจำลองฮิดเดนมาร์คอฟ ในที่นี้จะใช้วิธีการ Flat start คือทุกหน่วยเสียงจะสร้างจากไฟล์ต้นแบบตัวเดียวกัน แล้วนำมาคำนวณค่าโดยใช้คำสั่ง HCompV ผลที่ได้คือ แบบจำลองฮิดเดนมาร์คอฟที่เฉลี่ยค่าความน่าจะเป็นให้เท่าๆกันสำหรับทุกหน่วยเสียง เก็บอยู่ในไฟล์ proto5s แสดงในภาพที่ 3.3 แล้วทำการคัดลอก (copy) ไปเป็นหน่วยเสียงแต่ละตัว ซึ่งต้องมีอยู่ในไฟล์ที่กำกับไฟล์เสียงไว้ แสดงในภาพที่ 3.4

ภาพที่ 3.3

ตัวอย่างของไฟล์ proto5s

```
~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_E_D_A><DIAGC>
~h "proto5s"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 39
 -6.919414e+00 -3.986485e+00 2.004288e+00 -3.946909e+00 -2.346894e+00 -5.416295e
<VARIANCE> 39
 3.845368e+01 5.648943e+01 6.281944e+01 6.733347e+01 6.601381e+01 5.630132e+01 5
<GCONST> 1.167378e+02
<STATE> 3
<MEAN> 39
 -6.919414e+00 -3.986485e+00 2.004288e+00 -3.946909e+00 -2.346894e+00 -5.416295e
<VARIANCE> 39
 3.845368e+01 5.648943e+01 6.281944e+01 6.733347e+01 6.601381e+01 5.630132e+01 5
<GCONST> 1.167378e+02
<STATE> 4
<MEAN> 39
 -6.919414e+00 -3.986485e+00 2.004288e+00 -3.946909e+00 -2.346894e+00 -5.416295e
<VARIANCE> 39
 3.845368e+01 5.648943e+01 6.281944e+01 6.733347e+01 6.601381e+01 5.630132e+01 5
<GCONST> 1.167378e+02
<TRANSP> 5
 0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
 0.000000e+00 6.000000e-01 4.000000e-01 0.000000e+00 0.000000e+00
 0.000000e+00 0.000000e+00 6.000000e-01 4.000000e-01 0.000000e+00
 0.000000e+00 0.000000e+00 0.000000e+00 6.000000e-01 4.000000e-01
 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
<ENDHMM>
```

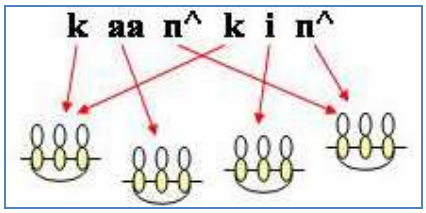
ภาพที่ 3.4
ตัวอย่างของหน่วยเสียงแต่ละตัวที่มีอยู่ในไฟล์ที่กำกับไฟล์เสียงไว้



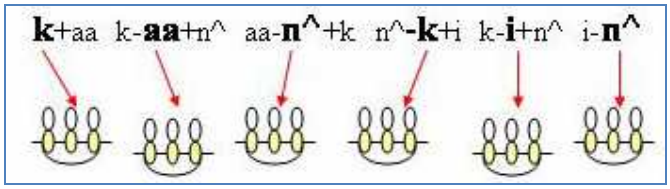

1.2.4 ทำการประเมินค่าใหม่ (Re-estimate) ของแต่ละหน่วยเสียงแบบจำลองฮิดเดนมาร์คอฟ ด้วยไฟล์ค่าลักษณะสำคัญทั้งหมดใหม่ ด้วยคำสั่ง HERest โดยจะทำการประเมินค่าใหม่ 3 ครั้ง เพื่อทำการปรับค่าพารามิเตอร์ต่างๆ ในแบบจำลองฮิดเดนมาร์คอฟ หลังจากการประเมินค่าแล้วจะได้แบบจำลองฮิดเดนมาร์คอฟของหน่วยเสียงเดี่ยว

1.2.5 เพื่อความถูกต้องแม่นยำมากขึ้น จะใช้หน่วยเสียงเรียงสาม (Triphone) แทนหน่วยเสียงเดี่ยว โดยหน่วยเสียงเดี่ยว นั้นคือ 1 แบบจำลองฮิดเดนมาร์คอฟ จะแทน 1 หน่วยเสียง โดยไม่สนใจหน่วยเสียงรอบข้าง (Context-Independent phone model) หน่วยเสียงเรียงสาม คือ 1 แบบจำลองฮิดเดนมาร์คอฟ จะแทน 3 หน่วยเสียง โดยจะสนใจหน่วยเสียงรอบข้าง (Context-dependent phone model) ตัวอย่างของหน่วยเสียงของคำว่า “การ กิน” คือ “k aa n^ k i n^” ภาพที่ 3.5 แสดงแบบจำลองหน่วยเสียงเดี่ยวของคำว่า “การกิน” ภาพที่ 3.6 แสดงแบบจำลองหน่วยเสียงเรียงสามของคำว่า “การกิน”

ภาพที่ 3.5
แบบจำลองหน่วยเสียงเดี่ยวของคำว่า “การกิน”




ภาพที่ 3.6
แบบจำลองหน่วยเสียงเรียงสามของคำว่า “การกิน”



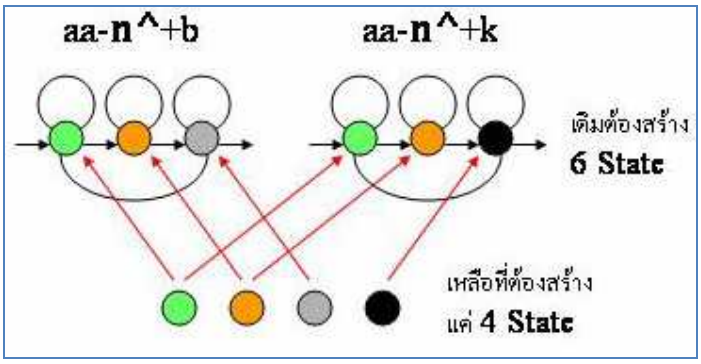

1.2.6 ทำการแปลงแบบจำลองหน่วยเสียงเดี่ยวให้เป็นแบบจำลองหน่วยเสียงเรียงสาม ด้วยคำสั่ง HLEd แล้วทำการสร้างรูปแบบของแบบจำลองฮิดเดนมาร์คอฟของหน่วยเสียงเรียงสาม (Triphone HMM) ด้วยคำสั่ง HHEd

1.2.7 ทำการประเมินค่าใหม่ของหน่วยเสียงเรียงสามอีก 3 ครั้ง หลังจากการประเมินค่าใหม่แล้วจะได้แบบจำลองฮิดเดนมาร์คอฟของหน่วยเสียงเรียงสาม

1.2.8 จะเห็นว่าหน่วยเสียงเรียงสามที่เกิดขึ้นนั้นมีมากมาย ในขณะที่แบบจำลองบางตัวนั้นเกิดขึ้นเพียงครั้งเดียว หน่วยเสียงเรียงสามบางตัวจะมีองค์ประกอบคล้ายๆกัน เช่น “aa-n^+b” ในคำว่า “k aa n^ b aa n” (การ บ้าน) กับ “aa-n^+k” ในคำว่า “k aa n^ k i n” (การ กิน) ดังนั้นทำการลดจำนวนการฝึกฝนให้น้อยลง โดยใช้แบบจำลองเสตทร่วมกัน (Tied-state model) แสดงดังภาพที่ 3.7 ในขั้นตอนการทำแบบจำลองเสตทร่วมกันนี้ไม่มีคำสั่งใน HTK ดังนั้นทางเนคเทคได้เขียนสคริปท์ genfulltri.pl เพื่อใช้ในขั้นตอนนี้

ภาพที่ 3.7

ตัวอย่างของแบบจำลองเสตทร่วมกัน




1.2.9 ทำการสร้างรูปแบบของแบบจำลองฮิดเดนมาร์คอฟเสตทร่วมกัน (Tied-state HMM) ด้วยคำสั่ง HHEd

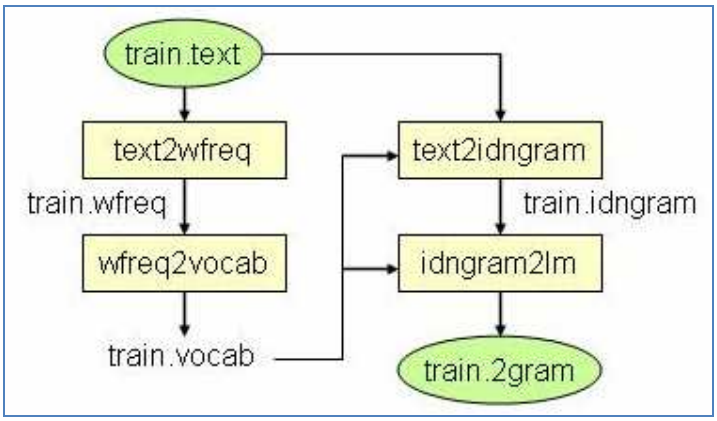
1.2.10 ทำการประเมินค่าใหม่ของแบบจำลองฮิดเดนมาร์คอฟเสตทร่วมกันอีก 3 ครั้ง หลังจากการประเมินค่าใหม่แล้วจะได้แบบจำลองฮิดเดนมาร์คอฟเสตทร่วมกัน ซึ่งหลังจากขั้นตอนนี้แล้วสามารถนำแบบจำลองไปใช้งานได้ แต่นักวิจัยยังสามารถพัฒนาให้ได้ผลดีมากขึ้น โดยการเพิ่มจำนวนขององค์ประกอบแบบเกาส์ (Gaussian mixture) เพื่อให้ค่ามีความต่อเนื่องมากขึ้น ทำการเพิ่มจำนวนขององค์ประกอบแบบเกาส์ จาก 1 ไป 2, จาก 2 ไป 4, จาก 4 ไป 8, จาก 8 ไป 16 ด้วยคำสั่ง HHed สุดท้ายจะได้แบบจำลองหน่วยเสียงของระบบ BN-LVCSR ซึ่งจะนำไปใช้ต่อไป

## 1.3 กำหนดแบบจำลองภาษา

สำหรับการฝึกฝนแบบจำลองภาษาจะใช้ CMU SLM ซึ่งเป็นเครื่องมือที่ใช้ในการเตรียมแบบจำลองภาษาเอ็นแกรม (N-gram Language model) การสร้างแบบจำลองภาษาเอ็นแกรมด้วย CMU-SLM มีขั้นตอน ดังภาพที่ 3.8

ภาพที่ 3.8

หลักการสร้าง 2-gram ด้วย CMU-SLM




1.3.1 แสดงคำที่มีทั้งหมดใน train.text พร้อมทั้งความถี่ของแต่ละคำด้วยคำสั่ง “text2wfreq < train.text > train.text” โดยผลลัพธ์ที่ได้ในขั้นตอนนี้จะเก็บอยู่ในไฟล์ train.wfreq

1.3.2 แสดงรายชื่อของคำศัพท์ (Vocabulary) ที่มีทั้งหมดใน train.text ด้วยคำสั่ง “wfreq2vocab < train.wfreq > train.vocab” โดยผลลัพธ์ที่ได้ในขั้นตอนนี้จะเก็บอยู่ในไฟล์ train.vocab

1.3.3 สร้างดัชนี (Index) ของ N-gram ที่เกิดใน train.text ด้วยคำสั่ง “text2idngram -vocab train.vocab -n 2 < train.text > train.idngram” ในที่นี้ N=2 คือ ไบแกรม (Bigram) โดยผลลัพธ์ที่ได้ในขั้นตอนนี้จะเก็บอยู่ในไฟล์ train.idngram

1.3.4 สร้างแบบจำลอง train.2gram ด้วยคำสั่ง “idngram2lm -vocab train.vocab -n 2 -idngram train.idngram -arpa train.2gram”

1.3.5 สร้างรูปแบบข้อความย้อนกลับ (Reverse text) สำหรับ train.text เพื่อการทำไตรแกรม (Trigram) ด้วยคำสั่ง “reverse.pl < train.text > train.revtext” โดยผลลัพธ์ที่ได้ในขั้นตอนนี้จะเก็บอยู่ในไฟล์ train.revtext

1.3.6 สร้างดัชนี (Index) ของ N-gram ที่เกิดใน train.revtext ด้วยคำสั่ง “text2idngram -vocab train.vocab -n 3 < train.revtext > train.revid3gram” ในที่นี้ N=3 คือ ไตรแกรม (Trigram) โดยผลลัพธ์ที่ได้ในขั้นตอนนี้จะเก็บอยู่ในไฟล์ train.revid3gram

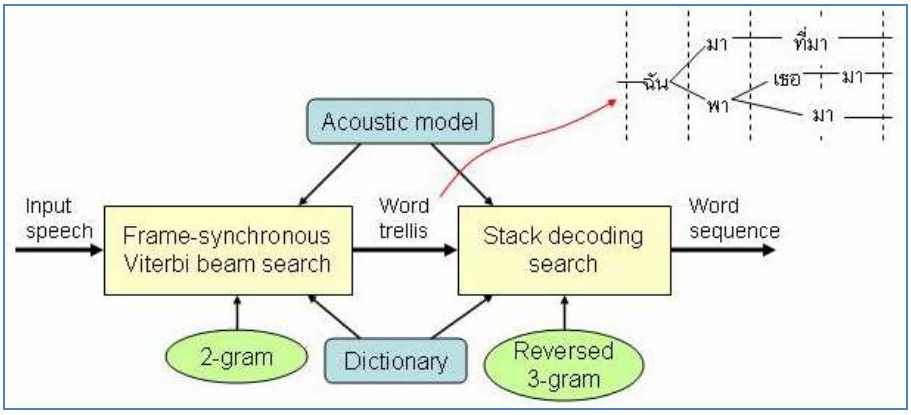
1.3.7 สร้างแบบจำลอง train.rev3gram ด้วยคำสั่ง “idngram2lm -vocab train.vocab -n 3 -idngram train.idngram -arpa train.rev3gram”

ซึ่งแบบจำลองภาษา train.2gram และ train.rev3gram จะนำมาใช้ในขั้นตอนการถอดรหัสเสียงต่อไป

## 1.4 **การถอดรหัสเสียง** (Speech decoder)

สำหรับการถอดรหัสเสียงนั้นจะใช้ JULIUS ซึ่งเป็นเครื่องมือในการถอดรหัสเสียงแบบหลายขั้น (Multi-pass) ขั้นตอนในการถอดรหัสเสียงนั้น แสดงดังภาพที่ 3.9

ภาพที่ 3.9
แสดงการถอดรหัสเสียงโดยใช้ JULIUS




ขั้นที่ 1 (1st pass)

ใช้ train.2gram และ Viterbi search ในการสร้างโครงสร้างคำ (word trellis) ซึ่งก็คือลำดับของคำ (Word sequence) ที่เป็นไปได้ในทุกเส้นทาง (path) ถ้าเส้นทางไหนมีค่าความน่าจะเป็นน้อยกว่าค่าเริ่มต้น (Threshold) ที่กำหนด ก็จะไม่อยู่ในโครงสร้างคำ เป็นการจำกัดขอบเขตของการค้นหาในขั้นที่ 2 ต่อไป

ขั้นที่ 2 (2nd pass)

ใช้ train.rev3gram คำนวณค่าความน่าจะเป็นของลำดับของคำด้วย Stack search และจะได้ลำดับของคำที่มีค่าความน่าจะเป็นสูงที่สุด

### 1.5 การฝึกฝนข้อมูลหน่วยเสียงสำหรับระบบ BN-LVCSR

ในการทดลองนี้ทำการฝึกฝนข้อมูลสำหรับระบบ BN-LVCSR ให้เข้ากับชุดข้อมูลการทดสอบ ซึ่งส่วนของแบบจำลองเสียงนั้น เบื้องต้นผู้วิจัยได้ทำการทดสอบชุดข้อมูล LOTUS Corpus ขนาดเล็ก และ LOTUS รวมกับคลังข้อมูลข่าวภาษาไทย (BN Corpus)

ตารางที่ 3.1

ตารางเปรียบเทียบเปอร์เซ็นต์ความถูกต้องของ LOTUS และ BN+LOTUS

| ข้อมูลที่ใช้ฝึกฝนแบบจำลองหน่วยเสียง | เปอร์เซ็นต์ความถูกต้อง |
|---|---|
| LOTUS Corpus | 55.23% |
| BN+LOTUS Corpus | 43.19% |

จากตารางที่ 3.1 แสดงเปอร์เซ็นต์ความถูกต้องระหว่าง LOTUS และ BN+LOTUSผลลัพธ์ที่ได้แสดงให้เห็นว่าข้อมูล LOTUS Corpus อย่างเดียวจะให้ผลความถูกต้องมากกว่า LOTUS รวมกับคลังข้อมูลข่าวภาษาไทย เพราะลักษณะของเสียงในชุดข้อมูลการทดสอบคล้ายคลึงกับ LOTUS Corpus ซึ่งเป็นแบบข้อมูลจากการอ่าน (Reading speech) ส่วนคลังข้อมูลข่าวภาษาไทยเป็นข้อมูลแบบสนทนา (Spontaneous speech) ซึ่งจะมีเสียงรบกวน (noise) มากทำให้มีผลต่อการฝึกฝนของระบบ จากผลลัพธ์ดังกล่าวจะเลือก LOTUS Corpus ความยาว 50 ชั่วโมง มาใช้ในการสร้างและฝึกฝนแบบจำลองหน่วยเสียง

### 1.6 กำหนดแบบจำลองภาษาที่เหมาะสม

เลือกชุดข้อมูลแบบต่างๆในการทดสอบกับแบบจำลองหน่วยเสียงที่ได้จากข้อ 1.5 ข้อมูลที่ใช้สำหรับฝึกฝนแบบจำลองภาษาจะใช้คลังข้อความของ LOTUS Corpus และ คลังข้อมูลข่าวภาษาไทย โดยคลังข้อมูลข่าวภาษาไทยที่นำมาใช้นี้ได้นำเอาคำที่มีคำอ่านอยู่ในพจนานุกรมเสียงอ่าน เช่น คำว่า [laugh] ซึ่งเป็นเสียงหัวเราะในข่าว โดย

จะนำเอาออกไปทั้งประโยค ดังนั้นจะเหลือข้อมูลของคลังข้อมูลข่าวภาษาไทย 9,115 ประโยค จากทั้งหมด 13,044 ประโยค

ในการหาแบบจำลองภาษาที่เหมาะสมนั้นจะทำการเปรียบเทียบข้อมูลที่แตกต่างกัน 3 แบบ เพื่อที่จะหาว่าข้อมูลในการฝึกฝนแบบจำลองภาษานั้น ควรจะใช้คลังข้อความของ LOTUS Corpus, คลังข้อมูลข่าวภาษาไทยเพียงอย่างเดียว หรือ LOTUS Corpus รวมกับคลังข้อมูลข่าวภาษาไทย ที่จะเหมาะสมกับแบบจำลองภาษามากที่สุด

ข้อมูลที่ใช้ฝึกฝนแบบจำลองภาษาจะเปรียบเทียบข้อมูลที่แตกต่างกัน 3 แบบ ได้แก่

1. ข้อมูลของ LOTUS Corpus ซึ่งประกอบด้วย 20,984 ประโยค 390,584 คำ 5,412 คำศัพท์ โดยแบบจำลองนี้เป็นระบบต้นแบบ (Baseline system)
2. ข้อมูลของคลังข้อมูลข่าวภาษาไทยซึ่งประกอบด้วย 9,115 ประโยค 170,219 คำ 7,123 คำศัพท์
3. ข้อมูลของ LOTUS และ คลังข้อมูลข่าวภาษาไทย ซึ่งประกอบด้วย 30,099 ประโยค 560,803 คำ 9,602 คำศัพท์

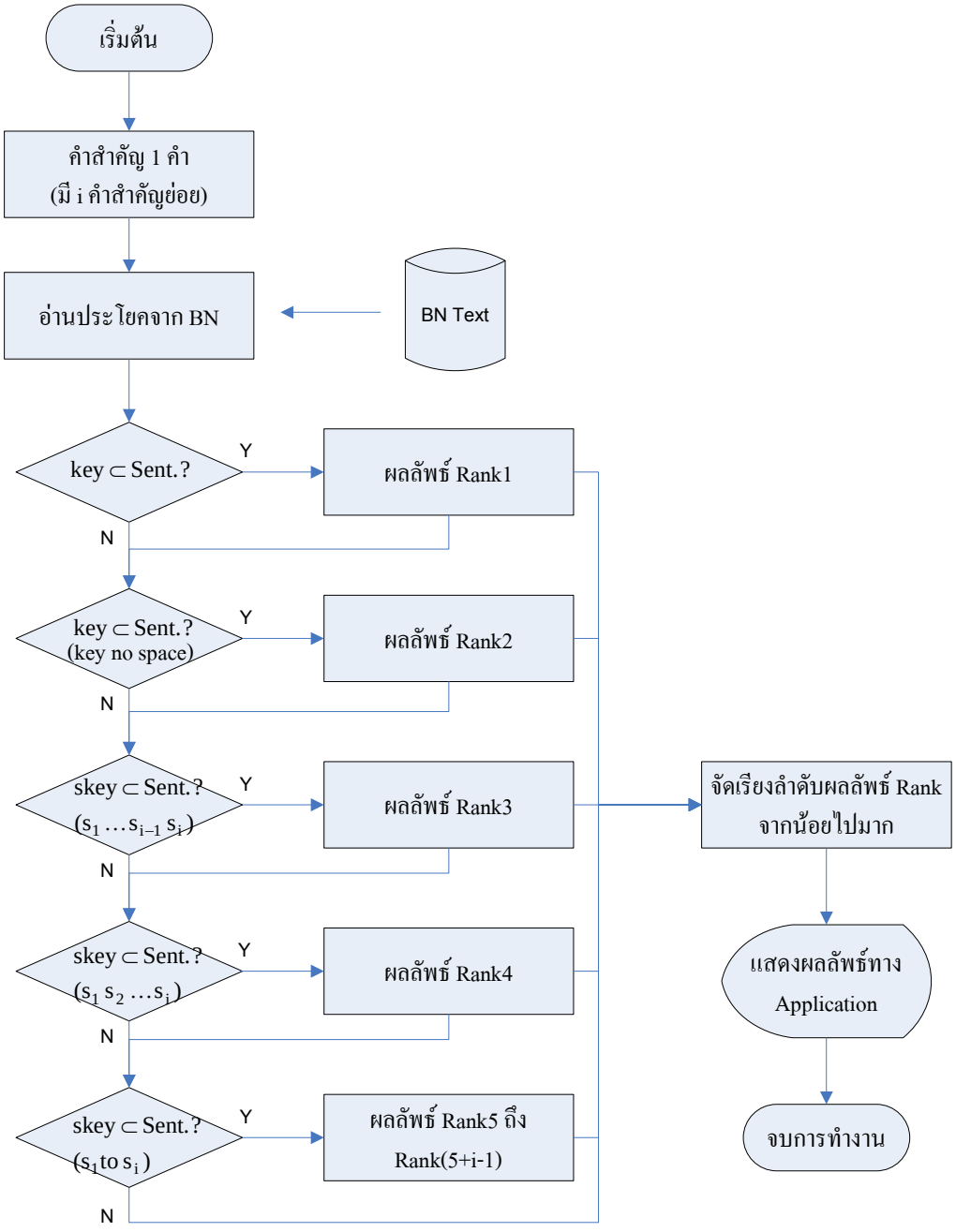
## 2. ขั้นตอนการค้นหาโดยใช้คำสำคัญ

จากขั้นตอนการสร้างระบบ BN-LVCSR เราจะได้รายการของคำที่ถอดรหัสมาจากข้อมูลเข้าที่เป็นเสียงจำนวน 10 คำ ซึ่งใช้เป็นคำสำคัญในการค้นหาข้อมูลในขั้นต่อไป

ในแต่ละคำสำคัญที่ได้จากระบบ LVCSR นั้น อาจจะแยกเป็นคำสำคัญย่อย (Sub Keyword) ได้หลายคำ เช่น คำสำคัญว่า “กระทรวง ศึกษาธิการ” จะมีคำสำคัญย่อย 2 คำ คือ คำว่า “กระทรวง” และ “ศึกษาธิการ”

ขั้นตอนการค้นหาคำสำคัญ 1 คำในคลังข้อมูลข่าวภาษาไทย จะแสดงได้ดังภาพที่ 3.10

ภาพที่ 3.10
ขั้นตอนการค้นหาคำสำคัญ1 คำในคลังข้อมูลข่าวภาษาไทย




จากภาพที่ 3.10 แสดงการทำงานของการค้นหาคำสำคัญ 1 คำ (จะแทนด้วย key) ในคำสำคัญอาจจะมีคำสำคัญย่อย (จะแทนด้วย skey) แล้วนำไปค้นหาในคลังข้อมูลข่าวภาษาไทย

(จะแทนด้วย Sent.) โดยผลลัพธ์ที่ได้ในแต่ละเงื่อนไข Rank1, 2...n คือความสำคัญในการค้นหาเป็นอันดับที่ 1, 2...n ตามลำดับ แล้วจะนำมาเรียงลำดับอีกครั้งก่อนที่จะแสดงในแอพพลิเคชั่น ยกตัวอย่างการค้นหาคำสำคัญว่า “กระทรวง พาณิชย์”

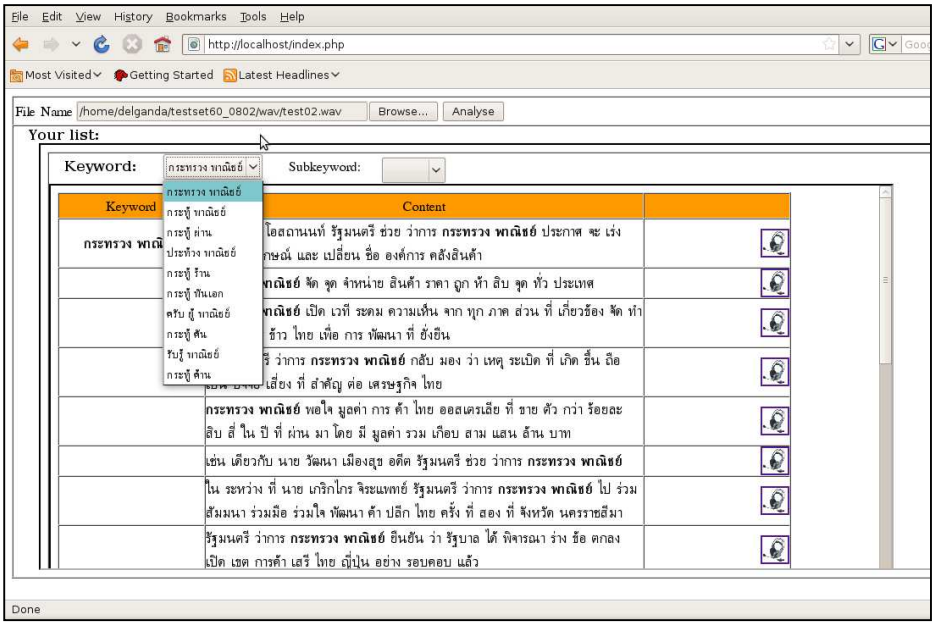
Rank 1 คือการค้นหาคำว่า “กระทรวง พาณิชย์”

Rank 2 คือการค้นหาคำว่า “กระทรวงพาณิชย์”

Rank 3 คือการค้นหาคำว่า “กระทรวง” และการค้นหาคำว่า “พาณิชย์”

ภาพที่ 3.11

ผลลัพธ์จากการค้นหาคำสำคัญว่า “กระทรวง พาณิชย์”




จากภาพที่ 3.11 แสดงตัวอย่างของการค้นหาข้อมูลด้วยเสียง โดยให้ข้อมูลเข้าเป็นไฟล์เสียง ในตัวอย่างนี้จะแสดงการค้นหาคำว่า “กระทรวง พาณิชย์” ซึ่งจะได้ผลลัพธ์ของการค้นหาไฟล์เสียงดังภาพ

## การประเมินผลการทดลอง

สำหรับค่าที่จะใช้เปรียบเทียบในการรู้จำเสียง จะใช้ เปอร์เซ็นต์ความถูกต้อง (% Correct) และ เปอร์เซ็นต์ความแม่นยำ (% Accuracy)

เปอร์เซ็นต์ความถูกต้อง (% Correct)

$$\%Correct = \frac{N - S - D}{N} \times 100$$

เปอร์เซ็นต์ความแม่นยำ (% Accuracy)

$$\%Accuracy = \frac{N - S - D - I}{N} \times 100$$

ซึ่ง N – จำนวนคำทั้งหมด

S – จำนวนคำที่มีอยู่แทนที่ด้วยคำอื่น

D – จำนวนคำที่หายไป

I – จำนวนคำอื่นที่แทรกเข้ามา

## สมมติฐานของงานวิจัย

การสืบค้นข้อมูลด้วยเสียงสำหรับข่าวภาษาไทยนี้จะอำนวยความสะดวกให้แก่ผู้ที่ต้องการหาข้อมูลแต่ไม่สามารถพิมพ์ข้อความได้ เช่น บุคคลทั่วไป คนพิการ หรือผู้ป่วย