

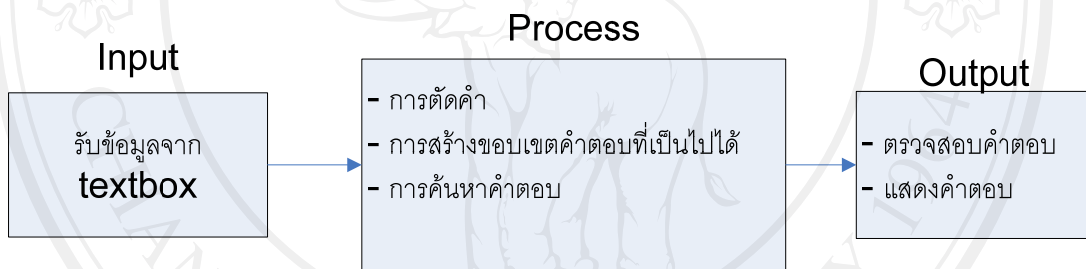
บทที่ 3

การออกแบบระบบงานและฐานข้อมูล

ระบบการพัฒนาต้นแบบแชตโรบอทภาษาไทย กรณีศึกษา เรื่องอัจฉริยะ เป็นระบบที่ใช้การนำเข้าข้อมูลที่เป็นอักขระ เพื่อเข้าสู่กระบวนการตัดคำ และเข้าสู่กระบวนการค้นหา โดยจะแบ่งงานออกเป็นสามส่วนหลัก ดังนี้

- 3.1 ภาพรวมของระบบที่จะทำการพัฒนา
- 3.2 การออกแบบผังงานระบบ
- 3.3 การออกแบบฐานข้อมูล

3.1 ภาพรวมของระบบที่จะทำการพัฒนา



รูป 3.1 ภาพรวมขั้นตอนการทำงานของระบบ

ในการพัฒนาต้นแบบแชตโรบอทภาษาไทย จะมีคำศัพท์ที่เตรียมไว้แล้วเบื้องต้นจำนวน 4,990 คำ โดยเตรียมคำศัพท์จากฐานความรู้ทั้งหมดที่มีในระบบเพื่อช่วยในการตัดคำ และมีฐานข้อมูลความรู้จำนวน 1,137 รายการ ซึ่งการทำงานแบ่งออกเป็น 3 ส่วนประกอบด้วย

- 1) การรับข้อมูลนำเข้า จะรับข้อมูลผ่านทางแป้นพิมพ์อักขระ
- 2) การประมวลผล ระบบจะทำการตัดคำจากข้อความที่รับเข้ามาโดยเปรียบเทียบกับคำศัพท์ที่เตรียมไว้ แล้วนำไปประมวลผล เพื่อสร้างขอบเขตคำตอบ และค้นหาคำตอบ
- 3) การตรวจสอบคำตอบและแสดงผลคำตอบ เป็นการตรวจสอบคำตอบแบบย้อนรอยกลับว่าเป็นคำตอบที่ถูกต้องหรือไม่ จากนั้นจะเตรียมข้อความเพื่อแสดงผลคำตอบต่อไป

3.1.1 การรับข้อมูลนำเข้า

การนำข้อมูลนำเข้าสู่ระบบจะใช้วิธีนำเข้าข้อมูลตัวอักษร (Text) ผ่านทางแป้นพิมพ์อักษร (Keyboard) จากนั้นจะนำข้อความที่ได้เข้าสู่กระบวนการตัดคำ (Word Segmentation) ในขั้นตอนต่อไป

3.1.2 การประมวลผล

การประมวลผลเป็นส่วนของโปรแกรมที่ทำการพัฒนา ซึ่งในขั้นตอนนี้จะนำข้อมูลนำเข้ามาประมวลผล โดยมีขั้นตอนดังนี้

1. การตัดคำ เป็นส่วนของการตัดคำจากข้อมูลที่ได้รับเข้ามาจากผู้ใช้งานระบบ โดยใช้วิธีการตัดคำแบบเทียบคำที่ยาวที่สุดก่อน เพื่อให้ได้คำและความหมายตามที่ต้องการ ส่วนคำที่ไม่สามารถตัดได้ในขั้นตอนนี้จะนำไปตัดในขั้นตอนการตัดด้วยกฎของภาษาไทยอีกครั้ง
2. การสร้างขอบเขตคำตอบที่เป็นไปได้ ขั้นตอนนี้จะนำคำที่ตัดได้ทั้งหมดไปค้นหาจากฐานข้อมูลเรื่องอัญมณี และเก็บรายการที่ค้นพบทั้งหมดลงในตารางเก็บข้อมูลชั่วคราว
3. การสร้างคำตอบ เป็นส่วนของการให้คะแนนรายการที่พบโดยทุกๆ คำที่พบจะให้คะแนน 1 คะแนน และรายการที่มีคะแนนสูงสุดคือคำตอบที่น่าจะเป็นไปได้ที่สุด หรือถ้ามีรายการที่มีคะแนนสูงสุดเท่ากันหลายรายการ จะนำผลลัพธ์ที่ได้เข้าสู่ขั้นตอนต่อไป

3.1.3 การตรวจสอบคำตอบและแสดงคำตอบ

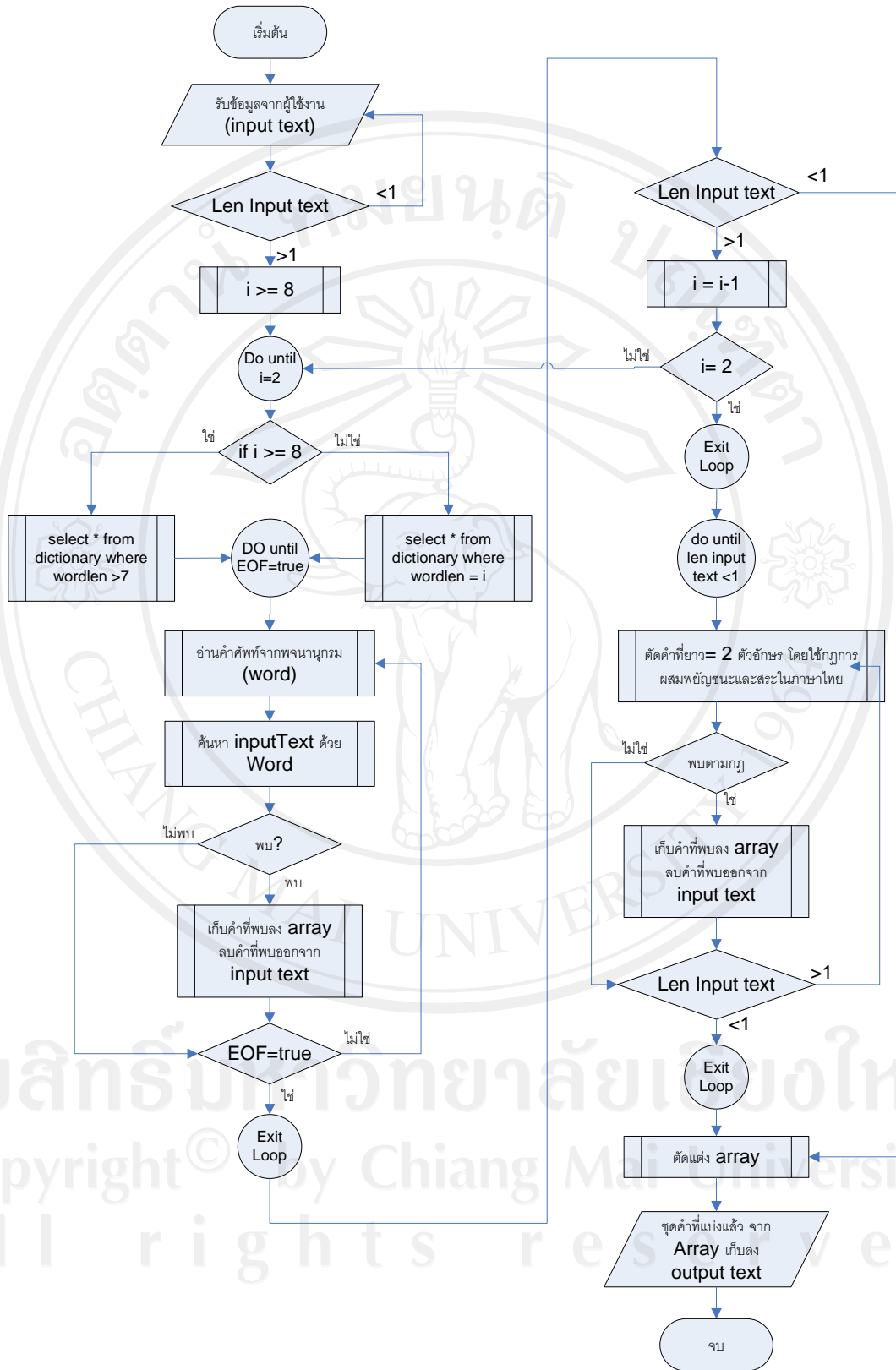
การตรวจสอบคำตอบ เป็นขั้นตอนการตรวจสอบว่าคำตอบที่ได้จะเป็นคำตอบที่ถูกต้องหรือไม่ ด้วยการตรวจสอบย้อนกลับ ว่าจำนวนคำที่พบในรายการที่ได้คะแนนสูงสุดเท่ากับ หรือ น้อยกว่า จำนวนคำที่ตัดได้ทั้งหมดมากน้อยเท่าใด หากเป็นคำตอบที่ถูกต้องและยอมรับได้ก็จะแสดงคำตอบจากรายการที่ค้นพบ แต่ถ้าเป็นคำตอบที่ยอมรับไม่ได้ หรือไม่มีคำตอบก็จะแสดงเป็นข้อความว่าไม่พบคำตอบ

3.2 การออกแบบผังงานระบบ

เป็นแผนภาพที่แสดงถึงผังงานระบบ ประกอบด้วยส่วนต่างๆ ดังต่อไปนี้

3.2.1 ส่วนการตัดคำ

การตัดคำจะใช้ 2 วิธีร่วมกัน คือการตัดคำโดยวิธีเทียบคำที่ยาวที่สุดก่อนและการตัดคำโดยกฎภาษาไทย ซึ่งคำที่มีมากกว่า 2 ตัวอักษรจะตัดโดยใช้วิธีการเทียบคำที่ยาวที่สุดก่อน จากนั้นจึงตัดคำจากประโยคที่เหลือด้วยกฎภาษาไทย



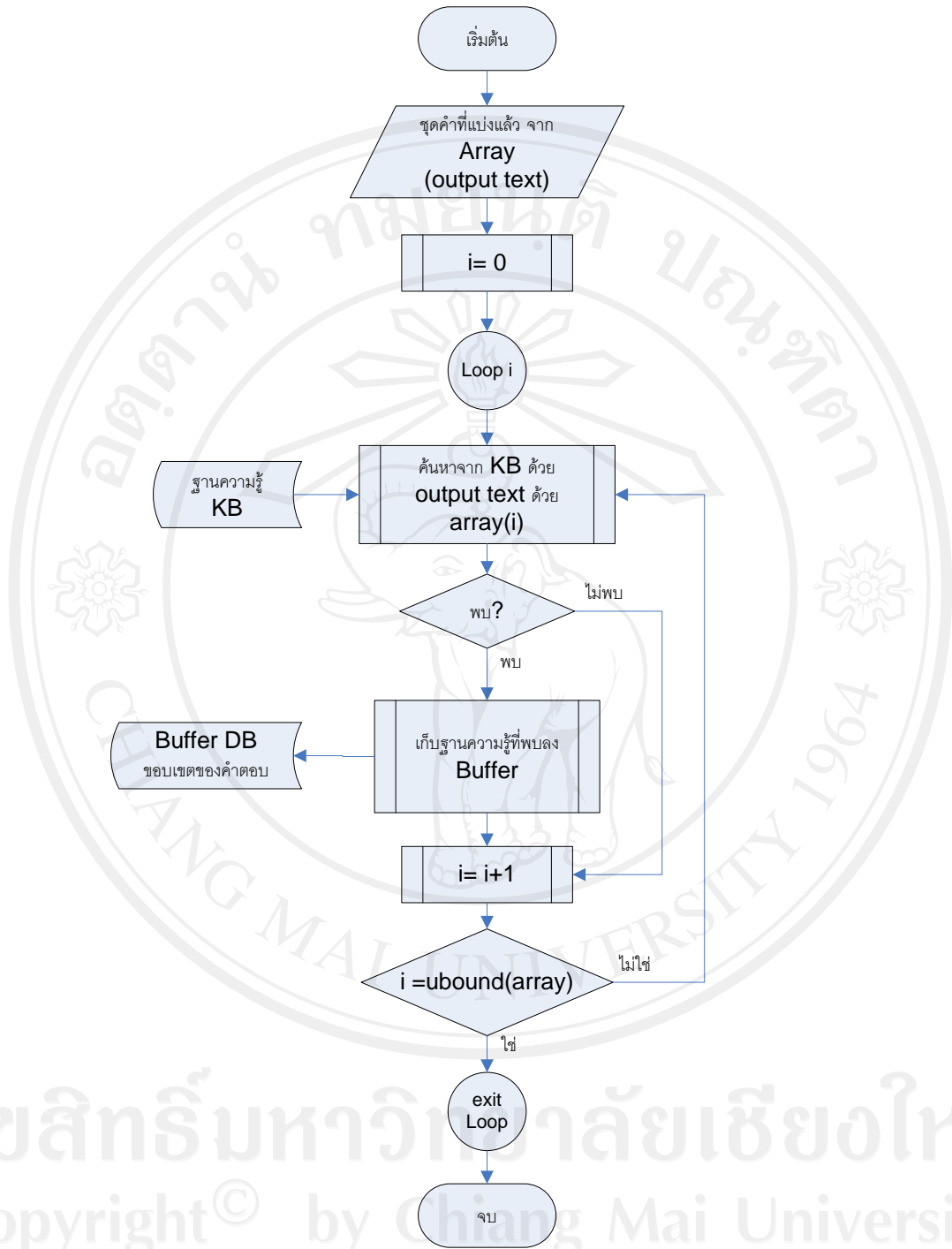
รูป 3.2 แสดงขั้นตอนการตัดคำ

จากรูป 3.2 เมื่อผู้ใช้งานป้อนข้อมูลคำถามที่เป็นข้อความเข้าสู่ระบบ ระบบจะทำงานดังนี้

- ตรวจสอบความยาวข้อความที่รับเข้ามาจากผู้ใช้งาน (input text) ถ้ามีมากกว่า 1 ตัวอักษร ระบบจะเริ่มทำงาน เช่นผู้ใช้งานป้อนข้อมูล “เพชรมีสีอะไรได้บ้าง”
- ระบบจะอ่านคำศัพท์จากฐานข้อมูล โดยเริ่มอ่านจากคำที่มีความยาวมากกว่า 8 ตัวอักษรทีละคำ แล้วนำไปเปรียบเทียบกับข้อความที่ผู้ใช้งานป้อนเข้ามา ถ้าพบจะเก็บคำที่พบลงหน่วยความจำ จากนั้นลบคำที่พบออกจากข้อความของผู้ใช้งาน ถ้าไม่พบก็จะไปสู่การตรวจสอบในรอบถัดไป โดยทำซ้ำจนกว่าความยาวข้อความของผู้ใช้งานที่ป้อนเข้ามาน้อยกว่า 1 หรือ เมื่อทำซ้ำไปจนถึงการอ่านคำศัพท์ที่มีความยาวเท่ากับ 3 ตัวอักษร ซึ่งในขั้นตอนนี้จะสามารถตัดคำจากประโยคที่ผู้ใช้งานป้อนเข้ามาได้เป็น |เพชร|มี|สี|อะไร|ได้|บ้าง| (ให้ “|” แทนสัญลักษณ์การตัดคำ) โดยคำว่า มี และ สี ไม่สามารถตัดได้ในขั้นตอนนี้ เนื่องจากในพจนานุกรมในการศึกษานี้ไม่มีคำว่า มี สี
- เมื่อระบบอ่านคำจากฐานข้อมูลเพื่อนำมาตรวจสอบจนถึงความยาวเท่ากับสามตัวอักษร และข้อความที่ผู้ใช้งานป้อนเข้ามายังมีความยาวมากกว่า 1 ตัวอักษร จากตัวอย่างประโยค “เพชรมีสีอะไรได้บ้าง” คำที่ยังตัดไม่ได้คือ มีสี ระบบตัดคำจะใช้กฎภาษาไทยในการช่วยตัดโดยคำว่า “มี” จะใช้กฎการผสมพยัญชนะและสระ ซึ่งจะตัดได้ดังนี้ |มี|สี|
- ผลลัพธ์ที่ได้จากขั้นตอนนี้คือ ชุดคำที่แบ่งจากข้อความที่ผู้ใช้งานป้อนเข้ามา โดยจะเรียกว่า ชุดคำที่ได้จากการตัดคำ จากตัวอย่างคือ |เพชร|มี|สี|อะไร|ได้|บ้าง|

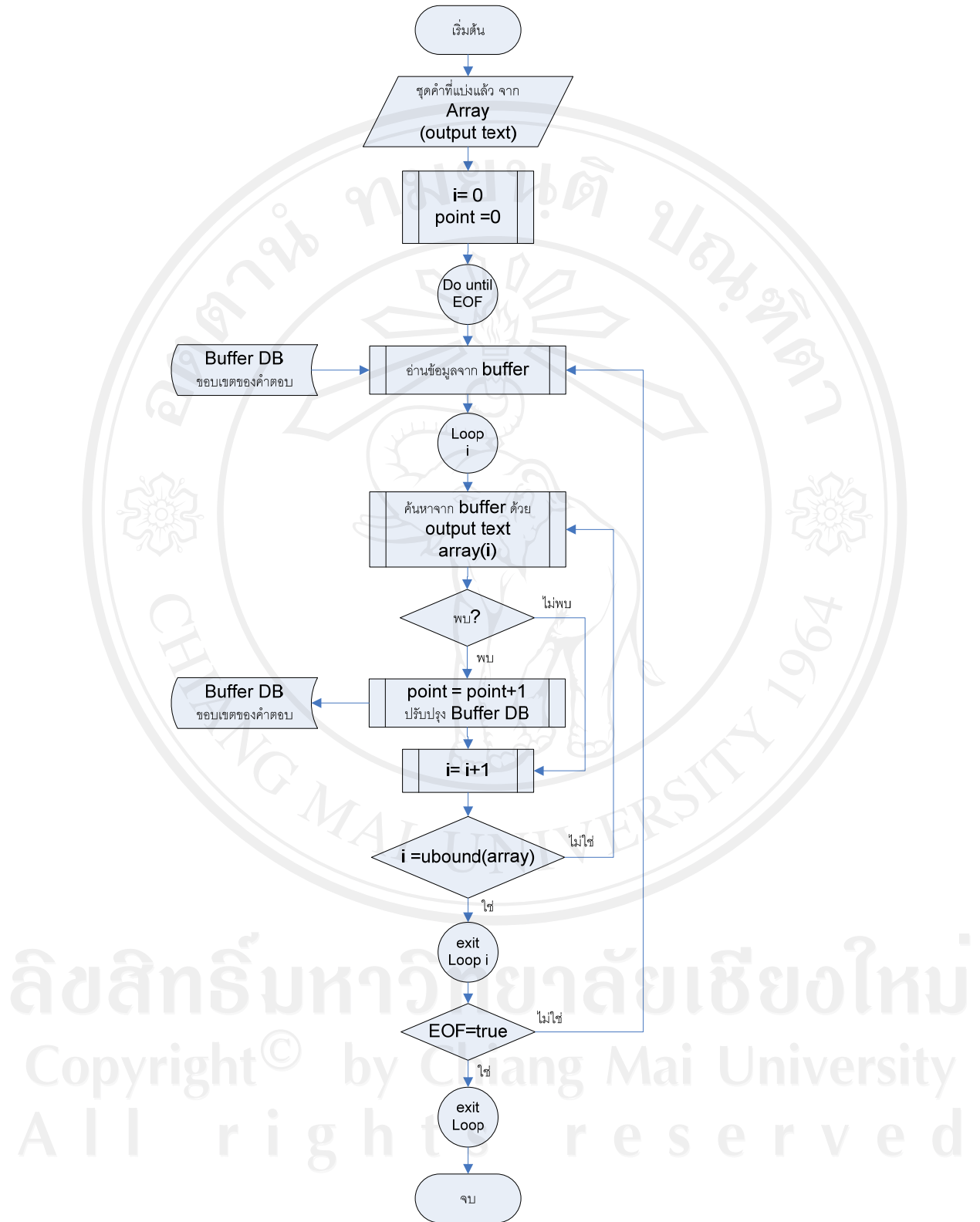
3.2.2 ส่วนการสร้างขอบเขตคำตอบ

จากรูป 3.3 หลังจากได้ชุดคำที่ได้จากการตัดคำแล้ว “|เพชร|มี|สี|อะไร|ได้|บ้าง|” จะนำคำที่ได้ทุกคำไปค้นหาจากฐานข้อมูลฐานความรู้ แล้วสร้างเป็นตารางฐานข้อมูลชั่วคราว เพื่อเป็นขอบเขตคำตอบ โดยนำคำว่า เพชร ไปค้นหาจากคำฐานข้อมูลอัญมณี ถ้าพบก็จะเก็บทุกรายการที่พบลงในตารางฐานข้อมูลชั่วคราว จากนั้นก็จะนำคำที่เหลือไปค้นหาและเก็บลงตารางฐานข้อมูลชั่วคราวเช่นเดียวกับคำว่า เพชร



รูป 3.3 แสดงขั้นตอนการสร้างขอบเขตคำตอบ

3.2.3 ส่วนการสร้างคำตอบ



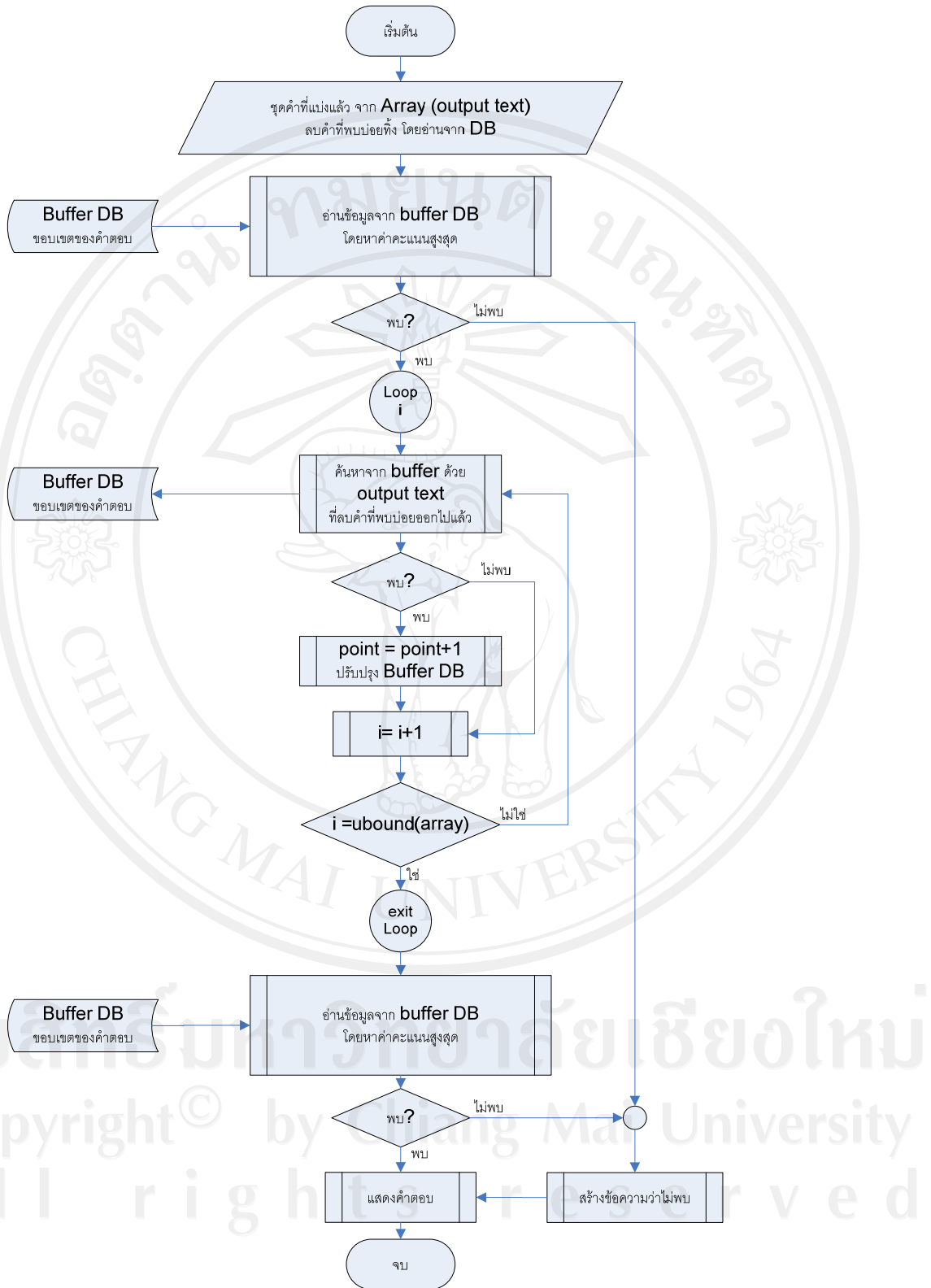
รูป 3.4 แสดงขั้นตอนการสร้างคำตอบ

จากรูป 3.4 เมื่อได้ขอบเขตของคำตอบทั้งหมด ระบบจะทำการสร้างคำตอบโดยการอ่านคำ จากชุดคำที่ได้จากการตัดคำทีละคำ แล้วนำมาเทียบกับคำตอบจากรางฐานข้อมูลชั่วคราวทีละ รายการ ถ้าพบจะให้คะแนนรายการนั้นเป็น 1 ถ้าไม่พบจะให้คะแนนเป็น 0 และทำซ้ำจนกว่าจะหมด ซึ่งสันนิษฐานได้ว่ารายการใดที่ได้คะแนนสูงสุดรายการนั้นอาจจะเป็นคำตอบ โดยจะนำ รายการที่ได้ไปเข้าขั้นตอนการตรวจสอบคำตอบต่อไป และในกรณีที่มีคะแนนสูงสุดเท่ากัน หลาย รายการ ก็จะนำไปเข้าขั้นตอนการตรวจสอบคำตอบเช่นเดียวกัน

3.2.4 ส่วนการตรวจสอบและแสดงคำตอบ

จากรูป 3.5 ในขั้นตอนนี้จะเป็นการตรวจสอบคำตอบที่ได้คะแนนสูงสุด ว่าจะเป็นการตอบที่ ถูกต้องหรือไม่โดยมีขั้นตอนดังนี้

- อ่านฐานข้อมูลคำที่เป็นคำสร้อยหรือคำที่พบบ่อย เช่น ครับ ค่ะ จ๊ะ นะจ๊ะ แล้วนำมา เปรียบเทียบกับชุดคำที่ได้จากการตัดคำ ถ้าพบจะตัดคำนั้นออกไป
- นำคำที่เหลือมาค้นหาจากรางฐานความรู้ชั่วคราวที่มีคะแนนสูงสุด ถ้าพบจะ ไปสู่ ขั้นตอนการตรวจสอบแบบย้อนรอยกลับไป ถ้าไม่พบจะออกไปยังขั้นตอนการสร้างคำตอบว่า ไม่พบ ซึ่งในการศึกษานี้จะใช้ข้อความว่า “ไม่ทราบคำตอบ แต่คุณสามารถหาคำตอบเพิ่มเติมได้จาก ลิงค์นี้ คลิ๊กที่นี่” โดยจะเชื่อมโยงลิงค์กลับเครื่องมือค้นหาของ google.com
- เมื่อพบคำตอบที่ได้คะแนนสูงสุด 1 รายการ หรือมากกว่า ระบบจะตรวจสอบคำตอบอีก ครั้งด้วยการ นำชุดคำที่ลบคำที่ใช้บ่อยออกไปแล้ว มาค้นหาจากรายการในตารางฐานความรู้ ชั่วคราว โดยค้นหาทีละคำ ถ้าพบจะให้คะแนนเพิ่มเป็นบวก 1 ถ้าไม่พบบวกเพิ่มเป็น 0
- ค้นหารายการที่ได้คะแนนสูงสุด ถ้าพบก็จะแสดงคำตอบตามรายการนั้น แต่ถ้าไม่พบจะ แสดงคำตอบ “ไม่ทราบคำตอบ แต่คุณสามารถหาคำตอบเพิ่มเติมได้จากลิงค์นี้ คลิ๊กที่นี่”



รูป 3.5 แสดงขั้นตอนการตรวจสอบคำตอบและการแสดงคำตอบ

3.3 การออกแบบระบบฐานข้อมูล

กระบวนการออกแบบฐานข้อมูลของการพัฒนาต้นแบบแชตโรบอทภาษาไทย ผู้ศึกษาได้กำหนดชนิดของข้อมูลแอสเซตที่ใช้ในฐานข้อมูลของระบบฯ ดังแสดงในตาราง 3.1 ดังนี้

ตาราง 3.1 ชนิดของข้อมูลแอสเซต ที่ใช้ในฐานข้อมูลของระบบ

ชนิดของข้อมูล	คำอธิบาย	ตัวอย่าง
Text	ข้อมูลที่เป็น String ขนาดเก็บมีขนาดไม่เกิน 255 ตัวอักษร	Bangkok
Number	เก็บค่าจำนวนเต็ม มีตั้งแต่ -2,147,483,648 ถึง +2,147,483,647	12345
Memo	ข้อมูลที่เป็น String ขนาดเก็บมีขนาดไม่เกิน 65,536 ตัวอักษร	โครงสร้าง
AutoNumber	ข้อมูลที่สร้างขึ้นอัตโนมัติ	123

ตาราง 3.2 รายละเอียดเพิ่มข้อมูลทั้งหมด

ลำดับ	ชื่อเพิ่มข้อมูล	ประเภทตารางข้อมูล	รายละเอียด
1	BotCreator	Reference Table	ข้อมูลผู้สร้างโรบอต
2	BrainTemp	Master Table	ข้อมูลขอบเขตของคำตอบ
3	ChatHistory	Transaction Table	ข้อมูลประวัติการสนทนา
4	knowledge	Master Table	ข้อมูลฐานความรู้
5	SegLog	Reference Table	ข้อมูลประวัติการตัดคำ
6	vocab	Master Table	ข้อมูลคำศัพท์
7	vocabNotuse	Master Table	ข้อมูลคำศัพท์ที่ไม่นำมาตัดคำ

ตาราง 3.3 รายละเอียดเพิ่มข้อมูล BotCreator

ชื่อเพิ่มข้อมูล	BotCreator			
คำอธิบาย	ข้อมูลผู้สร้างโรบอต			
คีย์หลัก	BotId			
คีย์นอก	-			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
BotId	Text	20	รหัสของโรบอต	200567432
Email	Text	50	อีเมลของผู้สร้างโรบอต	narong@yahoo.com
Password	Text	20	รหัสผ่านของผู้สร้างบอท	1234
Nickname	Text	30	ชื่อเข้าใช้	Ezeman
Name	Text	50	ชื่อผู้สร้างโรบอต	ณรงค์
Surname	Text	50	นามสกุลผู้สร้างโรบอต	โคตรสกิจ
Address	Text	50	ที่อยู่ผู้สร้างโรบอต	95/8 หมู่ 2
Tambon	Text	50	ตำบลผู้สร้างโรบอต	สันผักหวาน
Amphor	Text	50	อำเภอผู้สร้างโรบอต	หางดง
Changwat	Text	50	จังหวัดผู้สร้างโรบอต	เชียงใหม่
Admin	Text	1	สถานะของผู้ดูแลระบบ	0 หมายถึงผู้ใช้ที่ ลงทะเบียน 1 หมายถึงผู้ดูแลระบบ
Botname	Text	50	ชื่อ โรบอต	นายรอบรู้
Botdescription	Text	255	คำบรรยายคุณสมบัติของ โรบอต	เชี่ยวชาญเรื่องอัญมณี
Xactive	Text	1	สถานะการสร้างบอท	0 หมายถึงยังไม่ได้ สร้างบอท 1 หมายถึงสร้างบอท แล้ว

ตาราง 3.3 รายละเอียดเพิ่มเติมข้อมูล BotCreator (ต่อ)

Rstatus	Text	1	สถานะของบอทที่สร้างแล้ว	0 หมายถึงสถานะของ โรบอทไม่พร้อมใช้ งาน 1 หมายถึงสถานะ ของโรบอทพร้อมใช้ งาน
userActive	Text	1	สถานะของผู้สร้างโรบอต	1

ตาราง 3.4 รายละเอียดเพิ่มเติมข้อมูล BrainTemp

ชื่อเพิ่มข้อมูล	BrainTemp			
คำอธิบาย	ข้อมูลบัพเฟอ์ของคำตอบ			
คีย์หลัก	KbID			
คีย์นอก	BotID,SessID			
หมายเหตุ	sessID รับค่ามาจาก session ของระบบ			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
KbID	Number	8	รหัสของฐานความรู้	16
question	Text	255	คำถามของฐานความรู้	เพชรมีสีอะไรบ้าง
answer	Memo	65,536	คำตอบของฐานความรู้	มีสีขาว สีน้ำเงิน
BotId	Text	20	รหัสของโรบอต	200567432
SessID	Text	30	รหัส session ของผู้ใช้	906132741
totalpoint	Number	4	ข้อมูลคะแนนของคำตอบ	5

ตาราง 3.5 รายละเอียดเพิ่มเติมข้อมูล ChatHistory

ชื่อเพิ่มข้อมูล	ChatHistory			
คำอธิบาย	ข้อมูลประวัติการสนทนา			
คีย์หลัก	ChatID			
คีย์นอก	BotId			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
ChatID	AutoNumber	8	รหัสฐานความรู้	2
ChatQuestion	Text	255	คำถามของฐานความรู้	เพชรมีสีอะไรบ้าง
ChatAnswer	Memo	65,536	คำตอบของฐานความรู้	มีสีขาว สีนํ้าเงิน
Chater	Text	20	ชื่อของผู้ถาม	You
ChatSession	Text	30	รหัส session ของระบบ	906132741
BotId	Text	20	รหัสของโรบอต	200567432
ChatDate	Text	30	วันที่และเวลา	7/29/2009 11:29:10 PM

ตาราง 3.6 รายละเอียดเพิ่มเติมข้อมูล knowledge

ชื่อเพิ่มข้อมูล	knowledge			
คำอธิบาย	ข้อมูลฐานความรู้			
คีย์หลัก	KbId			
คีย์นอก	BotId			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
KbId	AutoNumber	8	รหัสฐานความรู้	2
question	Text	255	คำถามของฐานความรู้	เพชรมีสีอะไรบ้าง
answer	Memo	65,536	คำตอบของฐานความรู้	มีสีขาว สีนํ้าเงิน
BotId	Text	20	รหัสของโรบอต	906132741

ตาราง 3.7 รายละเอียดเพิ่มข้อมูล SegLog

ชื่อเพิ่มข้อมูล	SegLog			
คำอธิบาย	ข้อมูลประวัติการตัดคำ			
คีย์หลัก	SegmentLogId			
คีย์นอก	-			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
SegmentLogId	AutoNumber	8	รหัสของประวัติการตัดคำ	17
SegmentLogFull	Text	255	ของประวัติการตัดคำแบบเต็ม	เพชร มี สี อะไร บ้าง
SegmentLogShort	Text	255	ของประวัติการตัดคำแบบสั้น	เพชร สี อะไร บ้าง
SegDateTime	Text	50	วันที่และเวลาของการเก็บประวัติ	8/1/2009 10:55:44 PM

ตาราง 3.8 รายละเอียดเพิ่มข้อมูล vocab

ชื่อเพิ่มข้อมูล	vocab			
คำอธิบาย	ข้อมูลคำศัพท์			
คีย์หลัก	ID			
คีย์นอก	BotId			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
ID	AutoNumber	8	รหัส	14
vocabword	Text	255	คำศัพท์	อัญมณี
wordlong	Number	8	จำนวนตัวอักษร	6
wordType	Text	4	ประเภทของคำศัพท์	N
BotId	Text	20	รหัสของโรบอด	906132741

ตาราง 3.9 รายละเอียดเพิ่มข้อมูล vocabNotuse

ชื่อเพิ่มข้อมูล	vocabNotuse			
คำอธิบาย	ข้อมูลคำศัพท์ที่ไม่นำมาตัดคำ			
คีย์หลัก	VocabID			
คีย์นอก	-			
หมายเหตุ	-			
ชื่อฟิลด์	ชนิด	ขนาด (ไบต์)	ความหมาย	ตัวอย่างข้อมูล
VocabID	AutoNumber	8	รหัสคำที่ไม่นำมาตัด	1
VocabWord	Text	30	คำที่ไม่นำมาตัด	กรับ,กะ,จีะ