

**สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง**

**การศึกษาเปรียบเทียบวิธีการค้นคืนเอกสารการบรรยายยูสเคส**

**A COMPARISON STUDY ON USE CASE DESCRIPTION  
RETRIEVAL METHODS**



ร.พ.

ร.พ. 17 ก

2550

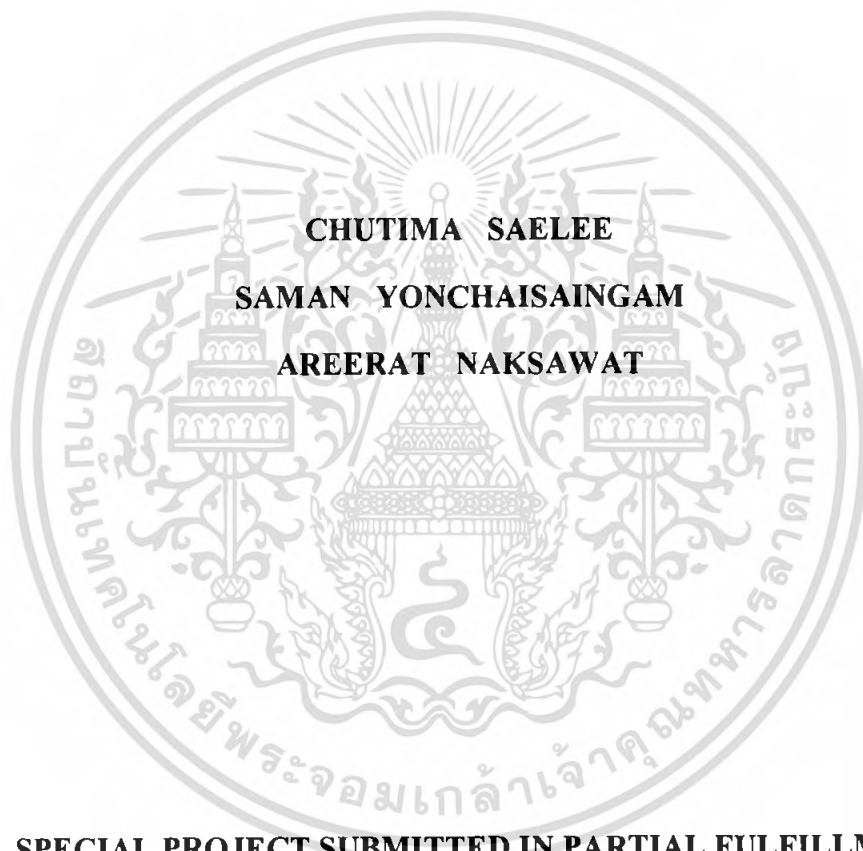
เลขหมู่.....  
เลขทะเบียน..... **82794**  
วัน,เดือน,ปี..... **23 ก.ค. 2551**

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต  
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ปีการศึกษา 2550

**119507166**  
b.....  
i.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำเผยแพร่

**A COMPARISON STUDY ON USE CASE DESCRIPTION  
RETRIEVAL METHODS**



**A SPECIAL PROJECT SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIRMENT FOR DEGREE OF BACHELOR OF SCIENCE  
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE  
FACULTY OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
ACADEMIC YEAR 2007**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**หัวข้อปัญหาพิเศษ**

การศึกษาเปรียบเทียบวิธีการค้นคืนเอกสารการบรรยายอุตสาหกรรม  
 A COMPARISON STUDY ON USE CASE DESCRIPTION  
 RETRIEVAL METHODS

**ชื่อนักศึกษา**

นางสาวชอุติมา แซ่หลี่ 47050323  
 นายสมาน ยนต์ชัยไทรงาม 47050351  
 นางสาวอารีรัตน์ นาคสวัสดิ์ 47050356

**ภาควิชา**

คณิตศาสตร์และวิทยาการคอมพิวเตอร์

**สาขาวิชา**

วิทยาการคอมพิวเตอร์

**อาจารย์ที่ปรึกษา**

อาจารย์อัคเดช อุดมชัยพร

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง อนุมัติให้รับปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ประจำปีการศึกษา 2550

คณะกรรมการสอบ	ลายมือชื่อ
รองศาสตราจารย์ ดร.วีระ บุญจริง ประธานกรรมการ	
อาจารย์สังกรศรีณีย์ ล่องชูผล กรรมการ	
อาจารย์อัคเดช อุดมชัยพร กรรมการและอาจารย์ที่ปรึกษา	



(รองศาสตราจารย์ไพโรจน์ พันธุ์พงษ์)

หัวหน้าภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

ลิขสิทธิ์ของภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์  
 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่ ทุกคนที่เป็นกำลังใจมาโดยตลอด  
ชุตินา



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การศึกษาเปรียบเทียบวิธีการค้นคืนเอกสารการบรรยายยูสเคส	
ชื่อนักศึกษา	นางสาวชุตติมา แซ่หลี่	47050323
	นายสมาน ынด้ชัยไทรงาม	47050351
	นางสาวอารีรัตน์ นาคสวัสดิ์	47050356
ปริญญา	วิทยาศาสตร์บัณฑิต	
ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์	
สาขาวิชา	วิทยาการคอมพิวเตอร์	
ปีการศึกษา	2550	
อาจารย์ที่ปรึกษา	อาจารย์อัคเดช อุดมชัยพร	

## บทคัดย่อ

ปัญหาพิเศษนี้ทำการศึกษาเปรียบเทียบวิธีการค้นคืนข้อกำหนดความต้องการซอฟต์แวร์ที่อยู่ในรูปแบบของเอกสารการบรรยายยูสเคส เนื่องจากการค้นคืนเอกสารการบรรยายยูสเคสนั้นสามารถทำได้หลายวิธี ปัญหาพิเศษนี้จึงได้ทำการทดลองเพื่อเปรียบเทียบวิธีการค้นคืนแบบต่างๆ ประกอบด้วยวิธีการถ่วงน้ำหนัก 3 วิธีคือ TF, IDF, TF-IDF และวิธีคำนวณค่าความคล้าย 5 วิธีคือ ไคซ์ แจ็คคาร์ด โคลไซน์ โอเวอร์แลป และอะซิมเมตริก ซึ่งจะได้การทดลองทั้งหมด 15 กรณี จากผลการทดลองสรุปได้ว่าวิธีการถ่วงน้ำหนักแบบ TF-IDF กับ วิธีคำนวณความคล้ายแบบโอเวอร์แลป ให้ผลลัพธ์ที่ดีที่สุด วิธีการดังกล่าวสามารถนำไปใช้ในการค้นคืนข้อกำหนดความต้องการซอฟต์แวร์เพื่อนำกลับมาใช้ใหม่ ซึ่งจะช่วยให้การพัฒนาซอฟต์แวร์ใช้ระยะเวลาและงบประมาณที่ลดลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<b>Title</b>	A COMPARISON STUDY ON USE CASE DESCRIPTION RETRIEVAL METHODS	
<b>Students</b>	Ms.Chutima Saelee	47050323
	Mr.Saman Yonchaisaingam	47050351
	Ms.Areerat Naksawat	47050356
<b>Degree</b>	Bachelor of Science	
<b>Department</b>	Mathematics and Computer Science, Faculty of Science	
<b>Programme</b>	Computer Science	
<b>Academic Year</b>	2007	
<b>Advisor</b>	Mr.Akadej Udomchaiporn	

## ABSTRACT

The purpose of this special problem is to study and compare methods of software requirements retrieval in a form of use case description. There are many retrieval methods which can be applied to software requirements, so the special problem sets experiments to compare 3 weighting methods which are TF, IDF, TF-IDF and 5 similarity computation methods which are Dice, Jaccard, Cosine, Overlap and Asymmetric.

Therefore, there are 15 cases for the experiments and the results indicate that TF-IDF and Overlap are the best methods for weighting and similarity computation respectively. The methods can be used to retrieve software requirements for reusing them, so time and cost of software development can be decreased.

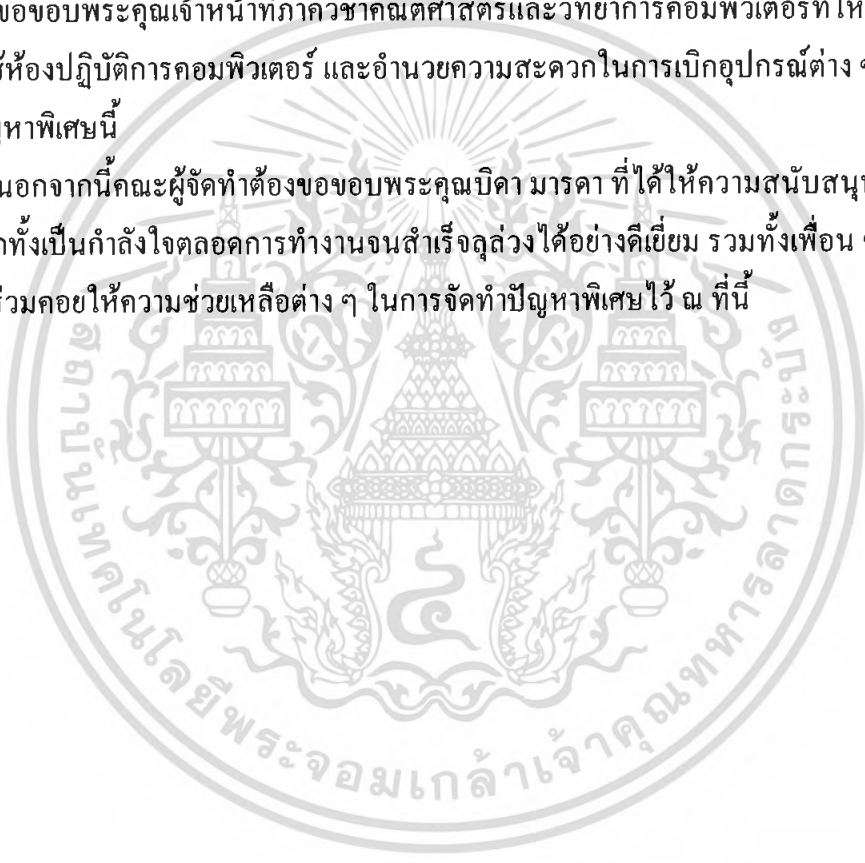
## กิตติกรรมประกาศ

ปัญหาพิเศษเรื่องการศึกษาเปรียบเทียบวิธีการค้นคืนเอกสารการบรรยายขุสเทศฉบับนี้ สามารถสำเร็จไปได้ด้วยดี คณะผู้จัดทำขอขอบพระคุณ อาจารย์อัคเดช อุดมชัยพร อาจารย์ที่ปรึกษา ปัญหาพิเศษฉบับนี้ที่กรุณาให้คำปรึกษาในการแก้ปัญหาด่าง ๆ รวมทั้งยังเป็นผู้ตรวจสอบความถูกต้องของปัญหาพิเศษฉบับนี้ได้เป็นอย่างดี

ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ทั้งทางด้านทฤษฎีและ ภาคปฏิบัติแก่คณะผู้จัดทำจนกระทั่งปัญหาพิเศษนี้สัมฤทธิ์ผลได้ด้วยดีทุกประการ

ขอขอบพระคุณเจ้าหน้าที่ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ที่ให้ความสะดวก ในการใช้ห้องปฏิบัติการคอมพิวเตอร์ และอำนวยความสะดวกในการเบิกอุปกรณ์ต่าง ๆ ที่ใช้ในการ จัดทำปัญหาพิเศษนี้

นอกจากนี้คณะผู้จัดทำต้องขอขอบพระคุณบิดา มารดา ที่ได้ให้ความสนับสนุนทางด้านทุน ทรัพย์ อีกทั้งเป็นกำลังใจตลอดการทำงานจนสำเร็จลุล่วงได้อย่างดีเยี่ยม รวมทั้งเพื่อน ๆ พี่ ๆ ทุกคน ที่มีส่วนร่วมคอยให้ความช่วยเหลือต่าง ๆ ในการจัดทำปัญหาพิเศษไว้ ณ ที่นี้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	i
บทคัดย่อภาษาอังกฤษ .....	ii
กิตติกรรมประกาศ .....	iii
สารบัญ.....	iv
สารบัญภาพ .....	vii
สารบัญตาราง.....	ix
<b>บทที่ 1 บทนำ</b> .....	<b>1</b>
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์.....	1
1.3 ขอบเขต .....	1
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b> .....	<b>3</b>
2.1 การจัดเก็บและการค้นคืนข้อมูล.....	3
2.1.1 การจัดเก็บข้อมูล.....	4
2.1.1.1 การสร้างดัชนี.....	4
2.1.1.2 การถ่วงน้ำหนัก.....	5
2.1.1.2.1 การถ่วงน้ำหนักตามความถี่ของคำ.....	6
2.1.1.2.2 การถ่วงน้ำหนักตามความถี่เอกสารแบบผกผัน.....	6
2.1.1.2.3 การถ่วงน้ำหนักตามความถี่ของคำและการถ่วงน้ำหนักตามความถี่เอกสารแบบผกผัน .....	6
2.1.2 การค้นคืนข้อมูล .....	7
2.1.2.1 การสร้างคิวรี.....	7
2.1.2.2 การคำนวณความคล้าย.....	7
2.1.2.3 การจัดลำดับผลลัพธ์จากการค้นคืน.....	9
2.1.3 การประเมินคุณภาพของวิธีการจัดเก็บและค้นคืนข้อมูล .....	9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

หน้า

2.2 เอกสารการบรรยายยูสเคส.....	11
2.2.1 ความหมายเบื้องต้นของเอกสารการบรรยายยูสเคส .....	11
<b>บทที่ 3 การดำเนินงานวิจัย.....</b>	<b>14</b>
3.1 ขั้นตอนการดำเนินงานวิจัย .....	14
3.1.1 การรวบรวมเอกสารการบรรยายยูสเคสและการสร้างดัชนี .....	17
3.1.2 การสร้างคิวรีและกระบวนการค้นคืน.....	18
3.1.3 การประเมินผลระบบ .....	20
3.2 การดำเนินการทดลอง.....	21
3.2.1 การทดลอง .....	21
3.3 การออกแบบฐานข้อมูล .....	23
3.3.1 ความสัมพันธ์ของแต่ละเอนทิตี .....	24
3.3.2 รายละเอียดของตารางต่างๆ.....	25
<b>บทที่ 4 ผลการทดลอง .....</b>	<b>30</b>
4.1 ค่าขีดแบ่งที่ได้จากการทดลอง.....	30
4.2 ค่าเฉลี่ยรีคอล ฟริชชัน และฮาร์โมนิกมีนของทุกคิวรีทั้ง 15 กรณี .....	34
4.3 ผลการวัดประสิทธิภาพของการค้นคืนของการค้นคืนโดยใช้ทฤษฎี การให้ความสำคัญสำหรับตัววัดประสิทธิภาพการค้นคืน .....	40
4.3.1 กรณีที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ายทั้งหมดคลบกับ ค่าส่วนเบี่ยงเบนมาตรฐาน(S.D) ของค่าเฉลี่ยของค่าความคล้ายทั้งหมด .....	40
4.3.1.1 กรณีให้ความสำคัญกับฟริชชันอย่างเดียว .....	40
4.3.1.2 กรณีให้ความสำคัญกับฟริชชันมากกว่ารีคอล.....	42
4.3.1.3 กรณีให้ความสำคัญกับฟริชชันเท่ากับรีคอล.....	44
4.3.2 กรณีที่ ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ายทั้งหมด.....	46
4.3.2.1 กรณีให้ความสำคัญกับฟริชชันอย่างเดียว .....	46
4.3.2.2 กรณีให้ความสำคัญกับฟริชชันมากกว่ารีคอล.....	48
4.3.2.3 กรณีให้ความสำคัญกับฟริชชันเท่ากับรีคอล.....	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

หน้า

4.3.3	กรณีที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ยทั้งหมดบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน(S.D.)ของค่าเฉลี่ยของค่าความคล้ยทั้งหมด .....	51
4.3.3.1	กรณีให้ความสำคัญกับปริชีข้นอย่างเดีว.....	51
4.3.3.2	กรณีให้ความสำคัญกับปริชีข้นมากกว่ารีคอด.....	54
4.3.3.3	กรณีให้ความสำคัญกับปริชีข้นเท่ากับรีคอด.....	56
<b>บทที่ 5</b>	<b>สรุปผลการวิจัย อภิปรายและข้อเสนอแนะ .....</b>	<b>58</b>
5.1	สรุปผลการวิจัย.....	58
5.2	การอภิปรายผลการวิจัย.....	60
5.3	ข้อเสนอแนะ.....	60
	<b>รายการอ้างอิง.....</b>	<b>61</b>
	<b>ภาคผนวก ก. รายละเอียดระบบของเอกสารการบรรยายยูสเคสที่เลือกใช้ในการวิจัย .....</b>	<b>62</b>
	<b>ภาคผนวก ข. การใช้งานระบบ.....</b>	<b>80</b>
ข.1	คู่มือการใช้งานระบบ .....	81
ข.1.1	ขั้นตอนการค้นคืน.....	81
ข.1.2	การเพิ่มเอกสารการบรรยายยูสเคสเข้ามาในระบบ.....	85
	<b>ภาคผนวก ค. หมายเลขเอกสารการบรรยายยูสเคสที่เกี่ยวข้องกับแต่ละหัวข้อของการทดลอง.....</b>	<b>104</b>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญภาพ

ภาพที่	หน้า
2.1 โครงสร้างพื้นฐานทางสถาปัตยกรรมของระบบคั่นคินข้อมูล.....	3
2.2 แสดงกระบวนการทำงานก่อนการสร้างดัชนี.....	5
2.3 การสร้างดัชนีผกผัน (Inverted Index) .....	5
2.4 ตัวอย่างของเอกสารการบรรยายยูสเคส.....	13
3.1 รูปแบบสำหรับการสร้างคิวรี .....	15
3.2 แผนภาพกิจกรรมของระบบทั้งหมด โดยสังเขป.....	16
3.3 แผนภาพกิจกรรมการสร้างดัชนี.....	17
3.4 แผนภาพกิจกรรมกระบวนการคั่นคิน.....	18
3.5 แผนภาพกิจกรรมการประเมินผล.....	21
3.6 แผนภาพความสัมพันธ์ของเอนติตีของระบบคั่นคินเอกสารการบรรยายยูสเคส .....	24
4.1 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยที่ค่าขีดแบ่ง เท่ากับค่าเฉลี่ยของทุกคิวรีลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับ พรีซิชันอย่างเดียวน เมื่อ $\alpha$ เท่ากับ 0.....	41
4.2 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยที่ค่าขีดแบ่ง เท่ากับค่าเฉลี่ยของทุกคิวรีลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับ พรีซิชันมากกว่ารีคอล เมื่อ $\alpha$ เท่ากับ 0.5 .....	43
4.3 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยที่ค่าขีดแบ่ง เท่ากับค่าเฉลี่ยของทุกคิวรีลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับ พรีซิชันเท่ากับรีคอล เมื่อ $\alpha$ เท่ากับ 1.....	45
4.4 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชันอย่างเดียวน เมื่อ $\alpha$ เท่ากับ 0.....	47
4.5 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชันมากกว่ารีคอล เมื่อ $\alpha$ เท่ากับ 0.5 .....	49
4.6 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชันเท่ากับรีคอล เมื่อ $\alpha$ เท่ากับ 1 .....	51
4.7 การเรียงลำดับของประสิทธิภาพการคั่นคินจากมากไปน้อย โดยที่ค่าขีดแบ่ง เท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับ พรีซิชันอย่างเดียวน $\alpha$ เท่ากับ 0.....	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.8 การเรียงลำดับของประสิทธิภาพการคั่นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอล $\alpha$ เท่ากับ 0.5 .....	55
4.9 การเรียงลำดับของประสิทธิภาพการคั่นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอล $\alpha$ เท่ากับ 1 .....	57
ข.1 หน้าจอการระบบคั่นคืน .....	81
ข.2 หน้าจอระบบคั่นคืนที่ผู้ใช้กรอกรายละเอียด .....	82
ข.3 หน้าจอผลลัพธ์จากการคั่นคืน .....	83
ข.4 หน้าจอตัวอย่างเอกสารการบรรยายยูสเคสที่ได้จากการคั่นคืน .....	84
ข.5 หน้าจอการเพิ่มเอกสารการบรรยายยูสเคสเข้ามาในระบบ .....	85
ข.6 แสดงการใส่ข้อมูลเอกสารการบรรยายยูสเคส .....	86
ข.7 แสดงสถานะการบันทึกข้อมูลเอกสารการบรรยายยูสเคส .....	87
ข.8 แสดงการใส่ข้อมูลเอกสารการบรรยายยูสเคสซ้ำกับที่มีอยู่แล้ว .....	87

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
2.1 ความสัมพันธ์ของเอกสารที่ถูกค้นคืนกับความเกี่ยวข้อง .....	9
3.1 กรณีที่เกิดขึ้นในการค้นคืนเอกสารการบรรยายยูสเคส .....	19
3.2 ระบบของเอกสารการบรรยายยูสเคสที่นำมาให้ทดลอง.....	22
3.3 รายละเอียดของเอกสารการบรรยายยูสเคส .....	25
3.4 ระบบของเอกสารการบรรยายยูสเคส.....	26
3.5 ส่วนประกอบย่อยของเอกสารการบรรยายยูสเคส.....	26
3.6 คำทั้งหมดในฐานข้อมูล .....	26
3.7 คำทั้งหมดในแต่ละเอกสารการบรรยายยูสเคส.....	27
3.8 รายการคำที่ไม่มีประโยชน์ในการค้นคืน .....	27
3.9 ผู้ทดสอบ.....	27
3.10 ข้อมูลคิวรีของผู้ทดสอบที่บันทึกไว้โดยใช้วิธีค้นคืนด้วยโครงสร้าง ของเอกสารการบรรยายยูสเคส.....	28
3.11 หัวข้อ .....	28
3.12 ผลลัพธ์ของการค้นคืน.....	29
3.13 จำนวนคำที่ปรากฏในแต่ละเอกสาร .....	29
4.1 ค่าขีดแบ่งที่ได้จากการทดลองเมื่อค่าขีดแบ่งเท่ากับMean - S.D.....	31
4.2 ค่าขีดแบ่งที่ได้จากการทดลองเมื่อค่าขีดแบ่งเท่ากับMean. ....	32
4.3 ค่าขีดแบ่งที่ได้จากการทดลองเมื่อค่าขีดแบ่งเท่ากับMean + S.D.....	33
4.4 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่าน้ำหนักแบบต่างๆ เมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน .....	34
4.5 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย.....	35
4.6 ผลการทดลองของวิธีวัดความคล้ายแบบแจ็คคาร์ด (Jaccard) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย.....	35
4.7 ผลการทดลองของวิธีวัดความคล้ายแบบ โคไซน์ (Cosine) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย .....	36
4.8 ผลการทดลองของวิธีวัดความคล้ายแบบ โอเวอร์แลป (Overlap) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.9 ผลการทดลองของวิธีวัดความคล้ายแบบอะซิมเมตริก (Asymmetric) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ย.....	37
4.10 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน.....	37
4.11 ผลการทดลองของวิธีวัดความคล้ายแบบแจ็คคาร์ด (Jaccard) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน .....	38
4.12 ผลการทดลองของวิธีวัดความคล้ายแบบโคไซน์ (Cosine)ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน .....	38
4.13 ผลการทดลองของวิธีวัดความคล้ายแบบ โอเวอร์แลป (Overlap) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน .....	39
4.14 ผลการทดลองของวิธีวัดความคล้ายแบบอะซิมเมตริก (Asymmetric) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน .....	39
4.15 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีลกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ $\alpha$ เท่ากับ 0 .....	40
4.16 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีลกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอล เมื่อ $\alpha$ เท่ากับ 0.5 .....	42
4.17 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีลกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอล เมื่อ $\alpha$ เท่ากับ 1 .....	44
4.18 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ $\alpha$ เท่ากับ 0 .....	46
4.19 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอล เมื่อ $\alpha$ เท่ากับ 0.5 .....	48
4.20 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อย โดยให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอล เมื่อ $\alpha$ เท่ากับ 1 .....	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.21 การเรียงลำดับของประสิทธิภาพการคั่นคืนจากมากไปน้อย โดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นอย่างเคียว $\alpha$ เท่ากับ 0.....	52
4.22 การเรียงลำดับของประสิทธิภาพการคั่นคืนจากมากไปน้อย โดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอลล $\alpha$ เท่ากับ 0.5 .....	54
4.23 การเรียงลำดับของประสิทธิภาพการคั่นคืนจากมากไปน้อย โดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกคิวรีบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอลล $\alpha$ เท่ากับ 1.....	56
5.1 การสรุปผลการทดลอง.....	59
ก.1 ระบบของเอกสารการบรรยายยูสเคสที่ใช้ในการทดลอง.....	63
ก.2 ICT blocking and filtering system.....	64
ก.3 Departmental information system for curriculum and course offerings management.....	65
ก.4 Resource management system.....	66
ก.5 Automatic restaurant lookup system.....	67
ก.6 Library management system.....	68
ก.7 Mobile phone selling on the internet system.....	69
ก.8 Payroll system.....	69
ก.9 Personal stock manager on PDA system.....	71
ก.10 Basic product trading system.....	72
ก.11 Mobile phone service paying system.....	72
ก.12 Basic MLM computerize system.....	73
ก.13 Content management and platform upgrade for the official website of Nation Channel.....	74
ก.14 Purchase, producing, and export system.....	75
ก.15 Management information system for insurance company.....	76
ก.16 E-learning system.....	77
ก.17 Management information system for an academic.....	78
ค.1 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการเรียนการสอน.....	89

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง (ต่อ)

ตารางที่	หน้า
ค.2 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการค้าขายสินค้า.....	91
ค.3 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการจัดการข้อมูลลูกค้า.....	92
ค.4 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการคำนวณการเงิน.....	93
ค.5 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการสร้างรายงาน .....	94



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญ

ปัจจุบันปัญหาเรื่องงบประมาณของการผลิตซอฟต์แวร์ที่อยู่ในระดับที่สูงนั้นเนื่องจากที่ขั้นตอนในการผลิตซอฟต์แวร์ในแต่ละขั้นนั้นขาดการวางแผนและการนำเอาส่วนประกอบของการออกแบบกลับมาใช้ใหม่อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งในส่วนของข้อกำหนดความต้องการซอฟต์แวร์ (Software Requirements Specification) ซึ่งข้อกำหนดความต้องการซอฟต์แวร์นี้มักจะเขียนอยู่ในรูปแบบของเอกสารการบรรยายยูสเคส (Use Case Description) ดังนั้นในการนำเอาเอกสารดังกล่าวกลับมาใช้ใหม่จะช่วยให้สามารถลดระยะเวลาและงบประมาณของการพัฒนาซอฟต์แวร์ลงได้ ซึ่งก็คือเราต้องสามารถค้นหาเอกสารการบรรยายยูสเคสได้ตรงกับความต้องการมากที่สุด ในปัจจุบันมีวิธีการค้นหาข้อมูลหลากหลายวิธี ซึ่งวิธีการต่าง ๆ นั้นได้ถูกพัฒนาให้ดีขึ้นมาเป็นลำดับแต่ยังไม่สามารถสรุปได้ว่าวิธีการให้น้ำหนักและการคำนวณความคล้ายแบบใดเป็นวิธีที่มีประสิทธิภาพสำหรับการค้นหาเอกสารการบรรยายยูสเคส ซึ่งเป็นที่มาที่ทำให้เกิดการศึกษาและการทดลองในปัญหาพิเศษนี้

### 1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษาวิจัยว่าวิธีการให้น้ำหนักและวิธีการคำนวณหาความคล้ายแบบใดที่ทำให้การค้นหาเอกสารการบรรยายยูสเคสมีประสิทธิภาพที่ดีที่สุด โดยใช้ตัววัดประสิทธิภาพการค้นหาคือเอฟเมเชอร์ (F-measure)

1.2.2 เพื่อพัฒนาซอฟต์แวร์ประยุกต์ซึ่งใช้วิธีการให้น้ำหนักและวิธีการคำนวณหาความคล้ายตามการทดลองเป็นตัวอย่างที่แสดงให้เห็นว่าวิธีที่ใช้ในการค้นหาดังกล่าวสามารถนำมาใช้งานได้จริง

### 1.3 ขอบเขต

1.3.1 วิธีการให้น้ำหนักและการคำนวณความคล้ายที่นำมาใช้ในปัญหาพิเศษนี้ จะเป็นวิธีการถ่วงน้ำหนักดัชนีและวิธีคำนวณค่าความคล้าย โดย วิธีการถ่วงน้ำหนักดัชนีที่นำมาใช้มีดังนี้ TF, IDF, TF-IDF และ วิธีการคำนวณค่าความคล้ายที่นำมาใช้ในการทดลองมีดังนี้ ไคซ์ แจ็คคาร์ด โทไชน์ โอเวอร์แลปและอะซิมเมตริก

1.3.2 วิธีการและข้อมูลที่นำมาทดลองในปัญหาพิเศษนี้จะต้องเป็นไปตามการทดลองของเอกสาร [3]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 ทำให้รู้่ววิธีในการให้น้ำหนักและการคำนวณความคล้ายกันวิธีใดทำให้ประสิทธิภาพในการค้นคืนที่เหมาะสมกับการใช้ในการค้นคืนเอกสารการบรรยายยุคศต
- 1.4.2 ช่วยลดระยะเวลาและงบประมาณในการพัฒนาซอฟต์แวร์ลงได้
- 1.4.3 ทำให้ได้ข้อมูลที่แม่นยำถูกต้องและตรงกับความต้องการของผู้ใช้งานมากขึ้น
- 1.4.4 นำแนวคิดไปพัฒนาหรือแทรกอยู่ในโปรแกรมประยุกต์อื่นได้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

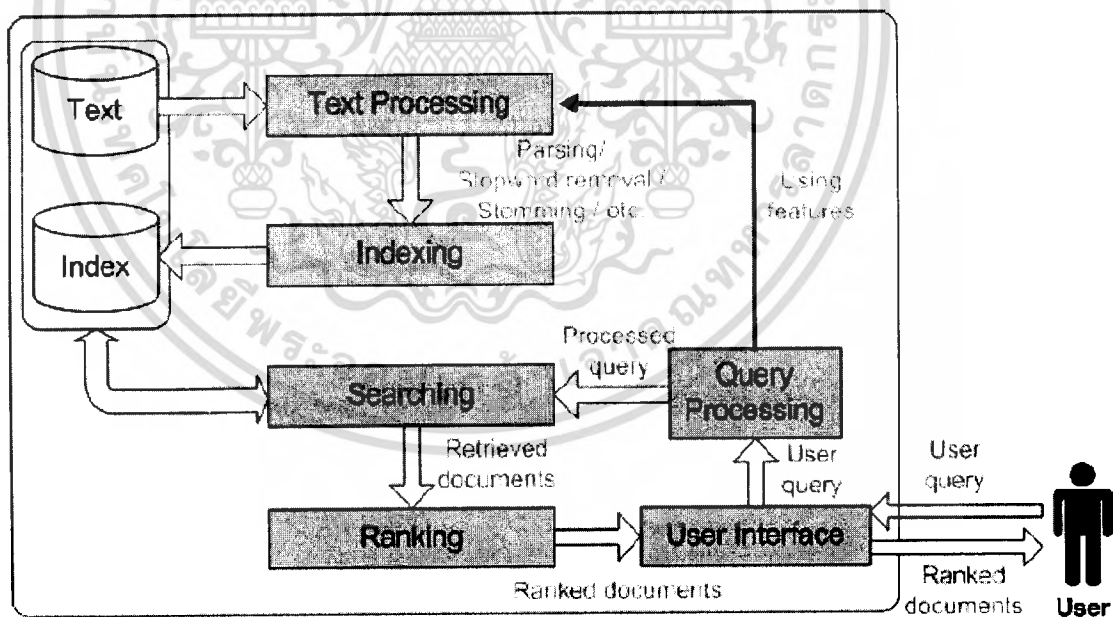
### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีของการจัดเก็บและค้นคืนข้อมูลแบบทั่วไปและจะกล่าวถึงส่วนประกอบต่างๆของเอกสารบรรณานุกรมซึ่งเป็นที่เกี่ยวข้องกัปัญหาพิเศษนี้

เนื่องจากข้อมูลต่างๆก่อนที่จะนำมาประมวลต้องมีการจัดเก็บที่ดีและมีประสิทธิภาพต่อการค้นคืนดังนั้นเราต้องรู้ขั้นตอนการเก็บและการนำข้อมูลเหล่านั้นกลับมาใช้ประโยชน์ได้อีกในภายหลัง

#### 2.1 การจัดเก็บและการค้นคืนข้อมูล (Information Storage and Retrieval)

การจัดเก็บข้อมูลเป็นการนำข้อมูลที่อยู่ในรูปแบบดิจิทัลเก็บลงในหน่วยเก็บข้อมูลเช่นฮาร์ดดิสก์ ซึ่งการจัดเก็บข้อมูลต่าง ๆ นั้นจำเป็นต้องมีการออกแบบการเก็บเพื่อให้สามารถค้นคืนได้อย่างสะดวก การค้นคืนข้อมูลจะมีการใช้สมการทางคณิตศาสตร์เพื่อนำมาใช้คำนวณค่าถ่วงน้ำหนักของคำต่างๆในแต่ละเอกสารและการใช้ในการคำนวณและวิเคราะห์ความคล้ายของเอกสารกับคิวรีเพื่อที่จะสามารถค้นคืนเอกสารได้ตรงกับความต้องการของผู้ใช้



ภาพที่ 2.1 โครงสร้างพื้นฐานทางสถาปัตยกรรมของระบบค้นคืนข้อมูล [6]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปเป็นกระบวนการค้นคืนเอกสารโดยผู้ใช้งานจะต้องใส่คิวรีเข้าไปในระบบ จากนั้นระบบจะทำการตัดคำและกระบวนการอื่นๆ เพื่อนำข้อมูลที่ได้ไปคำนวณทางทฤษฎีการค้นคืนข้อมูลและจะได้รายการของเอกสารตามที่เราต้องการ

การทำงานต่างๆในปัจจุบันต้องมีข้อมูลต่างๆเข้ามาเกี่ยวข้องกับเสมอดังนั้นเราต้องรู้จักวิธีการจัดเก็บข้อมูลต่างๆนั้นให้เหมาะสมกับประเภทข้อมูลแต่ละชนิดและมีขั้นตอนในการจัดเก็บที่ง่ายต่อการนำข้อมูลนั้นๆกลับมาใช้ประโยชน์อีกครั้ง

### 2.1.1 การจัดเก็บข้อมูล (Information Storage)

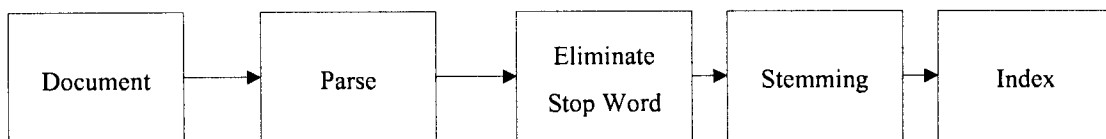
การจัดเก็บข้อมูลเป็นกระบวนการที่นำเอาไฟล์เอกสารต่างๆบันทึกในฐานข้อมูล โดยมีวิธีการในการนำเอาคำต่างๆที่อยู่ในแต่ละเอกสารมาผ่านกระบวนการดังนี้

#### 2.1.1.1 การสร้างดัชนี (Indexing)

ในกระบวนการสร้างดัชนีจะมีกระบวนการย่อยๆเพื่อที่จะใช้ในการสร้างดัชนีที่สามารถใช้ค้นหาคำที่ต้องการได้อย่างสะดวกรวดเร็ว โดยการวิเคราะห์และประมวลผลข้อความสำหรับการสร้างดัชนี การทำงานในส่วนนี้มีกระบวนการดังนี้คือ

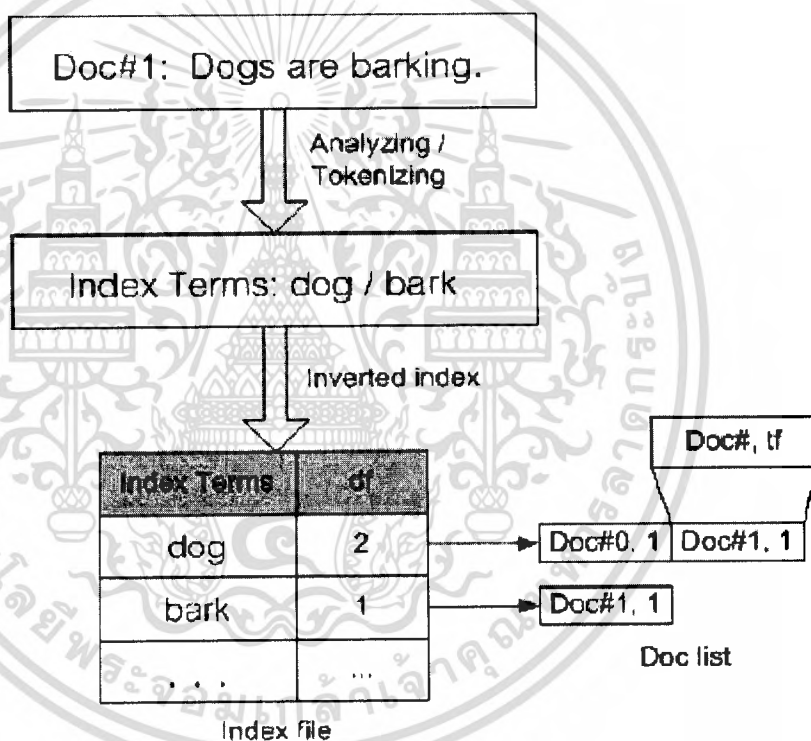
1. การตรวจการเรียงตัวของอักขระ(Parsing) และการตัดคำ(Tokenizing)
2. การแปลงคำให้อยู่ในรูปของรากศัพท์(Stemming) ส่วนนี้เป็นการลดทอนคำให้อยู่ในรูปแบบที่เป็นรากของคำศัพท์ เช่น คำว่า “computed”, “computer”, “computing” จะถูกแปลงเป็น “compute” ทั้งหมด เป็นต้น ซึ่งในการทดลองนี้ จะใช้ขั้นตอนวิธีของพอร์เตอร์ (Porter’s Algorithm)[3]
3. สตอปเวิร์ด (คำที่ไม่มีมีความหมายสำคัญและเกิดขึ้นบ่อยๆ) เช่น คำว่า “the” “on” “is” คำเหล่านี้ไม่สามารถระบุส่งต่างๆได้ชัดเจนจึงต้องกำจัดออกไป และการตัดคำเหล่านี้ออกไปก็ยังสามารถช่วยเพิ่มความเร็วในการสร้างดัชนีได้อีก ในการทดลองนี้มีการเก็บข้อมูลของสตอปเวิร์ดซึ่งมีทั้งหมด 425 คำ [4]

ในขั้นตอนนี้ระบบจะทำการสร้างดัชนีจากคลังเอกสาร โดยที่เอกสารที่จะนำมาสร้างดัชนีต้องนำมาผ่านในส่วนสำหรับประมวลผลข้อความก่อนเพื่อสกัดคำที่สำคัญไปสร้างดัชนีคำ เมื่อมีการเปลี่ยนแปลงและแก้ไขเอกสารหรือมีการเพิ่มเอกสาร ระบบจะต้องทำการสร้างดัชนีใหม่เพื่อให้สอดคล้องกับเอกสารชุดใหม่ ทั้งนี้เทคนิคการสร้างดัชนีที่มีประสิทธิภาพจะต้องมีความสามารถในการเพิ่มดัชนีในส่วนที่เปลี่ยนไปโดยไม่กระทบกับส่วนที่มีอยู่แล้ว เทคนิคนี้มีชื่อว่าการเพิ่มดัชนีซึ่งทำให้ไม่ต้องเสียเวลาในการแก้ไขข้อมูลดัชนีใหม่ทั้งหมดในกรณีที่มีการเปลี่ยนแปลงของเอกสารเพียงเล็กน้อย



ภาพที่ 2.2 แสดงกระบวนการทำงานก่อนการสร้างดัชนี

เทคนิคการสร้างดัชนีที่จะนำมาใช้นี้จะเป็นแบบดัชนีผกผัน(Inverted Index) วิธีการสร้างดัชนีแบบนี้เริ่มจากการนำเอกสารมาวิเคราะห์และแบ่งเป็นรายการของคำซึ่งอาจจะมีการกำจัดสตอปเวิร์ดและแปลงคำให้เป็นรากศัพท์ (Stemming) จากนั้นจึงนำมาเก็บเป็นไฟล์ดัชนีซึ่งมีค่าต่างๆ พร้อมทั้งจำนวนเอกสารที่คำนั้นๆ ปรากฏอยู่ และแต่ละคำก็จะมีข้อมูลซึ่งระบบรายการหมายเลขของเอกสารพร้อมทั้งจำนวนคำที่ปรากฏอยู่ในเอกสารนั้น ดังแสดงในภาพที่ 2.3



Note: df = document frequency , tf = term frequency

ภาพที่ 2.3 การสร้างดัชนีผกผัน (Inverted Index) [6]

#### 2.1.1.2 การถ่วงน้ำหนัก (Weighting)

การถ่วงน้ำหนักเป็นการจัดลำดับความสำคัญของคำศัพท์แต่ละคำว่ามีค่าความถี่มากน้อยแค่ไหนในเอกสารหนึ่งๆ ยิ่งมีความถี่สูงแสดงว่าเอกสารนั้นๆ มีโอกาสถูกค้นคืนได้มาก ซึ่งค่าที่ได้จากการถ่วงน้ำหนักของแต่ละคำศัพท์นั้นจะถูกเก็บไว้เพื่อนำไปคำนวณความคล้าย (Similarity) กับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คิวรี โดยหลังจากได้สร้างเซตของดัชนีจากกระบวนการสร้างดัชนี แต่ละดัชนีถูกถ่วงน้ำหนักด้วยวิธีการต่างๆดังนี้

#### 2.1.1.2.1 การถ่วงน้ำหนักตามความถี่ของคำ (Term Frequency Weighting)

เป็นพื้นฐานการถ่วงน้ำหนักบนความคิดที่สร้าง (คำ, วลี) ที่เกิดขึ้นบ่อยในตัวอักษรของเอกสาร ดังนั้นจึงเป็น การถ่วงน้ำหนักของเทอม k ในเอกสาร i จึงเท่ากับความถี่ของเทอม k ในเอกสาร i การถ่วงน้ำหนักแบบ TF (Term Frequency) [2] มีสูตรดังนี้

กำหนดให้

$WEIGHT_{ik}$  คือ การถ่วงน้ำหนักของเทอม k ในเอกสาร i

$FREQ_{ik}$  คือ ความถี่ของเทอม k ในเอกสาร i

$$WEIGHT_{ik} = FREQ_{ik} \quad (\text{สมการที่ 2.1})$$

#### 2.1.1.2.2 การถ่วงน้ำหนักตามความถี่เอกสารแบบผกผัน (Inverse Document Frequency Weighting)

การวัดความถี่ของคำที่เกิดขึ้นในเอกสารทำให้ไม่มีความแตกต่างระหว่างเทอมที่เกิดขึ้นในทุกๆเอกสารที่รวบรวมได้ จากการสังเกตระบุได้ว่าคำที่มีความถี่สูงๆจะทำให้ได้เอกสารจำนวนมากที่ไม่มีประโยชน์ในการค้นคืนแต่ได้เอกสารที่เกี่ยวข้องน้อยจึงต้องมีการใช้วิธีถ่วงน้ำหนักเอกสารตามความถี่แบบผกผัน (Inverse Document Frequency (IDF)) [2] ซึ่งมีวิธีการถ่วงน้ำหนัก ดังนี้

กำหนดให้

$WEIGHT_{ik}$  คือ การถ่วงน้ำหนักของเทอม k ในเอกสาร i

$FREQ_{ik}$  คือความถี่ของเทอม k ในเอกสาร i

$DOCFREQ_k$  คือความถี่หรือจำนวนคำหรือเทอมที่ k ในเอกสารทั้งหมด

$$WEIGHT_{ik} = \frac{FREQ_{ik}}{DOCFREQ_k} \quad (\text{สมการที่ 2.2})$$

#### 2.1.1.2.3 การถ่วงน้ำหนักตามความถี่ของคำและการใช้การถ่วงน้ำหนักตามความถี่เอกสารแบบผกผัน (Term Frequency – Inverse Document Frequency Weighting)

เป็นการประยุกต์วิธีทั้งสองวิธีของการถ่วงน้ำหนัก (TF • IDF) [2] ดังที่กล่าวข้างต้นมาเข้าไว้ด้วยกัน โดยมีสูตรการคำนวณดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้

$WEIGHT_{ik}$  คือ การถ่วงน้ำหนักของเทอม  $k$  ในเอกสาร  $i$

$FREQ_{ik}$  คือ ความถี่ของเทอม  $k$  ในเอกสาร  $i$

$DOCFREQ_k$  คือ ความถี่หรือจำนวนคำหรือเทอมที่  $k$  ในเอกสารทั้งหมด

$$WEIGHT_{ik} = \frac{FREQ_{ik}}{DOCFREQ_k} \cdot FREQ_{ik} = \frac{(FREQ_{ik})^2}{DOCFREQ_k} \quad (\text{สมการที่ 2.3})$$

### 2.1.2 การค้นคืนข้อมูล (Information Retrieval)

เป็นการนำข้อมูลที่จัดเก็บไว้โดยมีการออกแบบการจัดเก็บเพื่อให้ได้ข้อมูลที่ไม่ซ้ำซ้อนและมีการสร้างดัชนีแล้วมาผ่านกระบวนการคำนวณความคล้าย (Similarity Computation) เพื่อให้ได้เอกสารที่ตรงกับความต้องการของผู้ใช้งาน โดยผู้ใช้งานที่ต้องการค้นคืนข้อมูลต้องกรอกข้อมูลคิควิธีเพื่อใช้เป็นข้อมูลเพื่อจะนำไปใช้ในการวิเคราะห์และคำนวณความคล้าย

#### 2.1.2.1 การสร้างคิควิธี (Query Generation)

เป็นกระบวนการที่ผู้ใช้งานกรอกข้อมูลคิควิธีที่ต้องการลงในหน่วยรับอินพุตจากนั้นข้อมูลจะถูกนำไปคำนวณเพื่อถ่วงน้ำหนักให้กับคิควิธี โดยคิควิธี 1 ชุดก็มองเป็น 1 เอกสารจากนั้นก็จะนำค่าถ่วงน้ำหนักของคำอื่นๆซึ่งถูกสร้างไว้ในฐานข้อมูลและถูกคำนวณไว้ก่อนหน้านี้แล้วมาใช้ในการคำนวณความคล้ายกันของเอกสารกับคิควิธีดังกล่าว

#### 2.1.2.2 การคำนวณความคล้าย (Similarity Computation)

เป็นวิธีที่ใช้ในการวัดว่าคิควิธีที่เราใช้นั้นคล้ายกับเอกสารใดบ้างในฐานข้อมูลเพื่อนำค่าดังกล่าวไปใช้ในการจัดเรียงข้อมูลเพื่อแสดงตามลำดับความสำคัญของเอกสาร (Ranking) ที่คล้ายกับคิควิธีดังกล่าว

กำหนดให้

$E_{mi}$  คือ สมาชิก  $m$  ของ เอกสาร  $i$ .

$E_{mj}$  คือ สมาชิก  $m$  ของ เอกสาร  $j$ .

$Term_{mik}$  เป็นค่าที่ถูกถ่วงน้ำหนัก เมื่อเทอม  $k$  ปรากฏในสมาชิก  $m$  ของเอกสาร  $i$  และเป็น 0 เมื่อเทอม  $k$  ไม่ปรากฏในสมาชิก  $m$  ของเอกสาร  $i$ .

$Term_{mjk}$  เป็น 1 เมื่อเทอม  $k$  ปรากฏในสมาชิก  $m$  ของเอกสาร  $j$  และเป็น 0 เมื่อเทอม  $k$  ไม่ปรากฏในสมาชิก  $m$  ของเอกสาร  $i$ .

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- วิธีการแบบไดซ์ (Dice) [2] (สมการที่ 2.4)

$$Similarity(E_{mi}, E_{mj}) = \frac{2 \left[ \sum_{k=1}^l Term_{mik} \cdot Term_{mjk} \right]}{\sum_{k=1}^l Term_{mik} + \sum_{k=1}^l Term_{mjk}}$$

- วิธีการแบบแจ็กการ์ด (Jaccard) [2] (สมการที่ 2.5)

$$Similarity(E_{mi}, E_{mj}) = \frac{\sum_{k=1}^l (Term_{mik} \cdot Term_{mjk})}{\sum_{k=1}^l Term_{mik} + \sum_{k=1}^l Term_{mjk} - \sum_{k=1}^l (Term_{mik} \cdot Term_{mjk})}$$

- วิธีการแบบโคไซน์ (Cosine) [2] (สมการที่ 2.6)

$$Similarity(E_{mi}, E_{mj}) = \frac{\sum_{k=1}^l (Term_{mik} \cdot Term_{mjk})}{\sqrt{\sum_{k=1}^l (Term_{mik})^2 \cdot \sum_{k=1}^l (Term_{mjk})^2}}$$

- วิธีการแบบโอเวอร์แลป (Overlap) [2] (สมการที่ 2.7)

$$Similarity(E_{mi}, E_{mj}) = \frac{\sum_{k=1}^l (Term_{mik} \cdot Term_{mjk})}{\min\left(\sum_{k=1}^l Term_{mik}, \sum_{k=1}^l Term_{mjk}\right)}$$

- วิธีการแบบอะซิมเมตริกซ์ (Asymmetric) [2] (สมการที่ 2.8)

$$Similarity(E_{mi}, E_{mj}) = \frac{\sum_{k=1}^l \min(Term_{mik}, Term_{mjk})}{\sum_{k=1}^l Term_{mik}}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากได้คำนวณความคล้ายของเอกสารในฐานข้อมูลกับคิวรีแล้วแล้วต่อไปเราจะนำค่าความคล้ายที่คำนวณได้นั้น ไปจัดเรียงลำดับความสำคัญของเอกสาร

### 2.1.2.3 การจัดลำดับผลลัพธ์จากการค้นคืน (Ranking)

ในทางทฤษฎีเอกสารที่เกี่ยวข้องกับที่ผู้ใช้ต้องการมากที่สุด (High Relevance) ในขณะนั้นควรจะถูกเสนอให้ผู้ใช้เป็นอันดับแรกๆ แต่ในทางปฏิบัติเป็นการยากที่จะวัดความต้องการของผู้ใช้ที่แท้จริง เทคนิคการจัดเรียงลำดับของผลลัพธ์เอกสาร โดยทั่วไปส่วนใหญ่จะดูจากความถี่ของคำในเอกสาร เวลาที่แก้ไขเอกสารครั้งล่าสุด หรืออาจจะมีการวิเคราะห์ความนิยมในการถูกอ้างอิงจากเอกสารอื่นๆ เช่น ในกรณีของเอกสารบนเว็บ เป็นต้น

การจัดลำดับผลลัพธ์จากการค้นคืนก็เป็นการเรียงลำดับของเอกสารที่จะแสดงแก่ผู้ใช้งานตามค่าความคล้ายที่ได้คำนวณไว้ก่อนหน้านี้เพื่อให้ผู้ใช้ได้ข้อมูลที่ต้องการ โดยอยู่ในสมมติฐานที่ว่าเอกสารที่มีคำที่ตรงกับคิวรีมากจะเป็นเอกสารที่ผู้ใช้ต้องการที่สุด แต่วิธีการวิเคราะห์ความคล้ายและการถ่วงน้ำหนักในการค้นคืนปัจจุบันมีมากมายหลายวิธีเราจึงต้องมีเกณฑ์ในการประเมินเพื่อที่จะสามารถเปรียบเทียบได้ว่าวิธีการใดเหมาะสมที่สุด

### 2.1.3 การประเมินคุณภาพของวิธีการจัดเก็บและค้นคืนข้อมูล

การประเมินคุณภาพของของวิธีการจัดเก็บและค้นคืนข้อมูลก็เป็นอีกหัวข้อหนึ่งที่มีการศึกษากันอย่างลึกซึ้งและยังเป็นปัญหาวิจัยในการหาวิธีการประเมินผลที่เหมาะสม มีหลักการหลายอย่างที่นำไปได้สำหรับการเปรียบเทียบประสิทธิภาพของวิธีการจัดเก็บและการค้นคืนข้อมูล โดยทั่วไปวิธีสำหรับการประเมินการค้นคืนข้อมูล เป็นการเปรียบเทียบระหว่างผลการค้นคืนที่ได้จากระบบ กับผลการค้นคืนที่ได้จากผู้เชี่ยวชาญโดยใช้มนุษย์ค้นคืนเองโดยมีโคเมนของคิวรีเดียวกัน แต่ก็เป็นที่ยากที่จะสรุปได้ว่าเอกสารใดตรงกับความต้องการของผู้ใช้มากที่สุดเพราะเนื่องมาจากความหมายทางภาษาของคำแต่ละคำมีมากมายเช่นคำว่า Apple อาจหมายถึงผลไม้ หรืออาจหมายถึงบริษัทคอมพิวเตอร์ก็ได้ ในการเปรียบเทียบจะดูถึงค่าพรีซิชั่น (Precision) รีคอล (Recall) และเปรียบเทียบ โดยการประยุกต์ทั้งสองวิธีเข้าด้วยกันคือวิธี ฮาร์โมนิกมีน (Harmonic Mean) ดังจะอธิบายได้ดังนี้

### ตารางที่ 2.1 ความสัมพันธ์ของเอกสารที่ถูกค้นคืนกับความเกี่ยวข้อง

	เกี่ยวข้อง	ไม่เกี่ยวข้อง
ถูกค้นคืนออกมา	A	B
ไม่ถูกค้นคืนออกมา	C	D

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พรีซิชั่น (Precision) คือ สัดส่วนของจำนวนของเอกสารที่เกี่ยวข้องที่ค้นคืนได้เมื่อเปรียบเทียบกับจำนวนของเอกสารทั้งหมดที่ค้นออกมา

$$Precision = \frac{A}{A \cup B} \quad (\text{สมการที่ 2.9})$$

รีคอล (Recall) คือ สัดส่วนของจำนวนของเอกสารที่เกี่ยวข้องที่ค้นคืนได้เมื่อเปรียบเทียบกับจำนวนของเอกสารที่เกี่ยวข้องที่มีทั้งหมดในฐานข้อมูล

$$Recall = \frac{A}{A \cup C} \quad (\text{สมการที่ 2.10})$$

ฮาร์โมนิค มีน (Harmonic mean) เป็นการวัดที่รวมทั้งวิธี Recall และ Precision กำหนดให้

P = Precision และ R = Recall

$$Harmonic\ mean = \frac{2PR}{P+R} = \frac{2}{\frac{1}{R} + \frac{1}{P}} \quad (\text{สมการที่ 2.11})$$

เอฟเมเชอร์ (F-measure) เป็นการวัดที่รวมทั้งวิธี Recall และ Precision ได้ แต่เราสามารถให้ความสำคัญของ Recall และ Precision แยกต่างกันได้

$\alpha$  = ค่าถ่วงน้ำหนักของ Precision

$\alpha = 0$  เมื่อ ให้ค่าความสำคัญของ Precision อย่างเดียว

$\alpha = 0.5$  เมื่อ ให้ค่าความสำคัญของ Precision มากกว่า Recall

$\alpha = 1$  เมื่อ ให้ค่าความสำคัญของ Precision เท่ากับ Recall

$$F\ measure = \frac{(1+\alpha)PR}{\alpha P + R} \quad (\text{สมการที่ 2.12})$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2 เอกสารการบรรยายยูสเคส (Use Case Description)

ในการพัฒนาโปรแกรมในปัจจุบันจะใช้เอกสารการบรรยายยูสเคส ในการแปลงความต้องการของซอฟต์แวร์ของผู้ใช้งาน (Software Requirements Specification) ให้อยู่ในรูปแบบที่เข้าใจง่ายมีการบอกถึง ความต้องการของซอฟต์แวร์ ส่วนประกอบของซอฟต์แวร์ ผู้ที่จะติดต่อกับ ความต้องการซอฟต์แวร์ และอื่นๆ

### 2.2.1 ความหมายเบื้องต้นของเอกสารการบรรยายยูสเคส

เอกสารการบรรยายยูสเคสหมายถึงเอกสารที่ใช้แสดงหรืออธิบายรายละเอียดของยูสเคสซึ่งถูกเขียนในภาษาธรรมชาติ (Natural Language) เพื่อให้ง่ายต่อการเข้าใจโดยทั่วไปจะประกอบด้วย ส่วนประกอบ 9 ส่วนดังนี้

#### 1) ชื่อยูสเคส (Use Case Name)

ชื่อยูสเคสใช้ในการแยกความแตกต่างของเอกสารการบรรยายยูสเคสแต่ละอย่างกับ เอกสารบรรยายยูสเคสอื่น

#### 2) วัตถุประสงค์ (Objective)

เป็นส่วนประกอบของยูสเคสที่ใช้อธิบายเกี่ยวกับหน้าที่หลักของเอกสารการบรรยายยูสเคส นั้นๆ

#### 3) ผู้ใช้งานยูสเคส (Actor)

คือบุคคลหรือสิ่งต่างๆภายนอกระบบที่ติดต่อกับระบบที่สร้างขึ้น

#### 4) ความสัมพันธ์ (Relationship)

คือการปฏิสัมพันธ์ระหว่างส่วนประกอบต่างๆของโมเดลที่ออกแบบไว้ซึ่งส่วนนี้ ประกอบด้วย 4 ส่วนประกอบย่อยๆอีกดังนี้

- ความสัมพันธ์ (Association)

คือผู้ที่มีความเกี่ยวข้องกับยูสเคสนี้ทั้งการติดต่อหรือใช้งาน

- การรวม (Include)

คือสิ่งที่ต้องทำไปพร้อมกับความสัมพันธ์ที่เกิดขึ้น

- ส่วนเพิ่มเติม (Extend)

คือนอกเหนือจากที่กล่าวมาอาจมีการเพิ่มเติมได้อีกเป็นกรณีพิเศษ

- สายงาน (Generalization)

คืองานที่อยู่ในสายงานเดียวกัน

#### 5) เงื่อนไขก่อนการทำงาน (Precondition)

คือเงื่อนไขที่ต้องเป็นจริงก่อนที่ยูสเคสนั้นจะถูกเรียกไปใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**6) เงื่อนไขหลังการทำงาน (Postcondition)**

คือเงื่อนไขที่ต้องเป็นจริงเมื่อยูสเคสนั้นกระทำเสร็จสิ้นแล้ว

**7) ลำดับการทำงานปกติ (Normal flow of events)**

คือลำดับการเคลื่อนที่ไปของเหตุการณ์เมื่อมากระทำที่ยูสเคสนี้ การเขียนการไหลของเหตุการณ์ปกตินี้จะถูกสมมติว่าไม่มีข้อผิดพลาดเกิดขึ้น อาจเรียกอีกอย่างหนึ่งว่าเส้นทางพื้นฐาน (Basic path) หรือ เส้นทางปกติ (Normal path)

**8) เหตุการณ์ย่อย (Subflow)**

ในบางเอกสารการบรรยายยูสเคส การไหลของเหตุการณ์ปกติสามารถแบ่งเป็นกลุ่มของการไหลย่อยเพื่อให้ง่ายแก่การออกแบบ

**9) ลำดับทางเลือกการทำงาน (Alternative or exceptional flow of events)**

ใช้อธิบายเมื่อมีปัญหาจากการใช้เอกสารการบรรยายยูสเคสนี้หรือทางเลือกหรือวิธีการต่างๆที่ถูกนำมาใช้เมื่อเกิดปัญหา ส่วนประกอบนี้อาจเรียกอีกอย่างหนึ่งว่า เส้นทางทางเลือก (Alternative path)



<b>1507</b>		
<b>Use Case Name :</b>	Create examinations and examination solutions	
<b>Objective :</b>	This use case is to describe how an instructor creates examinations and their solutions.	
<b>Actor :</b>	An instructor, teacher	
<b>Relationship :</b>	<b>Association :</b>	An instructor, teacher
	<b>Include :</b>	-
	<b>Extend :</b>	-
	<b>Generalization :</b>	-
<b>Precondition :</b>	An instructor logged in to E-Learning system.	
<b>Postcondition :</b>	-	
<b>Normal Flow of Events :</b>	<ol style="list-style-type: none"> <li>1. An instructor selects a menu to create examination.</li> <li>2. The system displays a list of courses.</li> <li>3. An instructor selects a course.</li> <li>4. The system displays a list of subjects.</li> <li>5. An instructor selects a subject to create examination.</li> <li>6. An instructor defines a format of examination.</li> <li>7. An instructor defines details of examination such as time to test the examination, points.</li> <li>8. An instructor defines the number of topics in examination.</li> <li>9. An instructor writes the problems and their solutions.</li> <li>10. An instructor confirms the task.</li> </ol>	
<b>Subflow :</b>	-	
<b>Alternative or Exceptional Flow of Events :</b>	5a-1. An instructor cancels creating examination and its solutions. 5a-2. The system cancels creating examination and its solutions.	

**ภาพที่ 2.4 ตัวอย่างของเอกสารการบรรยายยูสเคส**

จากภาพที่ 2.4 เป็นตัวอย่างของเอกสารการบรรยายยูสเคสที่ผู้พัฒนาซอฟต์แวร์ส่วนใหญ่ในปัจจุบันนี้ใช้เป็นเครื่องมือในการพัฒนาดังกล่าว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### การดำเนินงานวิจัย

ในบทนี้จะอธิบายถึงวิธีการวิจัยรวมทั้งขั้นตอนวิธีการทดลองตามสมมติฐานที่ตั้งไว้ คือ การจัดเก็บและคั่นคืนเอกสารการบรรยายยูสเคสโดยใช้การคำนวณค่าน้ำหนักและค่าความคล้าย วิธีใดที่ทำให้ได้ผลลัพธ์ที่ตรงตามความต้องการของผู้ใช้มากที่สุด

#### 3.1 ขั้นตอนการดำเนินงานวิจัย

ในขั้นตอนของการวิจัยจะมีอยู่ 3 กระบวนการด้วยกัน คือ การเก็บรวบรวมเอกสารการบรรยายยูสเคสและการสร้างดัชนี การสร้างคิวิรีและกระบวนการคั่นคืน และการประเมินผล ระบบแต่ละขั้นตอนอธิบายอยู่ในรูปแผนภาพงาน ตามภาพที่ 3.2 และเพื่อเป็นการทดสอบผลลัพธ์ของงานวิจัยขั้นนี้ จึงได้มีการพัฒนาเครื่องมือขึ้นมาสำหรับการจัดเก็บเอกสารการบรรยายยูสเคส และการคั่นคืนเอกสารการบรรยายยูสเคส ขั้นตอนวิธีต่างๆสามารถอธิบายคร่าวๆได้ดังนี้

##### ขั้นที่ 1 : การเก็บรวบรวมเอกสารการบรรยายยูสเคสและการสร้างดัชนี

- 1) เอกสารการบรรยายยูสเคสที่พัฒนาจากขอบเขตงานตัวอย่างบางระบบ ถูกรวบรวมมาเก็บไว้
- 2) เอกสารการบรรยายยูสเคสจะถูกแปลงเป็นกลุ่มของดัชนีพร้อมค่าน้ำหนัก (อ่านรายละเอียดเพิ่มเติมในส่วน 3.1.1 และ 2.1.1.2)

##### ขั้นที่ 2 : การสร้างคิวิรี และกระบวนการคั่นคืน

- 3) ผู้ใช้งานจะต้องทำการสร้างคิวิรีและกำหนดค่าน้ำหนักให้แก่คิวิรี โดยใช้รูปแบบตามที่มีการกำหนดไว้ในการทดลอง ดังภาพที่ 3.1
- 4) คิวิรีในการคั่นคืนเอกสารการบรรยายยูสเคสของผู้ใช้ถูกแปลงเป็นกลุ่มดัชนี
- 5) เอกสารการบรรยายยูสเคสบางส่วนจากที่รวบรวมไว้จะถูกคั่นคืนขึ้นมา โดยขึ้นอยู่กับค่าความคล้าย (อ่านรายละเอียดเพิ่มเติมในส่วน 2.1.2.2) เทียบกับ คิวิรีของผู้ใช้
- 6) เอกสารการบรรยายยูสเคสที่ถูกคั่นคืนขึ้นมาจะแสดงผลแก่ผู้ใช้ (อ่านรายละเอียดเพิ่มเติมในส่วน 3.1.2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Use Case Retrieval System	
Query Use Case by Use Case Structure	
Weight By :	Select Method ▼
Similarity By :	Select Method ▼
Use Case Name:	<input type="text"/> Select Weigth Value ▼
Objective :	<input type="text"/> Select Weigth Value ▼
Actor :	<input type="text"/> Select Weigth Value ▼
Relationship :	Association : <input type="text"/> Select Weigth Value ▼
	Include : <input type="text"/> Select Weigth Value ▼
	Extend : <input type="text"/> Select Weigth Value ▼
	Generalization : <input type="text"/> Select Weigth Value ▼
Precondition :	<input type="text"/> Select Weigth Value ▼
Postcondition :	<input type="text"/> Select Weigth Value ▼
Normal Flow of Events :	<input type="text"/> Select Weigth Value ▼
Subflow :	<input type="text"/> Select Weigth Value ▼
Alternative or Exceptional Flow of Events :	<input type="text"/> Select Weigth Value ▼
<input type="button" value="Query"/> <input type="button" value="Reset"/>	
<a href="#">Back to Main Menu</a>	

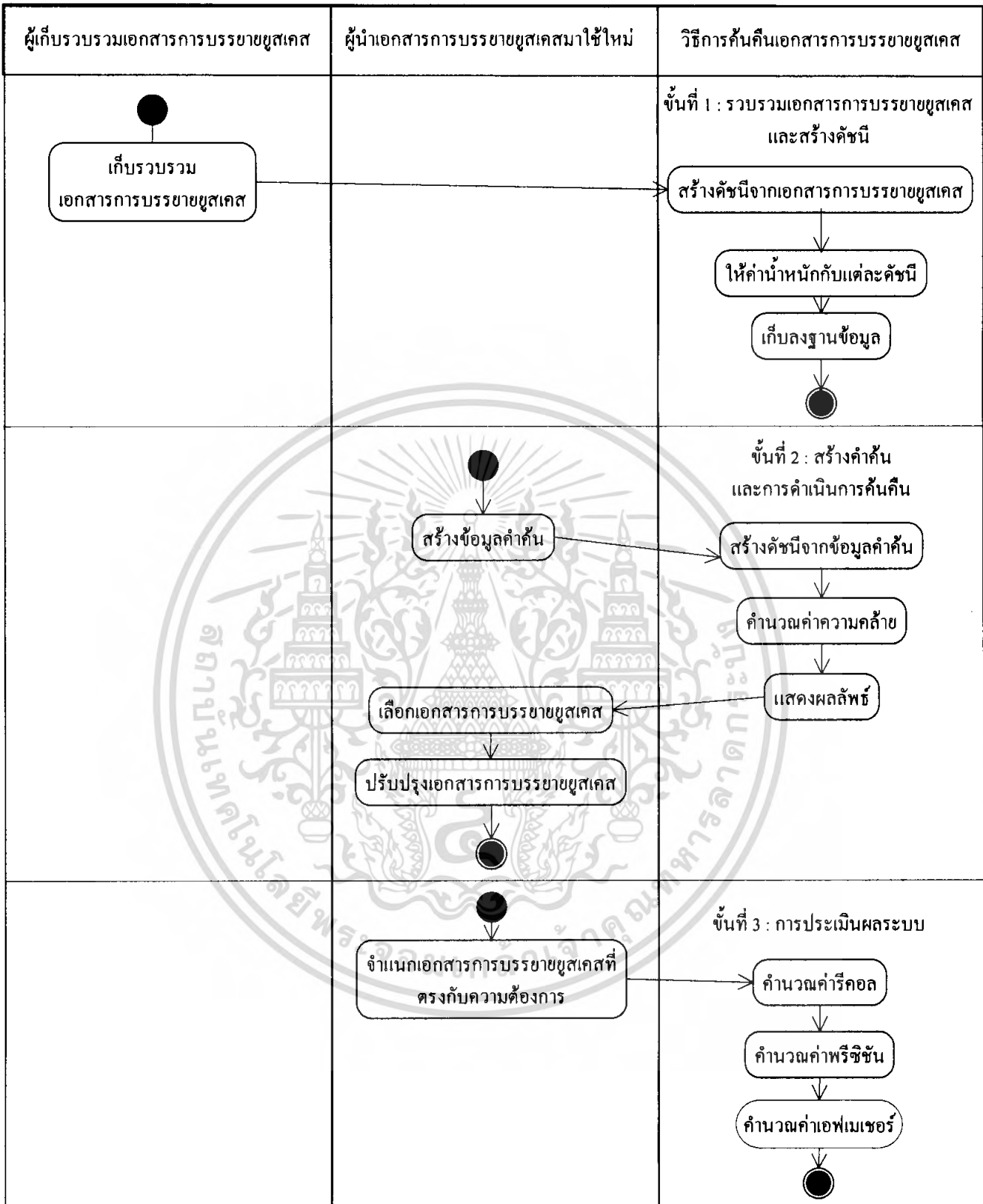
ภาพที่ 3.1 รูปแบบสำหรับการสร้างคิวรี

### ขั้นที่ 3 : การประเมินผลระบบ

- 7) ทำการประเมินผลเอกสารบรรยายขอบเขตที่ถูกค้นคืนมา ว่าเอกสารบรรยายขอบเขตอันไหนที่คล้ายกับคิวรีของผู้ใช้มากกว่ากัน เพื่อจะได้ทำการคำนวณค่ารีคอล (Recall) ค่าพรีซิชั่น (Precision) ค่าเฉลี่ยฮาร์โมนิก (Harmonic Mean) และค่าเอฟเมเชอร์ (F-measure) (อ่านรายละเอียดเพิ่มเติมในส่วน 3.1.3)

ซึ่งแต่ละขั้นตอนมีรายละเอียดการทำงาน แสดงได้ดังหัวข้อที่ 3.1.1 , 3.1.2 , 3.1.3

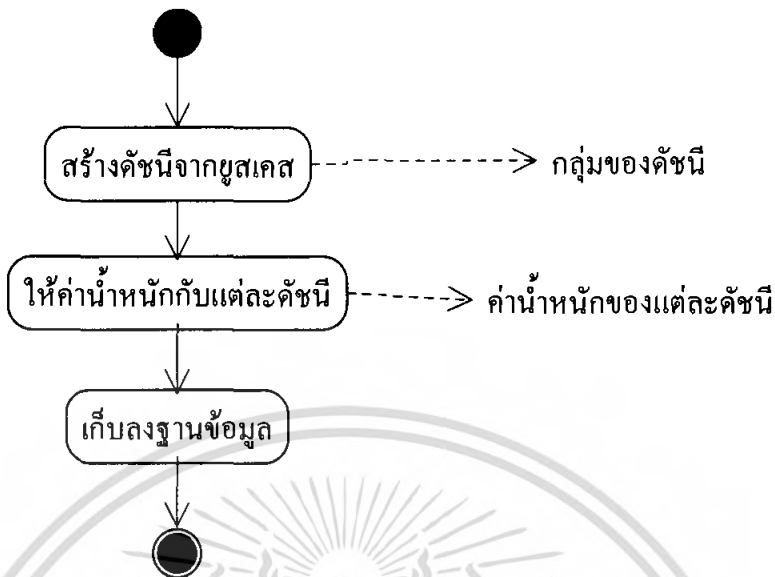
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 3.2 แผนภาพกิจกรรมของระบบทั้งหมดโดยสังเขป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 การเก็บรวบรวมเอกสารการบรรยายยูสเคส และการสร้างดัชนี  
กระบวนการในขั้นนี้อธิบายได้ดังภาพที่ 3.3



ภาพที่ 3.3 แผนภาพกิจกรรมการสร้างดัชนี

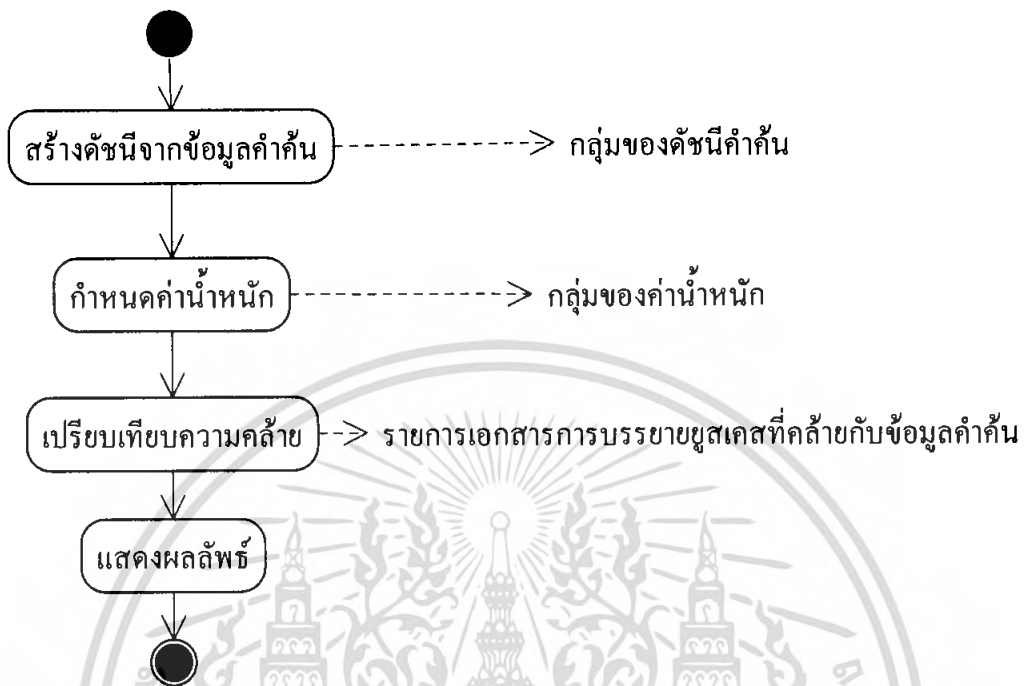
- หลังผู้ใช้รวบรวมเอกสารการบรรยายยูสเคสได้แล้ว เนื้อหาที่อยู่ในเอกสารการบรรยายยูสเคสจะถูกนำไปวิเคราะห์และแปลงเป็นกลุ่มของดัชนีโดยใช้กระบวนการสร้างดัชนีอัตโนมัติตามที่ระบุไว้ในส่วนที่ 2.1.1 ผลลัพธ์ที่ได้จากขั้นนี้คือ กลุ่มของดัชนี ซึ่งจะทำการสร้างและเก็บค่าถ่วงน้ำหนักแต่ละแบบของแต่ละดัชนีทั้งหมด 3 แบบ
- ในแบบที่ 1 ของดัชนี ดัชนีแต่ละตัวจะถูกให้น้ำหนักโดยวิธี TF(Term Frequency) ที่กล่าวถึงในส่วน 2.1.1.2.1)
- ในแบบที่ 2 ของดัชนี ดัชนีแต่ละตัวจะถูกให้น้ำหนักโดยวิธี IDF (Inverse Document Frequency) ที่กล่าวถึงในส่วน 2.1.1.2.2)
- ในแบบที่ 3 ของดัชนี ดัชนีแต่ละตัวจะถูกให้น้ำหนักโดยวิธี TF•IDF ที่กล่าวถึงในส่วน 2.1.1.2.3)
- ดัชนีพร้อมค่าน้ำหนักจะถูกเก็บไว้ในฐานข้อมูล

82794

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.2 การสร้างคิวิรีและกระบวนการค้นคืน

กระบวนการในขั้นนี้อธิบายได้ดังภาพที่ 3.4



ภาพที่ 3.4 แผนภาพกิจกรรมกระบวนการค้นคืน

- เริ่มแรกผู้ใช้จะต้องสร้างเอกสารการบรรยายชุดศาสตร์ อย่างง่าย ๆ ขึ้นมาก่อนเพื่อเป็นข้อมูลคิวิรีซึ่งเอกสารการบรรยายชุดศาสตร์ นั้นจะประกอบด้วยคำสำคัญจากนั้นมันจะถูกแปลงเป็นกลุ่มของดัชนีด้วยกระบวนการสร้างดัชนีอัตโนมัติเช่นเดียวกับในขั้นการสร้างดัชนี ซึ่งดัชนีมีการให้ค่าน้ำหนักทั้งหมด 3 แบบ
- ผู้ใช้กำหนดน้ำหนักความสำคัญให้กับแต่ละองค์ประกอบในเอกสารการบรรยายชุดศาสตร์ ที่เป็นคิวิรี และค่าน้ำหนักเหล่านี้จะใช้ในการคำนวณหาค่าความคล้าย
- ค่าคะแนนความคล้ายระหว่างแต่ละองค์ประกอบในเอกสารการบรรยายชุดศาสตร์ ที่เป็นคิวิรีและทุกเอกสารการบรรยายชุดศาสตร์ ในฐานะข้อมูลจะถูกคำนวณจากกรณี que ที่ผู้ใช้เลือกเพียงกรณีเดียวจากทั้งหมด 15 กรณี ซึ่งกรณีทั้ง 15 กรณีสามารถแสดงได้ดังตารางที่ 3.1 ผลลัพธ์ที่ได้จากขั้นนี้คือ ค่าคะแนนความคล้ายของแต่ละองค์ประกอบในคิวิรี และเอกสารการบรรยายชุดศาสตร์ ในฐานะข้อมูล

การทดลองนี้จะใช้สมการและวิธีต่างๆ ดังที่กล่าวไว้ในบทที่ 2 ซึ่งสามารถแบ่งออกได้เป็น 15 กรณี รายละเอียดต่างๆ ของสมการดูได้ที่ 2.1.1.2) และ 2.1.2.2) ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 กรณีที่เกิดขึ้นในการค้นคืนเอกสารการบรรยายยูสเคส

กรณีที่	วิธีถ่วงน้ำหนัก	วิธีคำนวณความคล้าย
1	TF	Dice
2	TF	Jaccard
3	TF	Cosine
4	TF	Overlap
5	TF	Asymmetric
6	IDF	Dice
7	IDF	Jaccard
8	IDF	Cosine
9	IDF	Overlap
10	IDF	Asymmetric
11	TF-IDF	Dice
12	TF-IDF	Jaccard
13	TF-IDF	Cosine
14	TF-IDF	Overlap
15	TF-IDF	Asymmetric

- ค่าคะแนนความคล้ายระหว่างเอกสารการบรรยายยูสเคส[1] ที่เป็นคิวรีและทุกเอกสารการบรรยายยูสเคส ในฐานข้อมูลจะถูกคำนวณด้วยสมการ สมการที่ 3.1 (คำอธิบายต่างๆของสมการดูได้ที่ 2.1.2.3) ผลลัพธ์ที่ได้จากขั้นนี้คือ ค่าคะแนนความคล้ายของระหว่างคิวรีและเอกสารการบรรยายยูสเคสที่เก็บอยู่ในฐานข้อมูล

$$Similarity(UC_i, Query_j) = \frac{\sum_{m=1}^n [Similarity(E_{m_i}, E_{m_j}) \cdot WE_m]}{TotalWeight} \quad (\text{สมการที่ 3.1})$$

กำหนดให้

$Similarity(UC_i, Query_j)$  คือ ค่าความคล้ายระหว่างคิวรีเอกสารการบรรยายยูสเคสของผู้ทดสอบและเอกสารการบรรยายยูสเคสที่รวบรวมไว้

$Similarity(E_{m_i}, E_{m_j})$  คือ ค่าความคล้ายระหว่างแต่ละสมาชิกของเอกสารการบรรยายยูสเคส  $i$  และคิวรีเอกสารการบรรยายยูสเคส  $j$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$WE_m$  คือ ค่าน้ำหนักของสมาชิก  $m$  ที่ถูกกำหนดโดยผู้ทดสอบ

$TotalWeight$  คือ ผลรวมของค่าน้ำหนัก ( $WE_m$ ) ของทุกองค์ประกอบ

- เอกสารการบรรยายยูสเคสที่มีค่าคะแนนความคล้ายจากการคำนวณด้วยสมการ (สมการที่ 3.1) เป็นค่าสูงสุดจะถูกแสดงผลออกมา ตามลำดับ
- ค่าขีดแบ่ง (Threshold) คือค่าที่กำหนดมาเพื่อใช้เป็นเกณฑ์ในการระบุว่าเอกสารการบรรยายยูสเคสมีค่าความคล้ายกับคิวรีมากน้อยเพียงไร โดยค่าขีดแบ่งคำนวณได้จากสมการ

$$\text{ค่าขีดแบ่ง} = \text{Mean} - \text{S.D.} \quad (\text{สมการที่ 3.2})$$

$$\text{ค่าขีดแบ่ง} = \text{Mean} \quad (\text{สมการที่ 3.3})$$

$$\text{ค่าขีดแบ่ง} = \text{Mean} + \text{S.D.} \quad (\text{สมการที่ 3.4})$$

กำหนดให้

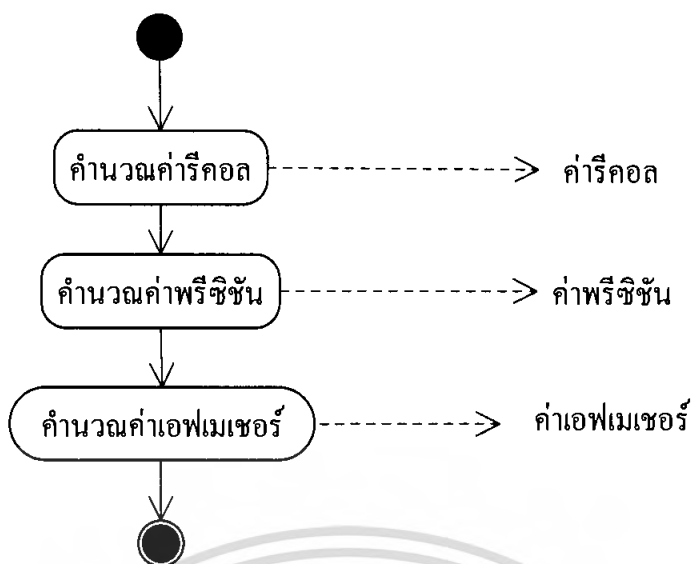
Mean คือ ค่าเฉลี่ยของค่าความคล้ายสำหรับทุกคิวรี

S.D. คือ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของค่าความคล้าย สำหรับทุกคิวรี

### 3.1.3 การประเมินผลระบบ

กระบวนการในขั้นตอนนี้อธิบายได้ดังภาพที่ 3.5

- ขั้นแรก ผู้ใช้จะต้องทำการระบุเอกสารการบรรยายยูสเคส ทั้งหมดที่ตรงกับความต้องการ ข้อมูลนี้จะนำไปคำนวณค่ารีคอด ค่าพรีซิชันและค่าเฉลี่ยฮาร์โมนิค
- ค่าทั้งหมดคำนวณได้โดยสมการ 2.9, 2.10, และ 2.12 รายละเอียดอยู่ในหัวข้อ 2.1.3



ภาพที่ 3.5 แผนภาพกิจกรรมการประเมินผล

### 3.2 การดำเนินการทดลอง

ขั้นตอนในการดำเนินการทดลองซึ่งมีการดำเนินงานดังนี้

#### 3.2.1 การทดลอง

จากสมมติฐานที่ได้กล่าวไว้แล้ว ทำให้สามารถทำการทดลองเพื่อทดสอบสมมติฐานได้โดยทำการทดลองกับ โปรแกรมการค้นคืนเอกสารการบรรยายชุดที่สร้างขึ้น โดยใช้กรณีที่เป็นไปได้ทั้งหมด 15 กรณี (ตามตารางที่ 3.1) ซึ่งการทดลองจะมีการควบคุมตัวแปรต่างๆ ดังนี้

- เอกสารการบรรยายชุดที่เก็บรวบรวม มีทั้งหมด 315 เอกสาร จาก 16 ระบบ ซึ่งแสดงได้ดังตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ระบบของเอกสารการบรรยายยูสเคสที่นำมาให้ทดลอง

โดเมน	ชื่อเอกสารการบรรยายยูสเคส
1	ICT blocking and filtering system
2	Departmental information system for curriculum and course offerings management
3	Resource management system
4	Automatic restaurant lookup system
5	Library management system
6	Mobile phone selling on the internet system
7	Payroll system
8	Personal stock manager on PDA system
9	Basic product trading system
10	Mobile phone service paying system
11	MLM computerize system
12	Content management and platform upgrade for the official website of Nation Channel
13	Purchase, producing, and export system
14	Management information system for insurance company
15	E-learning system
16	Management information system for an academic

- เอกสารที่ถูกเก็บรวบรวมจากระบบการออกแบบต่างๆ ซึ่งเขียนเป็นภาษาอังกฤษ
- ผู้ทำการทดลอง มีผู้ทดสอบ 10 คน ที่ทำการค้นคืนเอกสารการบรรยายยูสเคสซึ่งทุกคนศึกษาปริญญาโทในสาขาวิศวกรรมซอฟต์แวร์ และมีประสบการณ์ในการออกแบบเอกสารการบรรยายยูสเคสและมีความรู้ด้านภาษาอังกฤษเป็นอย่างดี เนื่องจากมีคะแนนสอบ TOEFL ประมาณ 500 คะแนน
  - หัวข้อที่ทำการทดลองค้นคืน มี 5 หัวข้อที่กำหนดให้ผู้ทดสอบทำการค้นคืนจากระบบการค้นคืนเอกสารการบรรยายเอกสารการบรรยายยูสเคส โดยผู้ทดสอบจะทำการสืบค้นเอกสารการบรรยายยูสเคสจากระบบของเอกสารการบรรยายยูสเคสที่มีอยู่ 16 ระบบ ซึ่งการสืบค้นเอกสารการบรรยายยูสเคสทำโดยการสร้างประโยคควิรีที่ สอดคล้องกับทั้ง 5 หัวข้อต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ก) ระบบการเรียน การสอน
  - ข) ระบบการค้าขายสินค้า
  - ค) การจัดการข้อมูลลูกค้า
  - ง) การคำนวณการเงิน
  - จ) การสร้างรายงาน
- ผู้ทำการทดสอบแต่ละคนสามารถทำการค้นคืนได้คนละ 5 หัวข้อ และแต่ละหัวข้อสามารถทำการค้นคืนได้ 5 ครั้ง โดยผู้ทดสอบสามารถสร้างคิวรีตามรูปแบบที่กำหนดดังภาพที่ 3.1

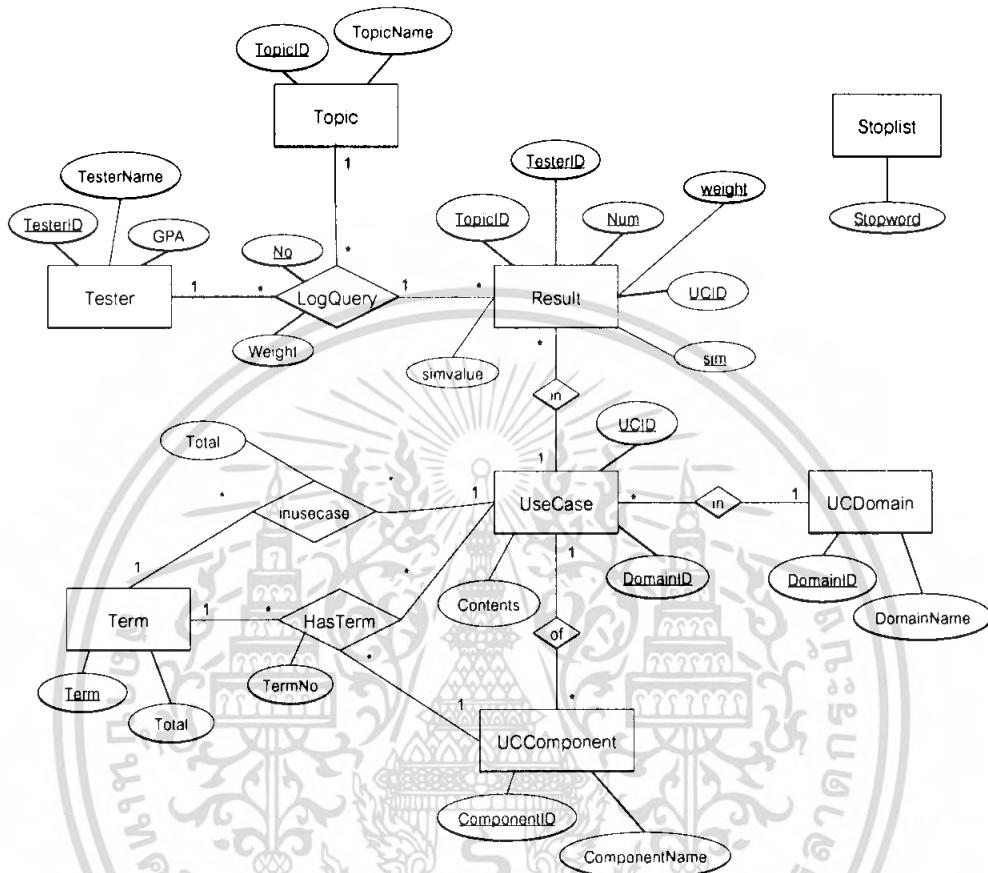
### 3.3 การออกแบบฐานข้อมูล

เอกสารการบรรยายยูสเคสที่อยู่ในระบบค้นคืนนั้นจะถูกแปลงและเก็บใน ระบบการจัดการฐานข้อมูลเชิงสัมพันธ์ ข้อมูลที่เกิดขึ้นในฐานข้อมูลไม่ใช่เพียงข้อมูลของเอกสารการบรรยายยูสเคสเท่านั้น แต่ยังมีข้อมูลการทดสอบระบบด้วย เช่น ข้อมูลของผู้ทำการทดลอง ข้อมูลของคิวรี และข้อมูลผลลัพธ์ ข้อมูลทั้งหมดที่เกิดขึ้นในระบบการจัดการฐานข้อมูลเชิงสัมพันธ์นี้ถูกออกแบบให้อยู่ในแผนภาพแสดงความสัมพันธ์ของเอนติตี (Entity Relationship Diagram) ซึ่งรายละเอียดอยู่ในหัวข้อ 3.3.1 ข้อมูลทั้งหมดในระบบการจัดการฐานข้อมูลเชิงสัมพันธ์สามารถแยกออกได้เป็น 12 ตาราง รายละเอียดของตารางต่างๆ มีรายละเอียดอยู่ดังด้านล่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.3.1 ความสัมพันธ์ของแต่ละเอนิตี

เอนิตีที่แสดงรายละเอียดและความสัมพันธ์ของแต่ละส่วนประกอบของการออกแบบฐานข้อมูล



ภาพที่ 3.6 แผนภาพความสัมพันธ์ของเอนิตีของระบบค้นคืนเอกสารการบรรยายยูลเศส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.3.2 รายละเอียดของตารางต่างๆ

จากการออกแบบระบบฐานข้อมูลเพื่อใช้จัดเก็บดัชนีของคำและเอกสารการบรรยาย ยูสเคสสามารถแจกแจงออกเป็นตารางได้ดังนี้

ตารางที่ 3.3 รายละเอียดของเอกสารการบรรยายยูสเคส

ชื่อตาราง : UseCase		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมรายละเอียดเอกสารการบรรยายยูสเคส		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
UCID	Char (4)	รหัสของเอกสารการบรรยายยูสเคส
UCDomainID	Char (2)	รหัสของระบบเอกสารการบรรยายยูสเคส
Name	Text	ชื่อของเอกสารการบรรยายยูสเคส
Objective	Text	วัตถุประสงค์ของเอกสารการบรรยายยูสเคส
Actor	Text	ผู้กระทำของเอกสารการบรรยายยูสเคส
Association	Text	ความสัมพันธ์ที่เชื่อมโยงของเอกสารการบรรยายยูสเคส
Include	Text	สิ่งที่รวมอยู่ในความสัมพันธ์ของเอกสารการบรรยายยูสเคส
Extend	Text	สิ่งที่เพิ่มอยู่ในความสัมพันธ์ของเอกสารการบรรยายยูสเคส
Generalization	Text	ความสัมพันธ์ทั่วไปของเอกสารการบรรยายยูสเคส
Precondition	Text	เงื่อนไขก่อนซึ่งต้องเป็นจริงก่อนที่เอกสารการบรรยายยูสเคสนี้จะถูกเรียกใช้
Postcondition	Text	เงื่อนไขหลังซึ่งต้องเป็นจริงหลังที่เรียกใช้เอกสารการบรรยายยูสเคสนี้แล้ว
Normalflow	Text	การไหลปกติของเอกสารการบรรยายยูสเคส
Subflow	Text	การไหลส่วนย่อยของเอกสารการบรรยายยูสเคส
Alternativeflow	Text	ทางเลือกการไหลของเอกสารการบรรยายยูสเคส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 ระบบของเอกสารการบรรยายยูสเคส

ชื่อตาราง : UCDomain		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมระบบของเอกสารการบรรยายยูสเคส		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
DomainID	Char (2)	รหัสของระบบเอกสารการบรรยายยูสเคส
DomainName	Char (50)	ชื่อของระบบเอกสารการบรรยายยูสเคส

ตารางที่ 3.5 ส่วนประกอบย่อยของเอกสารการบรรยายยูสเคส

ชื่อตาราง : UCComponent		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมแต่ละสมาชิกขององค์ประกอบเอกสารการบรรยายยูสเคส		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
ComponentID	Char (2)	รหัสขององค์ประกอบเอกสารการบรรยายยูสเคส
ComponentName	Char (50)	ชื่อขององค์ประกอบเอกสารการบรรยายยูสเคส

ตารางที่ 3.6 คำทั้งหมดในฐานข้อมูล

ชื่อตาราง : Term		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมคำทั้งหมดและจำนวนของคำที่ได้เก็บรวบรวม		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
Term	Char (20)	คำ
Total	Integer	จำนวนของคำที่ได้เก็บรวบรวม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ตารางที่ 3.7 คำทั้งหมดในแต่ละเอกสารการบรรยายยูสเคส

ชื่อตาราง : HasTerm		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมคำทั้งหมดและจำนวนของคำที่ได้เก็บรวบรวมในแต่ละเอกสารการบรรยายยูสเคส		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
Term	Char (20)	คำ
DomainID	Char (2)	รหัสของระบบเอกสารการบรรยายยูสเคส
UCID	Char (4)	รหัสของเอกสารการบรรยายยูสเคส
ComponentID	Char (2)	รหัสขององค์ประกอบเอกสารการบรรยายยูสเคส
TermNo	Integer	จำนวนของคำที่ปรากฏในเอกสารการบรรยายยูสเคส

### ตารางที่ 3.8 รายการคำที่ไม่มีประโยชน์ในการค้นคืน

ชื่อตาราง : Stoplist		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมคำที่ไม่มีความหมายสำคัญซึ่งต้องกำจัดทิ้งไป		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
Stopword	Char (20)	คำที่ไม่มีประโยชน์ในการค้นคืน

### ตารางที่ 3.9 ผู้ทดสอบ

ชื่อตาราง : Tester		
คำอธิบายตาราง : ตารางนี้เก็บรวบรวมผู้ทดสอบทั้งหมดในการทดลอง		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
TesterID	Char (2)	รหัสของผู้ทดสอบ
TesterName	Char (50)	ชื่อของผู้ทดสอบ
GPA	Char (4)	เกรดเฉลี่ยของผู้ทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.10 ข้อมูลคิวรีของผู้ทดสอบที่บันทึกไว้โดยใช้วิธีค้นคืนด้วยโครงสร้างของ เอกสารการ  
บรรยายยูสเคส

ชื่อตาราง : LogQuery		
รายละเอียดตาราง : ตารางนี้เก็บรวบรวมค่าทั้งหมดในคิวรีโดยใช้วิธีค้นคืนด้วยโครงสร้างของเอกสารการบรรยายยูสเคสของผู้ทดสอบในการทดลอง		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
TesterID	Char (2)	รหัสของผู้ทดสอบ
Topic	Char (1)	หมายเลขหัวข้อ
No	Integer	หมายเลขคิวรี
ComponentID	Char (2)	รหัสขององค์ประกอบเอกสารการบรรยายยูสเคส
Weight_user	Integer	ค่าน้ำหนักของแต่ละองค์ประกอบที่ผู้ใช้เลือก
query	Char (20)	ค่าที่เป็นคิวรี

ตารางที่ 3.11 หัวข้อ

ชื่อตาราง : Topic		
รายละเอียดตาราง : ตารางนี้เก็บรวบรวมหัวข้อทั้งหมดในการทดลอง		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
TopicID	Char (1)	รหัสของหัวข้อ
TopicName	Char (100)	ชื่อของหัวข้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.12 ผลลัพธ์ของการค้นคืน

ชื่อตาราง : Result		
รายละเอียดตาราง : ตารางที่เก็บรวบรวมผลลัพธ์ทั้งหมดของคิวรีเอกสารการบรรยายอุตสาหกรรมในการทดลอง		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
TesterID	Char (2)	รหัสของผู้ทดสอบ
Topic	Char (1)	หมายเลขหัวข้อ
No	Integer	จำนวนครั้งที่ของคิวรี
UCID	Char (4)	รหัสของเอกสารการบรรยายอุตสาหกรรมที่ถูกค้นคืน
Weight	Varchar(10)	วิธีการถ่วงน้ำหนัก
Sim	Varchar(20)	วิธีการหาค่าความคล้าย
Simvalue	Float	ค่าความคล้ายที่ได้

ตารางที่ 3.13 จำนวนคำที่ปรากฏในแต่ละเอกสาร

ชื่อตาราง : inusecase		
รายละเอียดตาราง : ตารางที่เก็บรวบรวมจำนวนคำที่ปรากฏในแต่ละเอกสารการบรรยายอุตสาหกรรม		
ชื่อ	ประเภทข้อมูล	คำอธิบาย
Term	Char (20)	คำแต่ละคำ
UCID	Char (4)	หมายเลขเอกสารการบรรยายอุตสาหกรรม
Total	Integer	จำนวนความถี่ของแต่ละคำในแต่ละเอกสารการบรรยายอุตสาหกรรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### ผลการทดลอง

ในบทนี้จะกล่าวถึงผลที่ได้จากการทดลอง ซึ่งประกอบด้วย ค่าขีดแบ่งของการทดลองและผลการวัดประสิทธิภาพของแต่ละกรณีโดยมีรายละเอียดดังต่อไปนี้

#### 4.1 ค่าขีดแบ่งที่ได้จากการทดลอง

ในการทดลองนี้จะต้องมีการระบุค่าขีดแบ่ง (Threshold) เพื่อใช้เป็นค่าที่ระบุว่าเอกสารใดควรที่จะถูกนำมาแสดงซึ่งเอกสารที่จะนำมาแสดงต้องมีค่าความคล้ายที่มากกว่าค่าขีดแบ่ง (ซึ่งได้กล่าวถึงค่าขีดแบ่งแล้วใน 3.1.2) ค่าขีดแบ่งดังกล่าวจะไม่เท่ากันตามแต่ละวิธีถ่วงน้ำหนักและวิธีการวิเคราะห์ความคล้าย แต่ในการทดลองนี้เรายังไม่รู้ว่ค่าขีดแบ่งที่คำนวณจากวิธีใดเหมาะสมดังนั้นเราต้องศึกษาหลายวิธีซึ่งวิธีการได้ระบุไว้ในสมการที่ 3.2 ถึงสมการที่ 3.4 แต่ในส่วนนี้จะขอทบทวนรายละเอียดใหม่ ซึ่งวิธีที่จะนำมาใช้ในการคำนวณค่าขีดแบ่งมีดังนี้

กำหนดให้ Mean เป็น ค่าเฉลี่ยของค่าความคล้ายสำหรับทุกคำค้น

S.D. เป็น ค่าส่วนเบี่ยงเบนมาตรฐาน(Standard Deviation) ของค่าความคล้าย

สำหรับทุกคำค้น

1. ค่าขีดแบ่งเท่ากับ  $\text{Mean} - \text{S.D.}$
2. ค่าขีดแบ่งเท่ากับ Mean
3. ค่าขีดแบ่งเท่ากับ  $\text{Mean} + \text{S.D.}$

ซึ่งได้แสดงในตารางที่ 4.1 ถึงสมการที่ 4.3 ตามลำดับ



ตารางที่ 4.2 ค่าชี้แจงที่ได้จากการทดลอง เมื่อ ค่าชี้แจงเท่ากับMean

วิธีการถ่วงน้ำหนัก	วิธีการคำนวณความคล้าย	ค่าชี้แจง
TF	Asymmetric	0.00112181
TF	Cosine	0.0109732
TF	Dice	0.00955699
TF	Jaccard	0.00576749
TF	Overlap	0.131985
IDF	Asymmetric	0.00188446
IDF	Cosine	0.00272437
IDF	Dice	0.000856685
IDF	Jaccard	0.000442794
IDF	Overlap	0.00175609
TF-IDF	Asymmetric	0.00303199
TF-IDF	Cosine	0.00508433
TF-IDF	Dice	0.00353408
TF-IDF	Jaccard	0.00206965
TF-IDF	Overlap	0.00721025

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ค่าขีดแบ่งที่ได้จากการทดลอง เมื่อ ค่าขีดแบ่งเท่ากับ Mean+S.D.

วิธีการถ่วงน้ำหนัก	วิธีการคำนวณความคล้าย	ค่าขีดแบ่ง
IDF	Asymmetric	0.0165199
IDF	Cosine	0.0159517
IDF	Dice	0.00483035
IDF	Jaccard	0.00258956
IDF	Overlap	0.00992216
TF	Asymmetric	0.00679506
TF	Cosine	0.0444485
TF	Dice	0.0393553
TF	Jaccard	0.0246415
TF	Overlap	0.633674
TF-IDF	Asymmetric	0.0259076
TF-IDF	Cosine	0.028295
TF-IDF	Dice	0.0206411
TF-IDF	Jaccard	0.013326
TF-IDF	Overlap	0.0476709

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.2 ค่าเฉลี่ยของรีคอล พรีซิชั่น และค่าเฉลี่ยฮาร์โมนิกของทุกคำค้นทั้ง 15 กรณี

จากการทดลองทำการค้นคืนเอกสารการบรรยายยูสเคสทำให้ได้เอกสารจำนวนหนึ่งที่มีความคล้ายมากกว่าค่าขีดแบ่ง ซึ่งเอกสารที่ค้นคืนได้นั้นอาจจะตรงความต้องการของผู้ใช้หรืออาจไม่ตรงกับความต้องการของผู้ใช้ ซึ่งทำให้เกิดค่า รีคอล พรีซิชั่น และค่าเฉลี่ยฮาร์โมนิกของเอกสารที่ถูกค้นคืนของแต่ละกรณีการทดลองทั้ง 15 กรณีโดยขึ้นอยู่กับค่าขีดแบ่งดังที่ได้แสดงในหัวข้อ 4.1 โดยมีรายละเอียดดังนี้

##### ▪ เมื่อค่าขีดแบ่งเท่ากับ Mean - S.D.

ซึ่งจะทำให้ไม่สามารถกำจัดเอกสารการบรรยายยูสเคสที่มีค่าความคล้ายน้อยๆซึ่งอาจไม่เกี่ยวข้องออกไปได้ซึ่งทำให้ได้เอกสารที่ไม่ค่อยจะเกี่ยวข้องมากเกินไปและมีความแม่นยำในกาค้นคืนน้อยลงด้วย ซึ่งจากการทดลองทุกกรณีจะให้ค่า รีคอล พรีซิชั่นและ ฮาร์โมนิกมินในหัวข้อการทดลองเดียวกันเป็นค่าเดียวกันหมดไม่ว่ากรณีที่ใช้ทดลองใดก็ตาม ซึ่งในที่นี้จะแสดงแค่วิธีเดียวคือวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่าน้ำหนักแบบต่างๆ ดังแสดงในตารางที่ 4.4 โดยที่วิธีอื่นๆก็ให้ค่าในการทำงานเดียวกัน

ตารางที่ 4.4 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ยลบด้วยค่าส่วนเบี่ยงเบนมาตรฐาน

ค่าความคล้าย	ค่าน้ำหนัก	TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
ไดซ์	หัวข้อ	R	P	H	R	P	H	R	P	H
	ก	0.4117	0.6726	0.4695	0.4117	0.6726	0.4695	0.4117	0.6726	0.4695
	ข	0.3398	0.2630	0.2665	0.3398	0.2630	0.2665	0.3398	0.2630	0.2665
	ค	0.7375	0.2646	0.3328	0.7375	0.2646	0.3328	0.7375	0.2646	0.3328
	ง	0.4635	0.5079	0.4390	0.4635	0.5079	0.4390	0.4635	0.5079	0.4390
	จ	0.5500	0.2497	0.3148	0.5500	0.2497	0.3148	0.5500	0.2497	0.3148
	ค่าเฉลี่ย	0.5005	0.3915	0.3645	0.5005	0.3915	0.3645	0.5005	0.3915	0.3645

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ■ เมื่อค่าจัดแบ่งเท่ากับ Mean

ถ้าค่าจัดแบ่งเท่ากับค่าเฉลี่ยจะทำให้ได้ค่าพรีซิชั่นที่สูงขึ้น กว่าการใช้ค่าจัดแบ่งที่เท่ากับค่าเฉลี่ยลบด้วยค่าส่วนเบี่ยงเบนมาตรฐาน ซึ่งจากการวิเคราะห์จะเห็นได้ว่าค่าเฉลี่ยนี้เหมาะกับการนำมาใช้เป็นค่าจัดแบ่งเพราะให้จำนวนของผลการค้นคืนของเอกสารที่เกี่ยวข้องที่พอเหมาะและมี ความแม่นยำที่สูง

ตารางที่ 4.5 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ย

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
ไดซ์	หัวข้อ									
	ก	0.3375	0.6712	0.4491	0.3899	0.6904	0.4984	0.3073	0.7164	0.4301
	ข	0.3352	0.2684	0.2981	0.2906	0.2460	0.2664	0.2766	0.2649	0.2706
	ค	0.7357	0.2698	0.3949	0.6554	0.2874	0.3996	0.6000	0.2991	0.3992
	ง	0.4625	0.5185	0.4889	0.4604	0.5278	0.4918	0.4521	0.5556	0.4985
	จ	0.5359	0.2536	0.3443	0.5172	0.2700	0.3548	0.4781	0.2686	0.3440
	ค่าเฉลี่ย	0.4814	0.3963	0.3950	0.4627	0.4043	0.4022	0.4228	0.4209	0.3885

ตารางที่ 4.6 ผลการทดลองของวิธีวัดความคล้ายแบบแจ็กการ์ด (Jaccard) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ย

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
แจ็กการ์ด	หัวข้อ									
	ก	0.3895	0.6925	0.4457	0.3895	0.6925	0.4986	0.3040	0.7312	0.4295
	ข	0.2906	0.2463	0.2958	0.2906	0.2463	0.2666	0.2664	0.2621	0.2642
	ค	0.6482	0.2867	0.3979	0.6482	0.2867	0.3976	0.5893	0.3042	0.4013
	ง	0.4604	0.5288	0.4903	0.4604	0.5288	0.4923	0.4469	0.5614	0.4976
	จ	0.5141	0.2705	0.3456	0.5141	0.2705	0.3545	0.4719	0.2724	0.3454
	ค่าเฉลี่ย	0.4783	0.3983	0.3950	0.4606	0.4050	0.4019	0.4157	0.4262	0.3876

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ผลการทดลองของวิธีวัดความคล้ายแบบโคไซน์ (Cosine) ซึ่งใช้ค่านำหนักแบบ  
ต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		หัวข้อ	R	P	H	R	P	H	R	P
โคไซน์	ก	0.3411	0.6756	0.3411	0.3702	0.7006	0.4844	0.2883	0.7230	0.4122
	ข	0.3336	0.2639	0.3336	0.2578	0.2450	0.2512	0.2672	0.2688	0.2680
	ค	0.7321	0.2696	0.7321	0.6179	0.3089	0.4118	0.6411	0.3194	0.4264
	ง	0.4625	0.5184	0.4625	0.4521	0.5427	0.4932	0.4000	0.5413	0.4601
	จ	0.5422	0.2519	0.5422	0.4906	0.2683	0.3469	0.4703	0.2711	0.3439
	ค่าเฉลี่ย	0.4823	0.3959	0.3950	0.4377	0.4131	0.3975	0.4134	0.4247	0.3821

ตารางที่ 4.8 ผลการทดลองของวิธีวัดความคล้ายแบบโอเวอร์แลป (Overlap) ซึ่งใช้ค่านำหนักแบบ  
ต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		หัวข้อ	R	P	H	R	P	H	R	P
โอเวอร์แลป	ก	0.3819	0.6969	0.4934	0.3819	0.6969	0.4934	0.3190	0.7363	0.4451
	ข	0.3063	0.2630	0.2830	0.3063	0.2630	0.2830	0.2883	0.2787	0.2834
	ค	0.6161	0.2938	0.3978	0.6161	0.2938	0.3978	0.5536	0.2927	0.3829
	ง	0.4563	0.5337	0.4920	0.4563	0.5337	0.4920	0.4417	0.5599	0.4938
	จ	0.4969	0.2657	0.3462	0.4969	0.2657	0.3462	0.4516	0.2647	0.3338
	ค่าเฉลี่ย	0.4731	0.3998	0.3951	0.4515	0.4106	0.4025	0.4108	0.4265	0.3878

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 ผลการทดลองของวิธีวัดความคล้ายแบบอะซิมเมตริก (Asymmetric) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ย

ค่าความคล้าย	ค่านำหนัก	TF			IDF			TF-IDF		
	หัวข้อ	R	P	H	R	P	H	R	P	H
อะซิมเมตริก	ก	0.2125	0.6393	0.3190	0.2125	0.6393	0.3190	0.1835	0.6560	0.2867
	ข	0.1992	0.3454	0.2527	0.1992	0.3454	0.2527	0.1734	0.3496	0.2318
	ค	0.3232	0.1900	0.2394	0.3232	0.1900	0.2394	0.2786	0.2029	0.2348
	ง	0.3031	0.4515	0.3627	0.3031	0.4515	0.3627	0.2969	0.4911	0.3700
	จ	0.2328	0.2197	0.2261	0.2328	0.2197	0.2261	0.2063	0.2354	0.2199
	ค่าเฉลี่ย	0.2542	0.3692	0.2800	0.2278	0.3811	0.2687	0.2277	0.3870	0.2687

■ เมื่อค่าขีดแบ่งเท่ากับ Mean + S.D.

เมื่อใช้ค่าขีดแบ่งเท่ากับค่าเฉลี่ยบวกด้วยค่าส่วนเบี่ยงเบนมาตรฐานจะทำให้ได้ค่าพรีซิชั่นที่สูงขึ้นขึ้นแต่ก็ทำให้ค่ารีคอลต่ำลงมากด้วยซึ่งทำให้อาจตัดเอกสารที่เกี่ยวข้องออกไปมากจนเกินไปหรือทำให้ค้นคืนออกมาได้น้อยมากๆและอาจจะไม่ตรงกับความต้องการของผู้ใช้ก็ได้

ตารางที่ 4.10 ผลการทดลองของวิธีวัดความคล้ายแบบไดซ์ (Dice) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน

ค่าความคล้าย	ค่านำหนัก	TF			IDF			TF-IDF		
	หัวข้อ	R	P	H	R	P	H	R	P	H
ไดซ์	ก	0.2448	0.7495	0.3468	0.2226	0.7726	0.3352	0.1927	0.8250	0.3025
	ข	0.2586	0.2767	0.2432	0.1945	0.3442	0.2404	0.1820	0.3485	0.2288
	ค	0.6464	0.3290	0.3959	0.2839	0.4092	0.3191	0.3107	0.4143	0.3376
	ง	0.3021	0.5093	0.3594	0.2823	0.5743	0.3607	0.2146	0.5467	0.2938
	จ	0.4391	0.2788	0.3183	0.3219	0.2673	0.2830	0.2313	0.2376	0.2259
	ค่าเฉลี่ย	0.3782	0.4287	0.3327	0.2610	0.4735	0.3077	0.2263	0.4744	0.2777

ตารางที่ 4.11 ผลการทดลองของวิธีวัดความคล้ายแบบแจ็กการ์ด (Jaccard) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน

ค่าน้ำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
แจ็กการ์ด	หัวข้อ									
	ก	0.2359	0.7582	0.3397	0.2190	0.7941	0.3316	0.1738	0.8211	0.2784
	ข	0.2461	0.2830	0.2381	0.1922	0.3527	0.2407	0.1766	0.3697	0.2301
	ค	0.6304	0.3469	0.4073	0.2750	0.4358	0.3204	0.2821	0.4602	0.3336
	ง	0.2750	0.4990	0.3355	0.2708	0.5738	0.3505	0.1771	0.5311	0.2537
	จ	0.4109	0.2681	0.3038	0.3188	0.2714	0.2843	0.2016	0.2296	0.2089
	ค่าเฉลี่ย	0.3597	0.4310	0.3249	0.2551	0.4855	0.3055	0.2022	0.4823	0.2609

ตารางที่ 4.12 ผลการทดลองของวิธีวัดความคล้ายแบบโคไซน์ (Cosine) ซึ่งใช้ค่าน้ำหนักแบบต่างๆเมื่อค่าขีดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน

ค่าน้ำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
โคไซน์	หัวข้อ									
	ก	0.2492	0.7383	0.3532	0.1927	0.7700	0.2993	0.2137	0.8199	0.3264
	ข	0.2813	0.2889	0.2581	0.1672	0.3383	0.2158	0.1844	0.3446	0.2294
	ค	0.6411	0.3010	0.3647	0.2446	0.3520	0.2687	0.3786	0.4099	0.3704
	ง	0.3156	0.5241	0.3663	0.1865	0.4997	0.2587	0.1823	0.5021	0.2560
	จ	0.4359	0.2721	0.3098	0.2828	0.2623	0.2618	0.2547	0.2660	0.2473
	ค่าเฉลี่ย	0.3846	0.4249	0.3304	0.4445	0.2148	0.2628	0.2427	0.4685	0.2859

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ผลการทดลองของวิธีวัดความคล้ายแบบโอเวอร์แลป (Overlap) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
โอเวอร์แลป	หัวข้อ									
	ก	0.2165	0.7809	0.3240	0.2185	0.7731	0.3306	0.1766	0.8245	0.2837
	ข	0.2375	0.3245	0.2516	0.1695	0.3156	0.2122	0.1633	0.3729	0.2176
	ค	0.4607	0.3263	0.3312	0.2589	0.4082	0.2966	0.3125	0.4418	0.3505
	ง	0.2010	0.4980	0.2672	0.2219	0.5272	0.2947	0.1656	0.5517	0.2403
	จ	0.2469	0.1948	0.2061	0.3031	0.2543	0.2646	0.1578	0.1955	0.1696
	ค่าเฉลี่ย	0.2725	0.4249	0.2760	0.4557	0.2334	0.2797	0.1952	0.4773	0.2523

ตารางที่ 4.14 ผลการทดลองของวิธีวัดความคล้ายแบบอะซิมเมตริก (Asymmetric) ซึ่งใช้ค่านำหนักแบบต่างๆเมื่อค่าจัดแบ่งเท่ากับค่าเฉลี่ยบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน

ค่านำหนัก ค่าความคล้าย		TF			IDF			TF-IDF		
		R	P	H	R	P	H	R	P	H
อะซิมเมตริก	หัวข้อ									
	ก	0.1048	0.5466	0.1582	0.0923	0.5499	0.1507	0.0968	0.5834	0.1580
	ข	0.1336	0.3444	0.1720	0.0938	0.3782	0.1436	0.1070	0.4020	0.1588
	ค	0.2464	0.2331	0.2017	0.1339	0.3086	0.1722	0.1786	0.3315	0.2168
	ง	0.2490	0.4681	0.2993	0.1635	0.4720	0.2272	0.1771	0.4893	0.2474
	จ	0.2047	0.2626	0.2134	0.1656	0.2948	0.2043	0.1609	0.2931	0.2013
	ค่าเฉลี่ย	0.1877	0.3709	0.2089	0.1298	0.4007	0.1796	0.1441	0.4199	0.1965

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.3 ผลการวัดประสิทธิภาพการค้นคืนของการค้นคืนโดยใช้ทฤษฎีการให้ความสำคัญสำหรับตัววัดประสิทธิภาพการค้นคืน

##### 4.3.1 กรณีที่ ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ายทั้งหมดลบกับค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ของค่าเฉลี่ยของค่าความคล้ายทั้งหมด

ในกรณีนี้จะเห็นได้ว่า ค่าเอฟเมเชอร์จะเท่ากันหมด (แสดงในตารางที่ 4.15) เพราะเนื่องจากได้ค่าขีดแบ่งที่ติดลบจึงสมมติเป็นศูนย์ (กล่าวถึงในตารางที่ 4.1) จึงทำให้ไม่สามารถจะกำจัดเอกสารที่มีความเกี่ยวข้องน้อยๆออกไปได้ซึ่งทำให้ค่าเอฟเมเชอร์เท่ากันทุกกรณี โดยการทดลองได้มีการให้ความสำคัญกับค่าพรีซิชั่นดังนี้

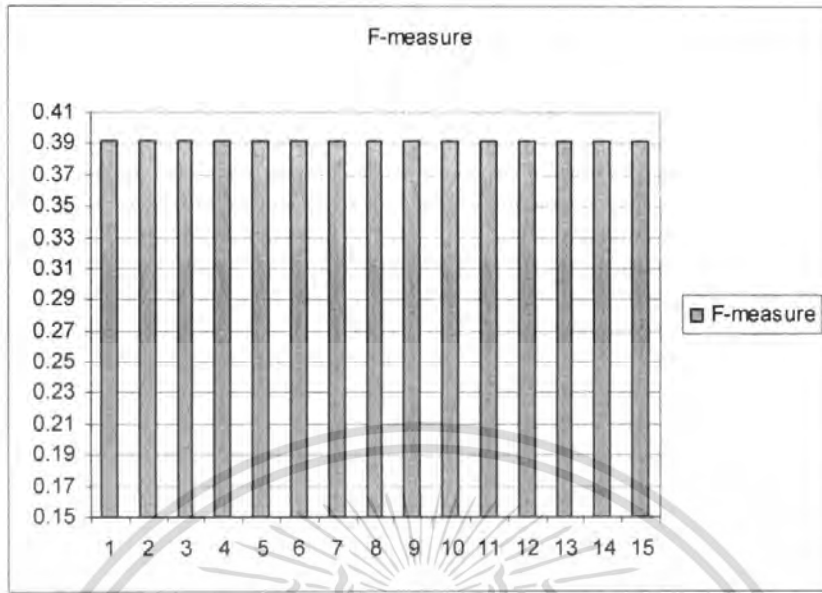
##### 4.3.1.1 กรณีให้ความสำคัญกับพรีซิชั่นอย่างเดียว

จากที่ได้กล่าวไว้ข้างต้นจะได้ผลของของการค้นคืนที่เท่ากันในทุกวิธีดังแสดงในตารางที่ 4.15

ตารางที่ 4.15 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ  $\alpha$  เท่ากับ 0 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.2)

ลำดับ	ค่านำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	TF-IDF	Asymmetric	0.391546
2	TF-IDF	Overlap	0.391546
3	IDF	Jaccard	0.391546
4	TF	Dice	0.391546
5	TF-IDF	Cosine	0.391546
6	IDF	Asymmetric	0.391546
7	IDF	Overlap	0.391546
8	TF	Jaccard	0.391546
9	TF-IDF	Dice	0.391546
10	IDF	Cosine	0.391546
11	TF	Asymmetric	0.391546
12	TF	Overlap	0.391546
13	TF-IDF	Jaccard	0.391546
14	IDF	Dice	0.391546
15	TF	Cosine	0.391546

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการเรียนการสอนเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของลิขสิทธิ์  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.1 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ  $\alpha$  เท่ากับ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.3.1.2 กรณีให้ความสำคัญกับพรีซิชันมากกว่ารีคอลล

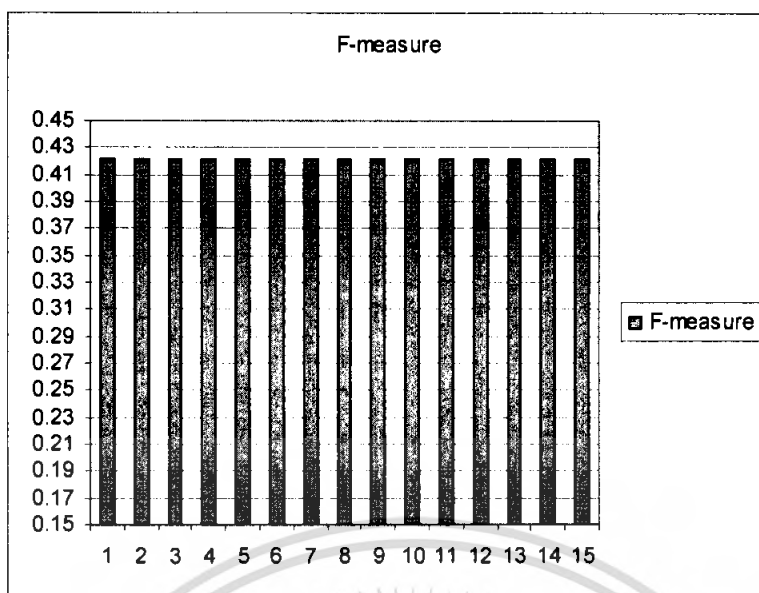
กรณีนี้ค่าเอฟเมเชอร์จะสูงขึ้นกว่า หัวข้อ 4.3.1.1 โดยแต่ละวิธีก็ยังคงให้ค่าเอฟเมเชอร์ที่เท่ากันอยู่ดีเพราะว่าไม่ได้ตัดเอกสารที่มีความเกี่ยวข้องน้อยๆ ออกไปจากตารางที่ 4.17 จะเห็นได้ว่าลำดับการแสดงผลไม่มีความสำคัญ เพราะค่าเอฟเมเชอร์เท่ากันหมดทุกกรณี

#### ตารางที่ 4.16 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่ง

เท่ากับค่าเฉลี่ยของทุกค่าค้นลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชันมากกว่ารีคอลล เมื่อ  $\alpha$  เท่ากับ 0.5 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.2)

ลำดับ	ค่าน้ำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	TF-IDF	Asymmetric	0.422184
2	TF-IDF	Overlap	0.422184
3	IDF	Jaccard	0.422184
4	TF	Dice	0.422184
5	TF-IDF	Cosine	0.422184
6	IDF	Asymmetric	0.422184
7	IDF	Overlap	0.422184
8	TF	Jaccard	0.422184
9	TF-IDF	Dice	0.422184
10	IDF	Cosine	0.422184
11	TF	Asymmetric	0.422184
12	TF	Overlap	0.422184
13	TF-IDF	Jaccard	0.422184
14	IDF	Dice	0.422184
15	TF	Cosine	0.422184

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.2 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นคืนกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอล เมื่อ  $\alpha$  เท่ากับ 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.3.1.3 กรณีให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอลล

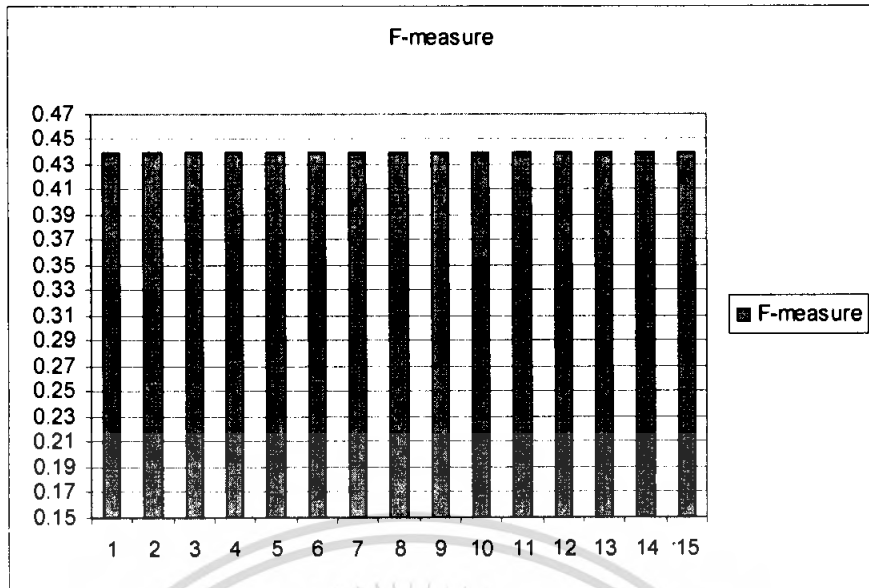
กรณีนี้ค่าเอฟเมเชอร์จะสูงขึ้นกว่า หัวข้อ 4.3.1.2 เล็กน้อยโดยแต่ละวิธีก็ ชังให้ค่าเอฟเมเชอร์ที่เท่ากันอยู่ดีเพราะว่าไม่ได้ตัดเอกสารที่มีความเกี่ยวข้อง น้อยๆออกไป จากตารางที่ 4.17 จะเห็นได้ว่าลำดับการแสดงผลไม่ม ี ความสำคัญเพราะค่าเอฟเมเชอร์เท่ากันหมดทุกกรณี

ตารางที่ 4.17 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่ง

เท่ากับค่าเฉลี่ยของทุกค่าค้นลกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับ พรีซิชั่นเท่ากับรีคอลล เมื่อ  $\alpha$  เท่ากับ 1 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.2)

ลำดับ	ค่านำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	TF-IDF	Asymmetric	0.439375
2	TF-IDF	Overlap	0.439375
3	IDF	Jaccard	0.439375
4	TF	Dice	0.439375
5	TF-IDF	Cosine	0.439375
6	IDF	Asymmetric	0.439375
7	IDF	Overlap	0.439375
8	TF	Jaccard	0.439375
9	TF-IDF	Dice	0.439375
10	IDF	Cosine	0.439375
11	TF	Asymmetric	0.439375
12	TF	Overlap	0.439375
13	TF-IDF	Jaccard	0.439375
14	IDF	Dice	0.439375
15	TF	Cosine	0.439375

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.3 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นลบกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคัล เมื่อ  $\alpha$  เท่ากับ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.3.2 กรณีที่ ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ายทั้งหมด

กรณีนี้ใช้ค่าเฉลี่ยเป็นค่าจัดแบ่งซึ่งทำให้สามารถกำจัดเอกสารที่มีความเกี่ยวข้องกับน้อยๆออกไปได้ทำให้แต่ละวิธีมีความแตกต่างกัน โดยจากการทดลองได้มีการให้ความสำคัญกับค่าพรีซิชั่นดังนี้

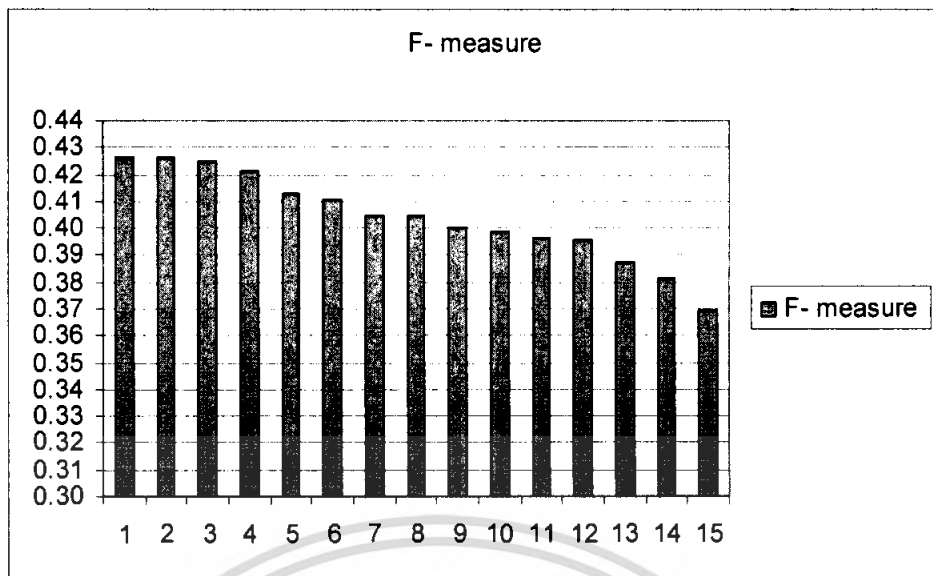
##### 4.3.2.1 กรณีให้ความสำคัญกับพรีซิชั่นอย่างเดียว

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

ตารางที่ 4.18 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นและให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ  $\alpha$  เท่ากับ 0 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.3)

ลำดับ	ค่าน้ำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	TF-IDF	Overlap	0.4265
2	TF-IDF	Jaccard	0.4262
3	TF-IDF	Cosine	0.4247
4	TF-IDF	Dice	0.4209
5	IDF	Cosine	0.4131
6	IDF	Overlap	0.4106
7	IDF	Jaccard	0.405
8	IDF	Dice	0.4043
9	TF	Overlap	0.3998
10	TF	Jaccard	0.3983
11	TF	Dice	0.3963
12	TF	Cosine	0.3959
13	TF-IDF	Asymmetric	0.387
14	IDF	Asymmetric	0.3811
15	TF	Asymmetric	0.3692

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.4 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นและให้ความสำคัญกับพรีซิชั่นอย่างเดียว เมื่อ  $\alpha$  เท่ากับ 0



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

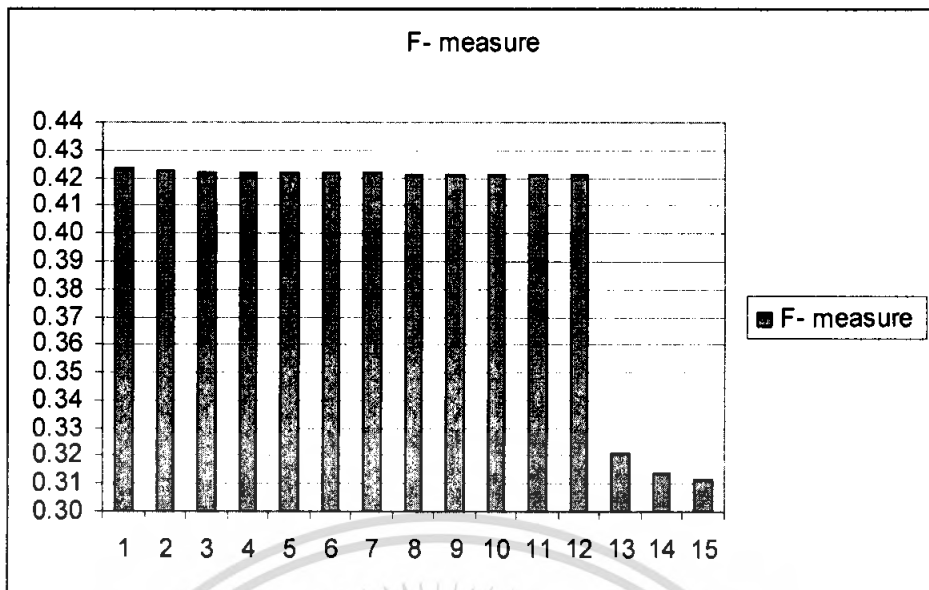
#### 4.3.2.2 กรณีให้ความสำคัญกับพรีซิชันมากกว่ารีคอล

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

ตารางที่ 4.19 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นและให้ความสำคัญกับพรีซิชันมากกว่ารีคอล เมื่อ  $\alpha$  เท่ากับ 0.5 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.3)

ลำดับ	ค่านำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	IDF	Overlap	0.4234
2	TF-IDF	Jaccard	0.4227
3	IDF	Dice	0.4221
4	IDF	Jaccard	0.4219
5	TF	Jaccard	0.4218
6	TF	Overlap	0.4216
7	TF-IDF	Dice	0.4215
8	TF-IDF	Overlap	0.4211
9	TF	Dice	0.4211
10	TF	Cosine	0.4210
11	IDF	Cosine	0.4210
12	TF-IDF	Cosine	0.4209
13	TF	Asymmetric	0.3208
14	TF-IDF	Asymmetric	0.3138
15	IDF	Asymmetric	0.3113

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.5 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกคำค้นและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคัล เมื่อ  $\alpha$  เท่ากับ 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

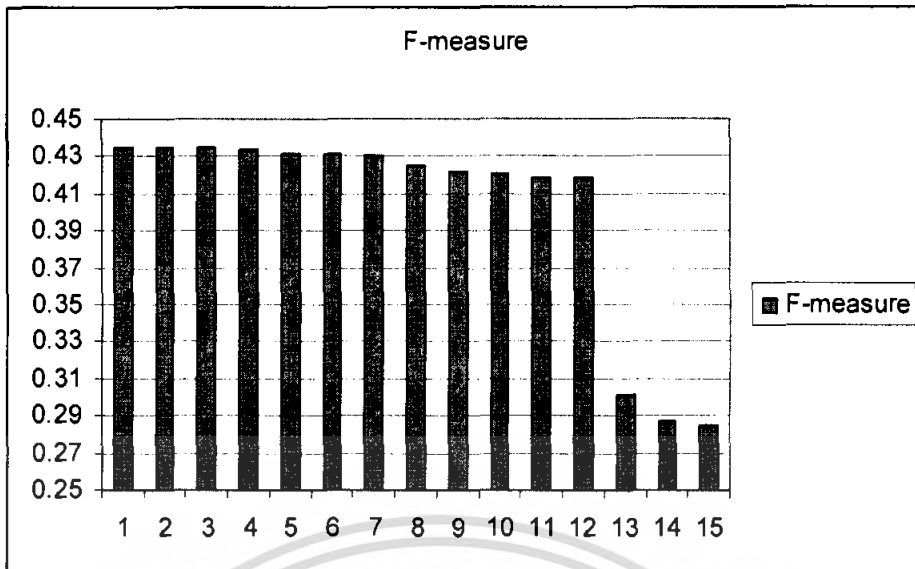
### 4.3.2.3 กรณีให้ความสำคัญกับพรีซิชันเท่ากับรีคอลล

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

ตารางที่ 4.20 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นและให้ความสำคัญกับพรีซิชันเท่ากับรีคอลล เมื่อ  $\alpha$  เท่ากับ 1 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.3)

ลำดับ	ค่าน้ำหนัก	ค่าความคล้าย	เอฟเมเจอร์
1	TF	Cosine	0.4348
2	TF	Dice	0.4347
3	TF	Jaccard	0.4347
4	TF	Overlap	0.4334
5	IDF	Dice	0.4315
6	IDF	Jaccard	0.4310
7	IDF	Overlap	0.4301
8	IDF	Cosine	0.4250
9	TF-IDF	Dice	0.4219
10	TF-IDF	Jaccard	0.4209
11	TF-IDF	Cosine	0.4190
12	TF-IDF	Overlap	0.4185
13	TF	Asymmetric	0.3011
14	TF-IDF	Asymmetric	0.2867
15	IDF	Asymmetric	0.2852

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.6 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกคำค้นและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคัล เมื่อ  $\alpha$  เท่ากับ 1

#### 4.3.3 กรณีที่ ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของค่าความคล้ายทั้งหมดบวกกับค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ของค่าเฉลี่ยของค่าความคล้ายทั้งหมด

ในกรณีนี้ได้มีการนำเอาค่าเฉลี่ยรวมกับค่าส่วนเบี่ยงเบนมาตรฐานของค่าความคล้ายทั้งหมดทำให้ได้ค่าขีดแบ่งที่สูงคั้งนั้นจึงทำให้สามารถคัดเอกสารที่มีความเกี่ยวข้องน้อยๆออกไปได้มากขึ้นในขณะที่เดียวกันก็ทำให้ค้นคืนเอกสารที่เกี่ยวข้องได้น้อยลง

การทดลองได้มีการให้ความสำคัญกับค่าพรีซิชั่นดังนี้

##### 4.3.3.1 กรณีให้ความสำคัญกับพรีซิชั่นอย่างเดียว

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

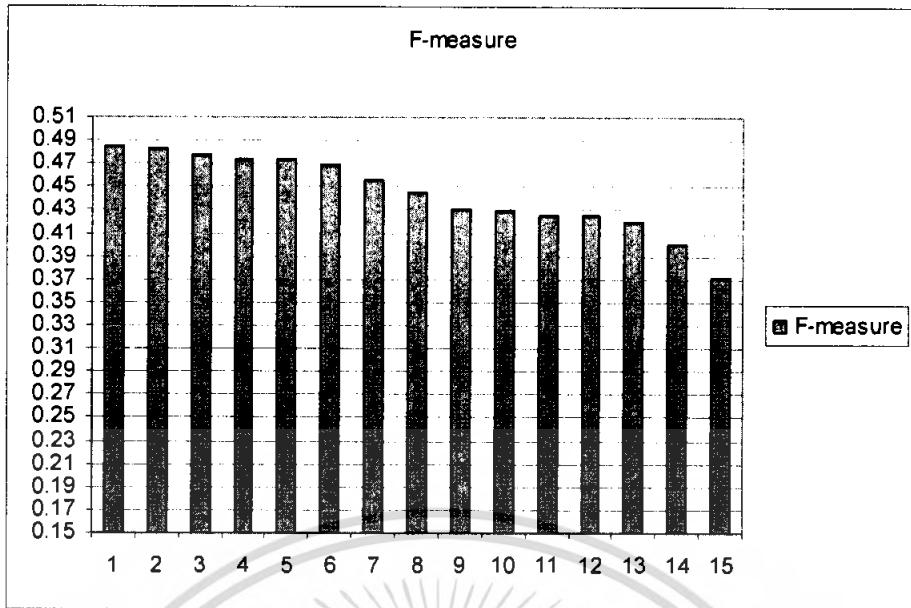
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.21 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชันอย่างเดียว เมื่อ  $\alpha$  เท่ากับ 0 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.4)

ลำดับ	ค่าน้ำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	IDF	Jaccard	0.485532
2	TF-IDF	Jaccard	0.482338
3	TF-IDF	Overlap	0.477269
4	TF-IDF	Dice	0.474425
5	IDF	Dice	0.473517
6	TF-IDF	Cosine	0.468483
7	IDF	Overlap	0.455676
8	IDF	Cosine	0.444474
9	TF	Jaccard	0.431038
10	TF	Dice	0.428684
11	TF	Overlap	0.424889
12	TF	Cosine	0.424879
13	TF-IDF	Asymmetric	0.419864
14	IDF	Asymmetric	0.400696
15	TF	Asymmetric	0.370935

จากตารางที่ 4.21 จะเห็นว่าวิธีการให้ค่าน้ำหนักแบบ IDF และวิธีให้ค่าความคล้ายแบบ แจ็คคาร์ดให้ค่าเอฟเมเชอร์ที่สูงที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.7 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าคำนวณกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นอย่างเดียวน เมื่อ  $\alpha$  เท่ากับ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

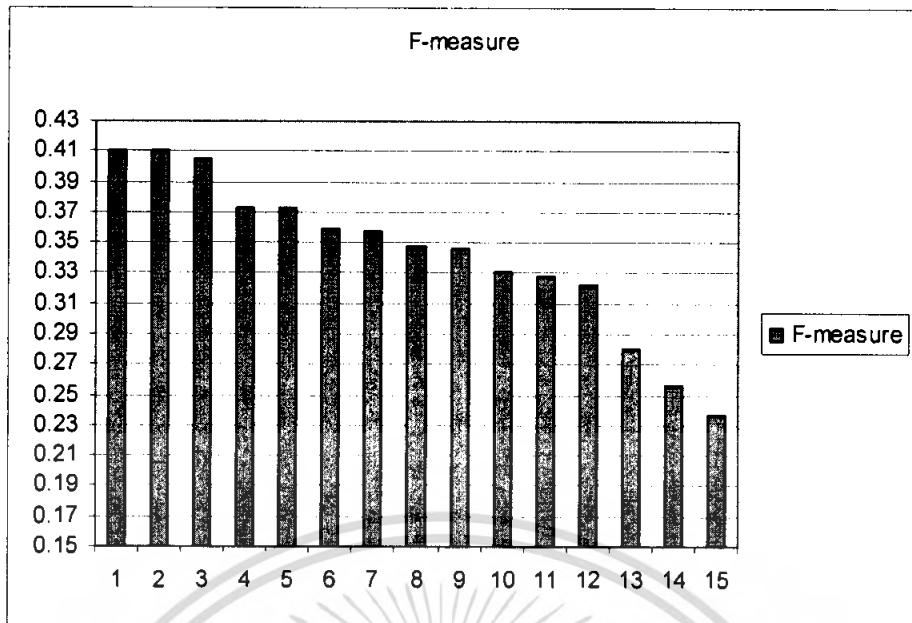
#### 4.3.3.2 กรณีให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอลล

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

ตารางที่ 4.22 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าจัดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคอลล เมื่อ  $\alpha$  เท่ากับ 0.5 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.4)

ลำดับ	ค่าน้ำหนัก	ค่าความคล้าย	เอฟเมเชอร์
1	TF	Cosine	0.410553
2	TF	Dice	0.410417
3	TF	Jaccard	0.404291
4	IDF	Jaccard	0.373202
5	IDF	Dice	0.372462
6	TF	Overlap	0.358149
7	TF-IDF	Cosine	0.357613
8	TF-IDF	Dice	0.347414
9	IDF	Overlap	0.34661
10	TF-IDF	Jaccard	0.329983
11	IDF	Cosine	0.327658
12	TF-IDF	Overlap	0.322084
13	TF	Asymmetric	0.279867
14	TF-IDF	Asymmetric	0.256321
15	IDF	Asymmetric	0.236345

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.8 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นมากกว่ารีคัลด์ เมื่อ  $\alpha$  เท่ากับ 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

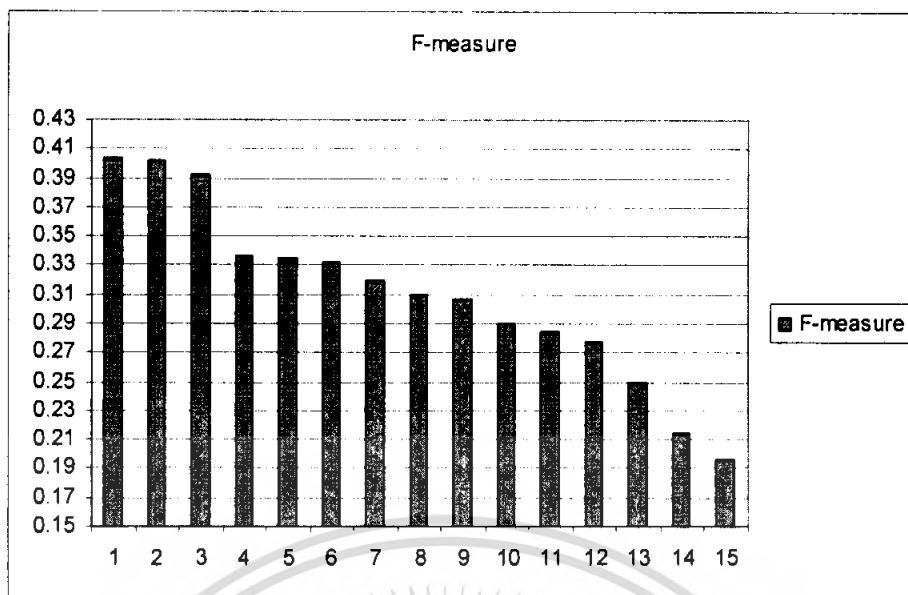
### 4.3.3.3 กรณีให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอลล

จะได้ผลของการเรียงลำดับการแสดงผลของกรณีการค้นคืนที่มีประสิทธิภาพจากมากไปน้อยดังนี้

ตารางที่ 4.23 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอลล เมื่อ  $\alpha$  เท่ากับ 1 (ซึ่งได้กล่าวถึงแล้วใน 2.1.3 และ สมการที่ 3.4)

ลำดับ	ค่านำหนัก	ค่าความคล้าย	เอฟเมเจอร์
1	TF	Cosine	0.403746
2	TF	Dice	0.401854
3	TF	Jaccard	0.392124
4	IDF	Dice	0.336549
5	IDF	Jaccard	0.334507
6	TF	Overlap	0.332069
7	TF-IDF	Cosine	0.319775
8	IDF	Overlap	0.309563
9	TF-IDF	Dice	0.3064
10	IDF	Cosine	0.289602
11	TF-IDF	Jaccard	0.284976
12	TF-IDF	Overlap	0.277043
13	TF	Asymmetric	0.249268
14	TF-IDF	Asymmetric	0.214539
15	IDF	Asymmetric	0.196124

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.9 การเรียงลำดับของประสิทธิภาพการค้นคืนจากมากไปน้อยโดยที่ค่าขีดแบ่งเท่ากับค่าเฉลี่ยของทุกค่าค้นบวกกับค่าส่วนเบี่ยงเบนมาตรฐานและให้ความสำคัญกับพรีซิชั่นเท่ากับรีคอลล เมื่อ  $\alpha$  เท่ากับ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### สรุปผลการวิจัย การอภิปราย และข้อเสนอแนะ

ปัญหาพิเศษนี้มีจุดประสงค์เพื่อศึกษาการเปรียบเทียบวิธีการถ่วงน้ำหนักและวิธีการคำนวณค่าความคล้ายว่าวิธีใดเหมาะสมในการค้นคืนตามโครงสร้างเอกสารการบรรยายยูสเคส ซึ่งรายละเอียดการสรุปผลการวิจัย การอภิปรายผลการวิจัยและข้อเสนอแนะมีดังต่อไปนี้

#### 5.1 สรุปผลการวิจัย

เนื่องจากความต้องการลดระยะเวลาและงบประมาณของการพัฒนาซอฟต์แวร์ ซึ่งก็คือต้องสามารถค้นคืนเอกสารการบรรยายยูสเคสได้ตรงกับความต้องการมากที่สุดเพื่อนำกลับมาใช้ใหม่ในปัจจุบันมีวิธีการค้นคืนข้อมูลมากมาย ซึ่งวิธีการต่าง ๆ นั้นได้ถูกพัฒนาให้ดีขึ้นมาเป็นลำดับซึ่งในบทนี้จะสรุปได้ว่าวิธีการให้น้ำหนักและการคำนวณค่าความคล้ายแบบใดเป็นวิธีที่มีประสิทธิภาพสำหรับการค้นคืนเอกสารการบรรยายยูสเคส

จากการศึกษาขั้นตอนและวิธีการ ในการหาวิธีการถ่วงน้ำหนักค่าและการวิเคราะห์ความคล้ายที่เหมาะสมกับการค้นคืนตาม โครงสร้างเอกสารการบรรยายยูสเคสนั้นต้องมีกระบวนการต่าง ๆ มากมายซึ่งขอสรุปโดยง่ายกล่าวคือ นำข้อมูลเอกสารการบรรยายยูสเคสผ่านกระบวนการตัดคำและสร้างดัชนีจากนั้นผู้ทดลองก็จะทำการค้นคืนและระบบก็จะทำการคำนวณค่าความคล้ายตามแต่ละวิธีที่เลือก โดยจะนำเฉพาะค่าความคล้ายที่มากกว่าค่าขีดแบ่งมาคำนวณว่าวิธีใดเหมาะสม

เนื่องจากการสรุปผลว่าวิธีการถ่วงน้ำหนักและการวิเคราะห์ความคล้ายแบบใดว่าเหมาะสม นั้น เราจะต้องวิเคราะห์จากค่าสถิติของการประเมินผลการค้นคืนทั้งค่า ฟรیشันและรีคอล โดยที่การสรุปผลนั้นขึ้นอยู่กับการให้ความสำคัญกับฟรیشันและรีคอลที่แตกต่างกันของกรณีค่าขีดแบ่งในแต่ละกรณี ซึ่งจะทำการศึกษาว่าค่าขีดแบ่งกรณีใดมีประสิทธิภาพเพื่อใช้วิเคราะห์ถึงความสามารถในการค้นคืนของวิธีต่างๆ ที่ใช้เป็นเกณฑ์ในการสรุปผลว่าวิธีการให้น้ำหนักแบบใดและค่าความคล้ายแบบใดที่มีประสิทธิภาพในการใช้ค้นคืนเอกสารการบรรยายยูสเคส ซึ่งในส่วนนี้เราจะใช้วิธีการวัดประสิทธิภาพโดยใช้ตัววัดประสิทธิภาพ เอฟเมเชอร์ (F-measure) ซึ่งทำให้สามารถกำหนดค่าความสำคัญดังกล่าวได้ซึ่งมีรายละเอียดดังต่อไปนี้

กำหนดให้  $N^*$  คือไม่สามารถระบุว่ามีวิธีใดเหมาะสมได้

เพราะวิธีคำนวณค่าขีดแบ่ง (Mean - S.D.) ไม่สามารถกำจัดเอกสารที่มีความเกี่ยวข้องน้อยๆ ออกไปได้จึงทำให้ค้นคืนออกมาได้ทั้งหมด กล่าวคือทุกวิธีให้ผลเท่ากันหมด

ตารางที่ 5.1 การสรุปผลการทดลอง

วิธีการคำนวณค่าจัดแบ่ง	การให้ความสำคัญของตัววัดประสิทธิภาพ ( $\alpha$ )	วิธีการค้นคืนที่มีเหมาะสม	
		วิธีถ่วงน้ำหนัก	วิธีการคำนวณความคล้าย
Mean – S.D.	$\alpha = 0$	N*	N*
	$\alpha = 0.5$	N*	N*
	$\alpha = 1$	N*	N*
Mean	$\alpha = 0$	TF-IDF	Overlap
	$\alpha = 0.5$	IDF	Overlap
	$\alpha = 1$	TF	Cosine
Mean + S.D.	$\alpha = 0$	IDF	Jaccard
	$\alpha = 0.5$	TF	Cosine
	$\alpha = 1$	TF	Cosine

วิธีการคำนวณค่าจัดแบ่งที่เหมาะสมนั้นคือ Mean เพราะจากการทดลองในบทที่ 4 ค่าจัดแบ่งที่เท่ากับค่าเฉลี่ยนี้จะให้ค่าเฉลี่ยรีคอลที่ไม่น่าจะเกินไป และให้เฉลี่ยค่าพรีซิชันที่สูง ทำให้ได้เอกสารที่มีตรงกับความต้องการและมีปริมาณที่พอเหมาะให้ผู้ใช้เลือกไปใช้ได้ แต่ถ้าใช้วิธีการคำนวณค่าจัดแบ่งวิธี Mean + S.D. จะทำให้เอกสารที่มีความเกี่ยวข้องบางเอกสารถูกกำจัดทิ้งไปได้ จึงทำให้ค่าพรีซิชันสูงมาก แต่สังเกตว่าแม้ค่าเฉลี่ยพรีซิชันจะสูงแต่ค่าเฉลี่ยรีคอลก็จะลดลงมากเช่นกัน

โดยทั่วไปจะพิจารณาการให้ความสำคัญของตัววัดประสิทธิภาพที่  $\alpha = 0$  เพราะเป็นการสนใจในค่าพรีซิชัน ซึ่งก็คือความแม่นยำในการค้นคืนเป็นหลัก ดังนั้นวิธีการที่เหมาะสมในการค้นคืนเอกสารการบรรยายขุสเทศก็คือวิธีถ่วงน้ำหนักแบบ TF-IDF และวิธีการคำนวณความคล้ายแบบโอเวอร์แลป (Overlap)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 5.2 การอภิปรายผลการวิจัย

จากผลการทดลองในบทที่ 4 เราใช้วิธีหาค่าขีดแบ่งหลายวิธีซึ่งค่าขีดแบ่งที่เหมาะสมที่สุดในการทดลองนี้คือ ค่าเฉลี่ย เพราะให้ค่ารีคอลที่สูง และค่าพรีซิชั่นที่สูง กว่ากรณีอื่น ซึ่งกรณีอื่นได้ค่ารีคอลและพรีซิชั่นที่สูงหรือต่ำเกินไป จากการวิเคราะห์ผลการทดลองในทุกๆกรณีจะเห็นได้ว่าวิธีการคำนวณค่าความคล้ายแบบอะซิมเมตริก (สมการที่ 2.8) ให้ค่าเอฟเมเชอร์ต่ำสุดในทุกๆกรณี

$$\text{Similarity}(E_{mi}, E_{mj}) = \frac{\sum_{k=1}^l \min(\text{Term}_{mik}, \text{Term}_{mjk})}{\sum_{k=1}^l \text{Term}_{mik}}$$

ทั้งนี้เนื่องจาก สูตรการคำนวณค่าความคล้ายนี้ ตัวเศษเป็นค่าผลรวมของเทอมที่น้อยที่สุด  $\min(\text{Term}_{mik}, \text{Term}_{mjk})$  ขณะที่ตัวส่วนก็ยังเป็นค่า ผลรวมของ  $\text{Term}_{mik}$  ซึ่งทำให้ ได้ตัวเศษที่น้อยกว่าวิธีอื่นและได้ตัวส่วนที่มากกว่านั้นค่าความคล้ายของวิธีนี้จึงน้อยกว่าวิธีที่เลือกมาทั้งหมด

การทดลองนี้มีหลายปัจจัยที่ใช้ในการค้นคืนทั้งจากวิธีการต่างๆที่เลือกมาและยังมีส่วนที่ผู้ใช้สามารถให้ค่าน้ำหนักกับแต่ละองค์ประกอบของเอกสารการบรรยายยูสเคส ดังนั้นผลที่ได้ อาจจะเหมาะสมในเฉพาะการศึกษาทดลองนี้

## 5.3 ข้อเสนอแนะ

การทดลองนี้เรายังสามารถเพิ่มประสิทธิผลในการค้นคืนได้อีกโดย

1) สามารถทำการตัดคำบางคำออกจากการทดลองอีก ยกตัวอย่างเช่น ตัวอักษรและต่อด้วยคำ กล่าวคือ หัวข้อของเอกสารการบรรยายยูสเคสบางองค์ประกอบมีหัวข้อ เช่น 1.2a แต่การตัดคำเราตัดตามช่องว่างของคำทำให้ตัวอักษร a ไม่ถูกตัดออกไปยังคงอยู่ในการสร้างดัชนีซึ่งทำให้ค้นหาได้ช้าลงและเปลืองเนื้อที่ในการจัดเก็บ อนึ่งยังทำให้การค้นคืนมีความแม่นยำน้อยลงอีกด้วย

2) สามารถปรับปรุงวิธีการในการตัดคำที่มีความถี่สูงๆและไม่สามารถระบุเอกสารใดๆได้ (Stopword) ได้อีกโดยศึกษาเชิงความหมายกับคำที่พบบ่อยหรือค้นหาจากรายงานการวิจัยที่เกี่ยวข้องเพื่อมาปรับปรุงประสิทธิภาพของการค้นคืนได้

## รายการอ้างอิง

- [1] A. Udomchaiporn ,N. Prompoon ,P. Kanongchaiyos, “*Software Requirements Retrieval Using Use Case Terms and Structure Similarity Computation*”, IEEE ,2006.
- [2] G. Salton and M. J. McGill., “*Retrieval Refiements, Introduction to Modern Information Retrieval*” , page 199-250.
- [3] B. Ribeiro-Neto and R. Baeza-Yates, “*Modern information retrieval, Addison-Wesley*”, 1999, P 433-436.
- [4] Frakes, William B. and Baeza-Yates, Ricardo, Information Retrieval: Data Structures and Algorithms, Chapter 7, “*Lexical Analysis and Stoplists*”, by Christopher Fox Prentice-Hall, 1992. (Revised Version - 1998)
- [5] B. Ribeiro-Neto and R. Baeza-Yates, “*Modern information retrieval, Addison-Wesley*”, 1999, P 433-436.
- [6] ชูชาติ หุฒิชัยศักดิ์, “*การพัฒนาบบค้นคืนข้อมูลโดยลูชัน*” ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (เนคเทค), 2006.

## ภาคผนวก ก.

### รายละเอียดระบบของเอกสารการบรรยายยุสเคสที่เลือกใช้ในการวิจัย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนนี้จะอธิบายรายละเอียดระบบของเอกสารการบรรยายยูสเคสที่ใช้ในการวิจัย ซึ่งประกอบด้วย 16 ระบบจากหลายๆระบบ ทุกระบบถูกเขียนโดยใช้ภาษาอังกฤษโดยมีรายละเอียดซึ่งจะอธิบายดังต่อไปนี้

ตารางที่ ก.1 ระบบของเอกสารการบรรยายยูสเคสที่เลือกมา

ระบบ	ชื่อเอกสารการบรรยายยูสเคส
1	ICT blocking and filtering system
2	Departmental information system for curriculum and course offerings management
3	Resource management system
4	Automatic restaurant lookup system
5	Library management system
6	Mobile phone selling on the internet system
7	Payroll system
8	Personal stock manager on PDA system
9	Basic product trading system
10	Mobile phone service paying system
11	MLM computerize system
12	Content management and platform upgrade for the official website of Nation Channel
13	Purchase, producing, and export system
14	Management information system for insurance company
15	E-learning system
16	Management information system for an academic

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ตารางที่ ก.2 ICT blocking and filtering system

ชื่อระบบ : ICT blocking and filtering system	
รายละเอียดระบบ : ระบบนี้ทำการตรวจจับและป้องกันสิ่งที่ไม่เหมาะสมบนอินเทอร์เน็ต และคอยตรวจคนเล่นเกมที่มีอายุต่ำกว่า 18 ปี ระบบนี้ช่วยผู้ปกครอง, ครู หรือ ตัวตรวจจับไซเบอร์ เพื่อลดความเสี่ยง จากอันตราย ข้อมูลที่ไม่เหมาะสม ซึ่งเด็ก ๆ อาจจะ ได้รับ	
หมายเลข	ชื่อเอกสารการบรรยายยูสเกส
0101	Register for a member
0102	Edit members information
0104	Cancel member status
0105	Request report about members
0106	Manage information for users
0107	Provide other information
0108	Display information
0109	Create token
0110	Create log file
0111	Verify log file
0112	Add game online users information
0113	Manage rules
0114	Receive log file from a game online system
0115	View monitoring reports
0116	Pre-Screening
0117	Manage a bad website
0118	Manage a good website
0119	Manage suspended keywords
0120	View checking blocked information reports

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ICT blocking and filtering system (ต่อ)

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0121	Notify a trace
0122	Initialize sending SMS system
0123	Notify SMS for emergency case
0124	Check notifying a trace
0125	View a notified trace report
0126	Define a system authority for a member
0127	View a report for defining authority for a member

ตารางที่ ก.3 Departmental information system for curriculum and course offerings management

ชื่อระบบ : Departmental information system for curriculum and course offerings management system	
รายละเอียดระบบ : ระบบนี้จัดการข้อมูลหลักสูตรในสถาบันการศึกษา	
Use case ID	Use case name
0201	Save course details
0202	Save course operation status
0203	Save course operation report
0204	Save course operation result
0205	View course operation details
0206	View course operation status
0207	View course operation report
0208	View course operation result
0209	Save subject details
0210	Save subject operation status
0211	Store subject report
0212	Save subject operation result
0213	View subject operation details

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.3 Departmental information system for curriculum and course offerings management (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0214	View subject operation status
0215	View subject operation result
0216	View subject operation report
0217	Save course information
0218	Edit course information
0219	View course information
0220	View course statistic information
0221	Save subject information
0222	Edit subject information
0223	View subject information
0224	View teaching schedule

**ตารางที่ ก.4 Resource management system**

ชื่อระบบ : Resource management system	
รายละเอียดระบบ : ระบบนี้เพื่อจัดการการดำเนินการในองค์กร ระบบนี้แบ่งเป็น 6 ระบบย่อย มีการจัดการลูกค้า , การจัดการการขาย, การจัดการพนักงาน , การจัดการวัตถุดิบ , การจัดการสินค้าคงคลัง และการจัดการคำสั่งซื้อ	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0301	Process new application
0302	Process add new customer
0303	Process customer verification
0304	Process Credit Customer Payment
0305	Process Cash Sales transactions
0306	Process Credit Sales
0307	Process Lay-Bye Sales
0308	Process Sales Returns

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.4 Resource management system (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0309	Capture employees details
0310	Process employee records
0311	Process existing supplier
0312	Process New Supplier
0313	Process Incoming Inventory
0314	Process Self-Manufactured Products
0315	Process Check Inventory Available
0316	Process Inventory Order
0317	Access control
0318	Process Generate Password
0319	Process Order Details
0320	Process Receive Order Details
0321	Process order payment
0322	Process Account Payments
0323	Generate Sales Figures
0324	Process Petty Cash Withdrawals
0325	Process Supplier Payments
0326	User Trail
0327	Supplier Order Trail
0328	Customer Payment Trail
0329	Customer Order Trail
0330	Inventory Trail

**ตารางที่ ก.5 Automatic restaurant lookup system**

<b>ชื่อระบบ :</b> Automatic restaurant lookup system
<b>รายละเอียดระบบ :</b> ระบบนี้ทำเพื่อกันหาร้านอาหารและอาหารที่อร่อย ร้านอาหารสามารถถูกค้นพบอย่างเป็นอิสระบนพื้นที่ต่างๆ ,ชนิดของอาหาร หรือ รสชาติ นอกจากนี้ ระบบนี้สามารถทำงานได้ทั้ง คอมพิวเตอร์บุคคลและ PDA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.5 Automatic restaurant lookup system (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0401	New customer
0402	Manage restaurant
0403	Manage customer
0404	Find a restaurant
0405	View restaurants information
0406	Vote restaurant
0407	Manage map
0408	Post webboard
0409	Manage webboard
0410	View FAQ
0411	Manage FAQ
0412	Log in
0413	Find restaurant on PDA
0414	View restaurant's information on PDA
0415	Vote restaurant on PDA

**ตารางที่ ก.6 Library management system**

<b>ชื่อระบบ :</b> Library management system	
<b>รายละเอียดระบบ :</b> ระบบนี้เพื่อจัดการงานบนห้องสมุดโดยใช้ RFID เทคโนโลยี ระบบนี้เตรียมการจัดการทรัพยากรและการจัดการบริการสำหรับบรรณารักษ์ นักศึกษาหรือผู้ใช้อื่น	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0501	Manage users information
0502	Summarize users report
0503	Manage library media information
0504	Define library media status
0505	Summarize library media
0506	Add members information

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ตารางที่ ก.6 Library management system (ต่อ)

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0507	Manage members information
0508	Summarize members
0509	Borrow library media
0510	Return library media
0511	Pay a fine

### ตารางที่ ก.7 Mobile phone selling on the internet system

ชื่อระบบ : Mobile phone selling on the internet system	
รายละเอียดระบบ : ระบบนี้ทำเพื่อขายโทรศัพท์มือถือบนอินเทอร์เน็ต ระบบนี้สามารถบริการค้นหาข้อมูลโทรศัพท์มือถือ ระบบการขาย ระบบการจ่ายเงิน และให้คำแนะนำเกี่ยวกับการเลือกใช้โทรศัพท์มือถือของลูกค้า	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0601	Search product's information
0602	View product's information
0603	Comment
0604	Compare products
0605	Purchase products
0606	Manage a product list
0607	Check purchase order
0608	Pay money
0609	Use credit card
0610	Transfer money
0611	Confirm transferring money
0612	Verify user's payment

### ตารางที่ ก.8 Payroll system

ชื่อระบบ : Payroll system	
รายละเอียดระบบ : ระบบนี้ใช้จัดตารางการทำงานของคนขายข้าวโพดคั่วในสถานที่ต่างๆและคำนวณเงินเดือนของคนขายข้าวโพดคั่วซึ่งขึ้นอยู่กับเวลาการทำงานจริง	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ทางการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.8 Payroll system (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
0701	Add new selling place
0702	Edit selling places information
0703	Cancel selling place
0704	Edit sale staff information
0705	Delete sale staff information
0706	View sale staff information
0707	Display work schedule
0708	Save working time
0709	Calculate staff salary
0710	Calculate social assurance fee
0711	Calculate over time fee
0712	Fine for working late
0713	Calculate special fee
0714	Fine for absent work.
0715	Decrease tax
0716	Collect additional money from selling place
0717	Calculate net salary
0718	Display net salary
0719	Create report for executive
0720	Create report for sale staff or seller
0721	View selling place information
0722	Add new sale staff or seller
0723	Select a sale staff or seller
0724	Calculate salary
0725	Manage salary
0727	Save taking money in advance
0728	View taking money in advance information

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.9 Personal stock manager on PDA system**

ชื่อระบบ : Personal stock manager on PDA system	
รายละเอียดระบบ : ระบบนี้ทำเพื่อจัดการข้อมูลเกี่ยวกับจำนวนสินค้าบน PDA ผู้ใช้สามารถดูข้อมูลจำนวนสินค้าแบบเวลาจริงได้ และสามารถแชร์ข้อมูลจำนวนสินค้าโดยใช้ PDA ได้	
หมายเลข	ชื่อเอกสารการบรรยายคุณสมบัติ
0801	Log-In
0802	Verify user's account
0803	Search by Abbreviate
0804	Search stock
0805	View top ten
0806	View most gain
0807	View most lost
0808	View most active
0809	View Portfolio
0810	View transaction
0811	View index
0812	View search results
0813	View other currencies
0814	View stock details
0815	View graph
0816	Buy stock
0817	Sell stock
0818	Verify PIN
0819	Check regulation
0820	Calculate price and commission
0821	Cancel transaction
0822	Check user regulation
0823	Maintain user account
0824	Maintain regulation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ตารางที่ ก.10 Basic product trading system

ชื่อระบบ : Basic product trading system	
รายละเอียดระบบ : ระบบนี้เพื่อจัดการการขายของบริษัทบนอินเทอร์เน็ต ระบบนี้รวมการจัดการลูกค้า การจัดการสินค้า การจัดการการขายและการจัดการใบสั่งสินค้าไว้ด้วย	
หมายเลข	ชื่อเอกสารการบรรยายคุณสมบัติ
0901	Download purchase order
0902	Maintain purchase order
0903	Maintain sale order
0904	Transfer purchase order to sale order
0905	Approve purchase order
0906	Approve price
0907	Approve credit
0908	Approve stock balance
0909	Maintain invoice
0910	Transfer sale order to invoice
0911	Confirm invoice
0912	Update stock balance
0913	Update customer credit
0914	Maintain customer credit
0915	Maintain product price
0916	Maintain product balance

### ตารางที่ ก.11 Mobile phone service paying system

ชื่อระบบ : Mobile phone service paying system	
รายละเอียดระบบ : ระบบนี้เพื่อจัดการบริการการจ่ายเงินโทรศัพท์มือถือของบริษัท ระบบนี้สามารถจัดการข้อมูลลูกค้า, คำนวณค่าธรรมเนียมของลูกค้าในแต่ละแพ็คเกจ และจัดการรายงานข้อมูลที่เกิดขึ้นได้	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.11 Mobile phone service paying system (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1001	Register telephone number
1002	Manage customer and telephone number information
1003	Manage customer information
1004	Manage telephone number information
1005	Manage service information
1006	Cancel service
1007	Manage service package
1008	Save service information
1009	Ask for information
1010	Check service bill
1011	Check remaining money
1012	Create reports
1013	Print paying report
1014	Print paying report by terminal network
1015	Print income summary report

**ตารางที่ ก.12 Basic MLM computerize system**

ชื่อระบบ : Basic MLM computerize system	
รายละเอียดระบบ : ระบบนี้ทำเพื่อจัดการข้อมูลระบบ MLM ระบบนี้สามารถจัดการข้อมูลลูกค้า, จัดการคลังสินค้า, จัดการการขายและคำนวณผลประโยชน์ของพนักงานได้	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1101	Sign in member
1102	Maintain member information
1103	Show child member
1104	Cancel member
1105	Adjust position
1106	Purchase order

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.12 Basic MLM computerize system (ต่อ)

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1107	Receive product
1108	Adjust product quantity
1109	Maintain information
1110	New product information
1111	Change product information
1112	New position information
1113	Change position information
1114	Order product
1115	Create invoice
1116	Save point value
1117	Print invoice
1118	Cancel invoice
1119	Print commission
1120	Calculate commission
1121	Promote position

ตารางที่ ก.13 Content management and platform upgrade for the official website of Nation Channel

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
<p><b>ชื่อระบบ :</b> Content management and platform upgrade for the official website of Nation Channel</p> <p><b>รายละเอียดระบบ :</b> ระบบนี้เพื่อจัดการเนื้อหาข่าวบนเว็บไซต์อย่างอัตโนมัติ เว็บไซต์นี้เตรียมข้อมูลข่าว, บริการไฟล์, บริการรูป, การออกอากาศ โดยหลักแล้วระบบนี้จัดการข้อมูลข่าวและการบริการต่างๆ.</p>	
1201	Login
1202	Manage Photo
1203	Load Photo Dialogue

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.13 Content management and platform upgrade for the official website of Nation Channel (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1204	Edit Schedule Table
1205	Reply Messages
1206	Manage Program
1207	Manage Today News
1208	Link Video Clip
1209	View Today News
1210	Post Photo Captions
1211	Leave message to anchorman
1212	Vote Poll
1213	Read Breaking News
1214	Activate Live

**ตารางที่ ก.14 Purchase, producing, and export system**

ชื่อระบบ : Purchase, producing, and export system	
รายละเอียดระบบ : ระบบนี้เพื่อจัดการการสั่งซื้อวัสดุ, สินค้า, สินค้านำเข้า, และสินค้าส่งออกของบริษัท	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1301	Purchase system
1302	Product system
1303	Warehouse and export system
1304	Check raw material
1305	Make Contract and Order Raw Material
1306	Track inventory and generate report
1307	Configure flow, hire rate and raw material
1308	Create barcode and preparation
1309	Produce
1310	Calculate hire and create report

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.14 Purchase, producing, and export system (ต่อ)**

หมายเลข	ชื่อเอกสารการบรรยายยุทธศาสตร์
1311	Check inventory
1312	Contact port
1313	Contact duty
1314	Export

**ตารางที่ ก.15 Management information system for insurance company**

ชื่อระบบ : Management information system for insurance company	
รายละเอียดระบบ : ระบบนี้สนับสนุนการขายประกันในบริษัทประกัน โดยหลักระบบนี้มีข้อมูลเกี่ยวกับรายละเอียดการทำประกันทั้งส่วนของพนักงานขายและลูกค้า นอกจากนี้ระบบนี้ยังสนับสนุนข้อมูลบนคอมพิวเตอร์บุคคลและเทคโนโลยีไร้สาย	
หมายเลข	ชื่อเอกสารการบรรยายยุทธศาสตร์
1401	Place a claim
1402	View claim status
1403	External approve claim
1404	Maintain company information
1405	View company information
1406	View product information
1407	Calculate an insurance application
1408	Verify underwriting rule
1409	Calculate premium
1410	Validate insurance application
1411	View agent information
1412	View agent holder information
1413	View policy holder information
1414	Authentication

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.16 E-learning system

ชื่อระบบ : E-learning system	
รายละเอียดระบบ : ระบบนี้ทำเพื่อสนับสนุนระบบ E-learning บนเว็บไซต์ นักเรียนและผู้ใช้ทั่วไปสามารถลงทะเบียนและศึกษาได้ผ่านเว็บไซต์ ระบบนี้สามารถจัดการหลักสูตร ข้อมูลนักเรียน, อาจารย์, ข้อมูลการสอบและการประเมินการเรียน	
หมายเลข	ชื่อเอกสารการบรรยายยูสเคส
1501	Register a new user for E-Learning system
1502	Generate password
1503	Log in to E-Learning system
1504	Change password
1505	Announce course information
1506	Upload materials
1507	Create examinations and examination solutions
1508	Post answer
1509	View course detail
1510	Register course
1511	Verify condition
1512	Study registered course
1513	Take examination and get result of the examination
1514	Record time
1515	View history's study information
1516	Post question
1517	View study result
1518	Follow study result
1519	View study result
1520	Number of learner or student
1521	Maintain employee information

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.17 Management information system for an academic**

ชื่อระบบ : Management information system for an academic	
รายละเอียดระบบ : ระบบนี้ทำเพื่อจัดการหลักสูตรในบางสถานศึกษา ระบบนี้ถูกแบ่งเป็น 3 ส่วนหลักๆคือ นักศึกษา คณะ และผู้บริหาร ระบบนี้สามารถจัดการข้อมูลนักศึกษา , ข้อมูลหลักสูตร, ตารางเรียนตารางสอนและข้อมูลรายวิชา	
หมายเลข	ชื่อเอกสารการบรรยายคุณสมบัติ
1601	Log-In
1602	Student Adjust Schedule
1603	Student View All Courses
1604	Student View Record
1605	Student View Schedule
1606	Faculty View Courses
1607	Faculty View Enrollment
1608	Faculty View Student Records
1609	Add another admin
1610	Add another building
1611	Add another class
1612	Add another course
1613	Add another department
1614	Add another faculty
1615	Add student
1616	Modify existing admin
1617	Add existing building
1618	Modify existing classroom
1619	Add existing course
1620	Modify existing department
1621	Admin Modify Enrollment List
1622	Modify existing faculty member
1623	Modify existing student

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.17 Management information system for an academic (ต่อ)

หมายเลข	ชื่อเอกสารการบรรยายผู้สแกน
1624	Remove administrator
1625	Remove building
1626	Remove classroom
1627	Remove course
1628	Remove department
1629	Remove faculty member
1630	Remove student



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

## การใช้งานระบบ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ข.1 คู่มือการใช้งานระบบ

### ข.1.1 ขั้นตอนการค้นคืน

ผู้ใช้งานสามารถเข้ามาใช้งานได้โดย ไปที่ โปรแกรม Internet Explorer พิมพ์ URL ไปที่ <http://127.0.0.1:8080/myproj> จะปรากฏหน้าต่างดังภาพที่ ข.1

Use Case Retrieval System	
Query Use Case by Use Case Structure	
Weight By :	Select Method ▼
Similarity By :	Select Method ▼
Use Case Name :	<input type="text"/> Select Weigth Value ▼
Objective :	<input type="text"/> Select Weigth Value ▼
Actor :	<input type="text"/> Select Weigth Value ▼
Relationship :	Association : <input type="text"/> Select Weigth Value ▼
	Include : <input type="text"/> Select Weigth Value ▼
	Extend : <input type="text"/> Select Weigth Value ▼
	Generalization : <input type="text"/> Select Weigth Value ▼
Precondition :	<input type="text"/> Select Weigth Value ▼
Postcondition :	<input type="text"/> Select Weigth Value ▼
Normal Flow of Events :	<input type="text"/> Select Weigth Value ▼
Subflow :	<input type="text"/> Select Weigth Value ▼
Alternative or Exceptional Flow of Events :	<input type="text"/> Select Weigth Value ▼
<input type="button" value="Query"/> <input type="button" value="Reset"/>	

ภาพที่ ข.1 หน้าจอการระบบค้นคืน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้ใช้งานเลือกวิธีถ่วงน้ำหนักค่า และวิธีการคำนวณค่าความคล้ายที่

Weight By : (Select method)

Similarity By : (Select method)

และคำค้นที่คอมพิวเตอร์เน้นที่ที่ต้องการ ดังแสดงที่ภาพ ข.2 ในส่วน หมายเลข 1

และใส่ค่าน้ำหนักที่ผู้ใช้ต้องการในคอมพิวเตอร์ที่เลือกด้วย ในส่วนที่ 2

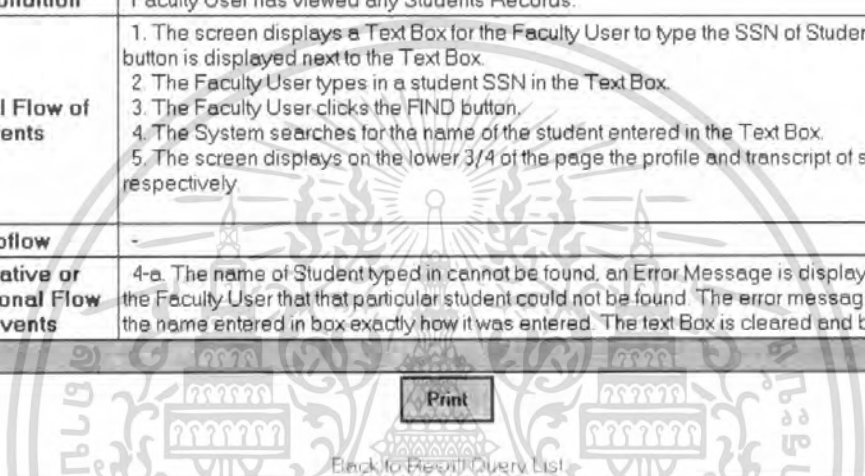
Use Case Retrieval System	
Query Use Case by Use Case Structure	
Weight By :	IDF <input type="text"/>
Similarity By :	Overlap Measure <input type="text"/>
Use Case Name :	student report <input type="text"/>
Objective :	4 <input type="text"/>
Actor :	Select Weighth Value <input type="text"/>
Relationship :	Association : Select Weighth Value <input type="text"/>
	Include : Select Weighth Value <input type="text"/>
	Extend : Select Weighth Value <input type="text"/>
	Generalization : Select Weighth Value <input type="text"/>
Precondition :	Select Weighth Value <input type="text"/>
Postcondition :	Select Weighth Value <input type="text"/>
Normal Flow of Events :	Select Weighth Value <input type="text"/>
Subflow :	Select Weighth Value <input type="text"/>
Alternative or Exceptional Flow of Events :	Select Weighth Value <input type="text"/>
<input type="button" value="Query"/> <input type="button" value="Reset"/>	

ภาพที่ ข.2 หน้าจอระบบค้นคืนที่ผู้ใช้กรอกรายละเอียด

เมื่อกรอกคำค้น วิธีการถ่วงน้ำหนัก วิธีการคำนวณค่าความคล้าย และค่าน้ำหนักของแต่ละคอมพิวเตอร์แล้ว ก็ กดปุ่ม **Query** เพื่อส่งข้อมูลไปค้นในฐานข้อมูลต่อไป หรือ กดปุ่ม **Reset** เมื่อ ต้องการยกเลิกค่าต่างๆแล้วล้างค่าทั้งหมดเพื่อเริ่มกรอกข้อมูลใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Use Case Retrieval System		
Result Query Use Case by Use Case Structure		
Use Case ID	1608	
Use Case Name	Faculty View Student Records	
Objective	Use Case to allow a Faculty User to view Students Records.	
Actor	a Faculty User	
Relationship	Association	a Faculty User
	Include	-
	Extend	-
	Generalization	-
Precondition	Faculty User has chosen to View Students Records.	
Postcondition	Faculty User has viewed any Students Records.	
Normal Flow of Events	<ol style="list-style-type: none"> <li>1. The screen displays a Text Box for the Faculty User to type the SSN of Student in. A FIND button is displayed next to the Text Box.</li> <li>2. The Faculty User types in a student SSN in the Text Box.</li> <li>3. The Faculty User clicks the FIND button.</li> <li>4. The System searches for the name of the student entered in the Text Box.</li> <li>5. The screen displays on the lower 3/4 of the page the profile and transcript of student respectively.</li> </ol>	
Subflow	-	
Alternative or Exceptional Flow of Events	4-a. The name of Student typed in cannot be found, an Error Message is displayed informing the Faculty User that that particular student could not be found. The error message also shows the name entered in box exactly how it was entered. The text Box is cleared and back to 1.	
		

ภาพที่ ข.4 หน้าจอตัวอย่างเอกสารการบรรยายยูสเคสที่ได้จากการค้นคืน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ข.1.2 การเพิ่มเอกสารการบรรยายยูสเคสเข้ามาในระบบ

เมื่อต้องการเพิ่มเอกสารการบรรยายยูสเคสในฐานะข้อมูลก็สามารถทำได้โดยเข้าไปที่

<http://127.0.0.1:8080/myproj/store.jsp> และจะมีหน้าจอดังภาพที่ ข.5

Use Case Retrieval System	
Add Use Case Collection	
Diagram ID :	<input type="text"/>
Use Case ID :	<input type="text"/>
Use Case Name :	<input type="text"/>
Objective :	<input type="text"/>
Actor :	<input type="text"/>
Relationship :	Association : <input type="text"/>
	Include : <input type="text"/>
	Extend : <input type="text"/>
	Generalization : <input type="text"/>
Precondition :	<input type="text"/>
Postcondition :	<input type="text"/>
Normal Flow of Events :	<input type="text"/>
Sub Flow :	<input type="text"/>
Alternative or Exceptional Flow of Events :	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

ภาพที่ ข.5 หน้าจอการเพิ่มเอกสารการบรรยายยูสเคสเข้ามาในระบบ

เมื่อมีหน้าจอ ดังภาพที่ ข.5 ขึ้นมาครบแล้วให้กรอกข้อมูลของเอกสารการบรรยายยูสเคสตามแต่ละคอม โปเน้นต์ที่ต้องการจัดเก็บซึ่งตรงกับชื่อในตาราง ดังแสดงในภาพที่ ข.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

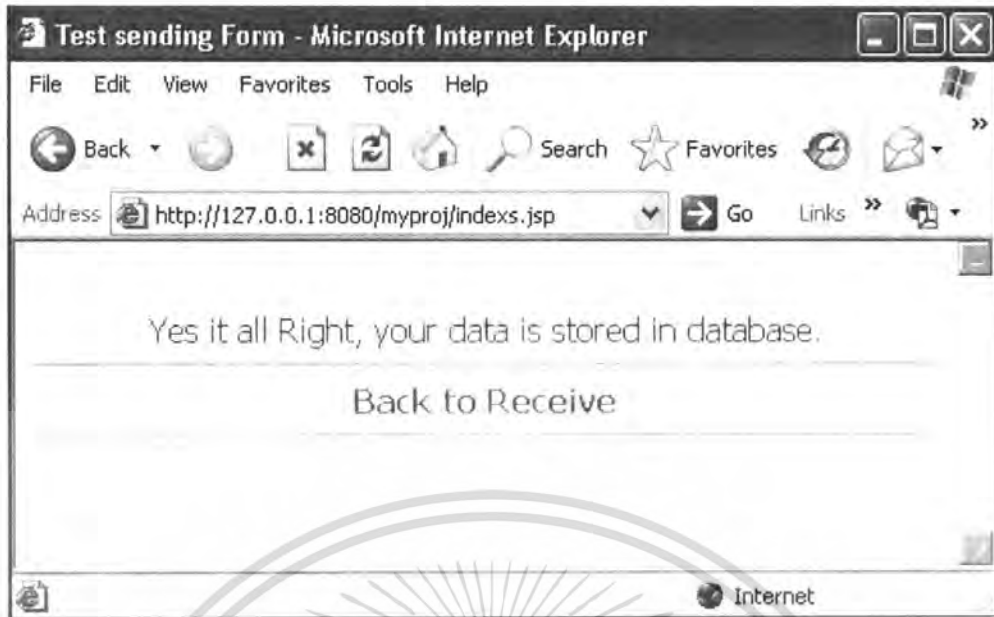
Use Case Retrieval System	
Add Use Case Collection	
Diagram ID :	<input type="text" value="17"/>
Use Case ID :	<input type="text" value="1701"/>
Use Case Name :	<input type="text" value="Request report about members"/>
Objective :	<input type="text" value="This use case is for a user or a member to request some reports about members."/>
Actor :	<input type="text" value="A user, A member"/>
Relationship :	Association : <input type="text" value="A user, A member"/>
	Include : <input type="text"/>
	Extend : <input type="text"/>
	Generalization : <input type="text" value="teacher, a parent, an internet inspector, an ISP, an administrator."/>
Precondition :	<input type="text"/>
Postcondition :	<input type="text"/>
Normal Flow of Events :	1. A user selects menu to view reports. 2. A user selects a type of reports.
Sub Flow :	<input type="text"/>
Alternative or Exceptional Flow of Events :	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

ภาพที่ ข.6 แสดงการใส่ข้อมูลเอกสารการบรรยายยูสเคส

หลังจากกรอกข้อมูลเรียบร้อยแล้วกดปุ่ม  เพื่อส่งข้อมูล ไปจัดเก็บในฐานข้อมูล หรือ กดปุ่ม  เพื่อล้างค่าที่กรอกไปและเริ่มต้นใหม่

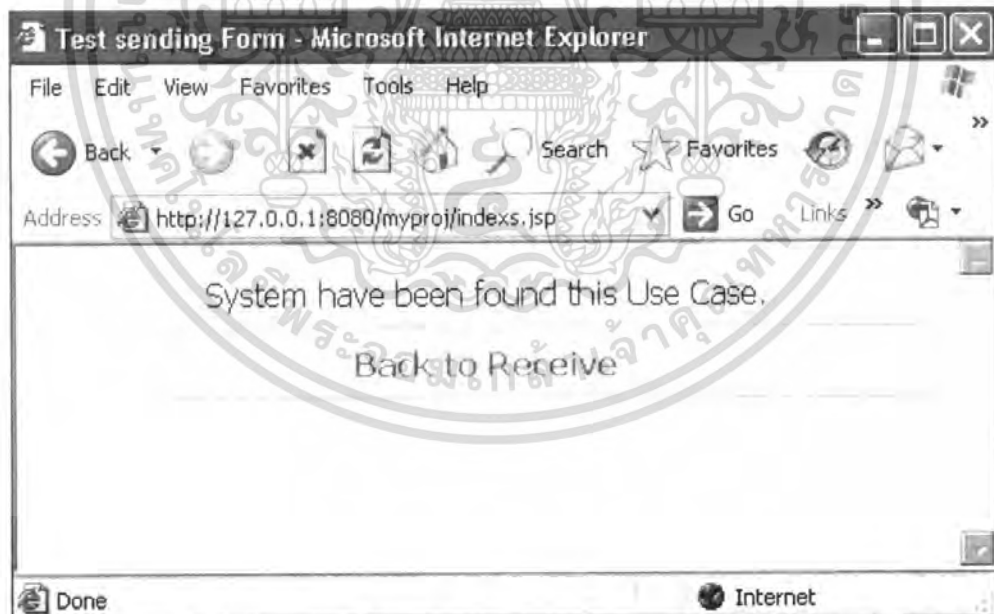
หลังจากกดปุ่ม  แล้วถ้า ไม่มีข้อผิดพลาดระบบจะพิมพ์ข้อความดังภาพที่ ข.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ ข.7 แสดงสถานะการบันทึกข้อมูลเอกสารการบรรยายยูสเคส

และถ้ามีข้อผิดพลาดเช่นมีการใส่ข้อมูลรหัสเอกสารการบรรยายยูสเคสซ้ำกับที่มีอยู่แล้ว ระบบก็จะแจ้งข้อความดังภาพที่ ข.8



ภาพที่ ข.8 แสดงการใส่ข้อมูลเอกสารการบรรยายยูสเคสซ้ำกับที่มีอยู่แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.

หมายเลขเอกสารการบรรยายยุคศตที่เกี่ยวข้องกับแต่ละหัวข้อของการทดลอง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ค.1 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการเรียนการสอน

หัวข้อ	หมายเลขเอกสารการบรรยายยยุสคต
ก) ระบบการเรียน การสอน	0201
	0202
	0203
	0204
	0205
	0206
	0207
	0208
	0209
	0210
	0211
	0212
	0213
	0214
	0215
	0216
	0217
	0218
	0219
	0220
0221	
0222	
0223	
0224	
1505	
1506	
1507	
1508	
1509	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ค.1 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการเรียนการสอน (ต่อ)

หัวข้อ	หมายเลขเอกสารการบรรยายยยusst
ก) ระบบการเรียน การสอน (ต่อ)	1510
	1513
	1515
	1516
	1517
	1518
	1519
	1520
	1602
	1603
	1604
	1605
	1606
	1607
	1608
	1610
	1611
	1612
	1613
	1614
	1615
	1617
	1618
	1619
	1620
	1622
1623	
1625	
1626	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ค.1 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการเรียนการสอน (ต่อ)

หัวข้อ	หมายเลขเอกสารการบรรยายยยุสเทศ
ก) ระบบการเรียน การสอน (ต่อ)	1627
	1628
	1629
	1630

ตารางที่ ค.2 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการค้าขายสินค้า

หัวข้อ	หมายเลขเอกสารการบรรยายยยุสเทศ
ข) ระบบการค้าขายสินค้า	0305
	0306
	0307
	0308
	0320
	0321
	0323
	0328
	0329
	0605
	0607
	0701
	0702
	0703
	0704
	0705
	0706
	0816
	0817
	0901
0902	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ตารางที่ ค.4 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการคำนวณการเงิน

หัวข้อ	หมายเลขเอกสารการบรรยายยยศาสตร์
ง) การคำนวณการเงิน	0321
	0323
	0325
	0608
	0609
	0610
	0709
	0710
	0711
	0712
	0713
	0714
	0715
	0716
	0717
	0724
	0820
	1010
	1011
	1120
1310	
1401	
1407	
1409	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ค.5 หมายเลขเอกสารที่เกี่ยวข้องกับหัวข้อระบบการสร้างรายงาน

หัวข้อ	หมายเลขเอกสารการบรรยายยุดศศ
จ) การสร้างรายงาน	0110
	0323
	0502
	0505
	0508
	0719
	0720
	0909
	0911
	1012
	1013
	1014
	1015
	1115
	1117
1119	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้