

โปรแกรมจัดการเอกสาร  
DOCUMENT MANAGEMENT PROGRAM

นนท์ปวีธ สามพวงทอง  
วิรัชชัย เรื่องชาติสมบัติ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต  
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ปีการศึกษา 2559

โปรแกรมจัดการเอกสาร  
DOCUMENT MANAGEMENT PROGRAM

นนท์ปวิธ สามพวงทอง  
วิรัชชัย เรื่องชาติสมบัติ

ปริญญาานิพนธ์เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต  
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ปีการศึกษา 2559

ปริญญานิพนธ์ปีการศึกษา 2559

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
เรื่อง โปรแกรมจัดการเอกสาร

DOCUMENT MANAGEMENT PROGRAM

ผู้จัดทำ

1. นายนนท์ปวีร์ สามพวงทอง รหัสนักศึกษา 56010631
2. นายวิรัชชัย เรืองชาติสมบัติ รหัสนักศึกษา 56011148



อาจารย์ที่ปรึกษา

(อาจารย์บัณฑิต พัสยา)

# โปรแกรมจัดการเอกสาร

นายหน้ปวิธ	สามพวงทอง	56010631
นายวิวิธชัย	เรืองชาติสมบัติ	56011148
อาจารย์บัณฑิต	พัสยา	อาจารย์ที่ปรึกษา
ปีการศึกษา 2559		

## บทคัดย่อ

การทำงานในหน่วยงานต่างๆ ต้องยอมรับว่าเอกสารเป็นหนึ่งในเรื่องที่สำคัญและใช้งานเป็นประจำอย่างหลีกเลี่ยงไม่ได้ ไม่ว่าจะเป็นเอกสารข้อมูล ประกาศ หรือ แบบฟอร์มต่างๆ ผู้คนเป็นจำนวนมากไม่น้อยมีปัญหาเกี่ยวกับเรื่องนี้ อาจจะเป็นเรื่องของการค้นหาเอกสาร ความสะดวกสบายในการเข้าถึงและเรียกใช้เอกสาร หรือแม้แต่ตัวผู้จัดทำเอกสารเหล่านั้นเองยังมีปัญหาเกี่ยวกับการเผยแพร่แจกจ่ายเอกสาร ผู้จัดทำจึงมีแนวคิดที่จะจัดการกับปัญหาเหล่านั้น โดยการจัดทำโปรแกรมที่สามารถช่วยเรื่องการจัดการเกี่ยวกับเอกสารขึ้นมา

โปรแกรมนี้เขียนด้วยภาษา python โดยโปรแกรมจะใช้หลักการ fuzzy logic ในการค้นหาเอกสารที่มีความเกี่ยวข้องกับที่ผู้ใช้งานต้องการมากที่สุด มีการออกแบบฐานข้อมูลที่ได้มาตรฐานตามนิยามของ SNF และสามารถเข้าถึงได้เฉพาะเอกสารที่มีสิทธิ์ในการเข้าถึง เหมาะกับการใช้ในบริษัทหรือองค์กรทุกรูปแบบ

# DOCUMENT MANAGEMENT PROGRAM

Mr. Nonpawit Sampuangthong 56010631

Mr. Vivitchai Ruangchartsombat 56011148

Mr. Bundit Pasaya Advisor

Academic Year 2016

## ABSTRACT

A doing work in company or organization, the document is one of the most important and most active either information document notice document or form. Many people have a problem about this in topic “search” “convenience in access document” or “distributed document”.

We are thinking of get rid of these problem by develop the application can help manage document.

This application development with python by using the fuzzy logic for search document that the most relevant which user wanted .Database designed by Fifth normal form (5NF) .Only those who are authorized can access the document. This application is convenient for any company or organization.

## กิตติกรรมประกาศ

ปริญญานิพนธ์เล่มนี้ สำเร็จได้โดยได้รับการช่วยเหลือและการให้คำปรึกษาจาก อาจารย์ บัณฑิต พัสยา อาจารย์ที่ปรึกษาโครงการ และความรู้ที่ได้ร่ำเรียนมาจากคณาจารย์ทุกท่าน ผู้จัดทำจึง ขอขอบพระคุณเป็นอย่างสูง

นนท์ปวีธ สามพวงทอง  
วิวิธชัย เรืองชาติสมบัติ

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย .....	I
บทคัดย่อภาษาอังกฤษ .....	II
กิตติกรรมประกาศ .....	III
สารบัญ .....	IV
สารบัญ(ต่อ) .....	V
สารบัญตาราง .....	VI
สารบัญรูป .....	VII
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มาของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 ขอบเขตของโครงการ.....	1
1.4 วิธีการดำเนินงาน.....	1
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.6 ข้อยกเว้นของโปรแกรม.....	2
1.7 ส่วนประกอบวิทยานิพนธ์.....	2
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	3
2.1 Fuzzy Logic.....	3
2.2 ตารางเก็บข้อมูล.....	8
2.3 Python.....	10
บทที่ 3 การออกแบบและการพัฒนา.....	12
3.1 สิ่งที่ต้องการจากระบบ.....	12
3.2 ภาพรวมของระบบ.....	13
3.3 รายละเอียดการทำงานของระบบ.....	13
3.4 ตารางเก็บข้อมูล.....	14

## สารบัญ(ต่อ)

	หน้า
บทที่ 4 วิธีการดำเนินงาน.....	16
4.1 การพัฒนาโปรแกรม.....	16
4.2 เทคโนโลยีที่ใช้.....	18
4.3 ผลของการดำเนินงาน.....	22
บทที่ 5 บทสรุป.....	26
5.1 สรุปผล.....	26
5.2 ปัญหาที่เกิดขึ้น.....	26
บรรณานุกรม.....	27
ภาคผนวก.....	28

## สารบัญตาราง

ตาราง	หน้า
ตาราง 2.1 ตาราง keyword relationship.....	7
ตาราง 3.1 ตาราง Keyword .....	14
ตาราง 3.2 ตาราง Document .....	14
ตาราง 3.3 ตาราง Document & Keyword .....	15
ตาราง 3.4 ตาราง Account.....	15
ตาราง 3.5 ตาราง Fuzzy.....	15

# สารบัญรูป

รูป	หน้า
รูปที่ 2.1 ตรรกะแบบจริงเท็จ (บูลีนลอจิก) และตรรกะแบบพีชชี (พีชชีลอจิก).....	4
รูปที่ 2.2 สมการ Index Term Relationship .....	5
รูปที่ 2.3 ยูเนียนของพีชชีเซต A และ B.....	6
รูปที่ 2.4 อินเตอร์เซกชันของพีชชีเซต A และ B.....	6
รูปที่ 2.5 คอมพลีเมนต์ของพีชชีเซต A.....	7
รูปที่ 3.1 ไดอะแกรมแสดงขั้นตอนการอัปเดตเอกสาร.....	13
รูปที่ 3.2 ไดอะแกรมแสดงขั้นตอนการค้นหาเอกสาร.....	14
รูปที่ 4.1 Class Diagram.....	17
รูปที่ 4.2 XAMPP.....	18
รูปที่ 4.3 phpMyAdmin.....	19
รูปที่ 4.4 FileZila.....	19
รูปที่ 4.5 kivy.....	20
รูปที่ 4.6 Login.....	22
รูปที่ 4.7 MyAccount.....	23
รูปที่ 4.8 Browse.....	24
รูปที่ 4.9 Upload.....	24
รูปที่ 4.10 Search.....	25

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของโครงการ

ในการทำงานต่างๆ เอกสารถือเป็นเรื่องที่สำคัญและใช้งานเป็นประจำอย่างหลีกเลี่ยงไม่ได้ ไม่ว่าจะเป็นเอกสารข้อมูล ประกาศ หรือ แบบฟอร์มต่างๆ ปัญหาที่เกิดขึ้นเกี่ยวกับเรื่องของเอกสาร อาจจะเป็นเรื่องของการค้นหาเอกสาร ความสะดวกสบายในการเข้าถึงและเรียกใช้เอกสาร หรือ แม้แต่ตัวผู้ที่จัดทำเอกสารเหล่านั้นเองยังมีปัญหาเกี่ยวกับการเผยแพร่แจกจ่ายเอกสาร ผู้จัดทำจึงมีแนวคิดที่จะจัดการกับปัญหาเหล่านั้น โดยการจัดทำโปรแกรมที่สามารถช่วยเรื่องของการจัดการเกี่ยวกับเอกสารขึ้นมา

### 1.2 วัตถุประสงค์ของโครงการ

- 1) เพื่อศึกษาและพัฒนาการสร้าง โปรแกรมบนระบบปฏิบัติการ window
- 2) เพื่อช่วยให้ผู้ใช้งานสามารถจัดการเอกสารได้รวดเร็วและมีประสิทธิภาพ
- 3) เพื่อช่วยอำนวยความสะดวกแก่การเข้าถึงเอกสาร
- 4) เพื่อเป็นประโยชน์ในการเผยแพร่และแจกจ่ายเอกสาร

### 1.3 ขอบเขตของโครงการ

- 1) ระบบสามารถอัปโหลดเอกสาร พร้อมทั้งจัดเก็บ keyword ที่มีในเอกสารนั้นๆ ไปยังฐานข้อมูลได้
- 2) ระบบสามารถค้นหาเอกสาร โดยจัดเรียงเรียงตามวันที่ที่อัปโหลดของเอกสารได้
- 3) ระบบสามารถค้นหาเอกสารที่มีความตรงประเด็นมากที่สุด จาก query ที่ผู้ใช้ส่งเข้าไปโดยคำนวณ โดยใช้แนวคิดของ fuzzy logic
- 4) ระบบมีการจำกัดสิทธิ์ในการเข้าถึงเอกสาร โดยมีการให้เฉพาะผู้ที่มีสิทธิ์จึงจะสามารถเข้าถึงเอกสารนั้นๆ ได้
- 5) สามารถดาวน์โหลดเอกสารจากระบบได้

### 1.4 วิธีการดำเนินงาน

- 1) กำหนดขอบเขตของการทำงาน วัตถุประสงค์ รวมทั้งภาษาที่จะใช้ในการเขียน โปรแกรม

- 2) วิเคราะห์ระบบจากการศึกษาหลักการ ทฤษฎี เอกสารที่เกี่ยวข้อง ปรัชญาอาจารย์ที่ปรึกษา  
โครงการ
- 3) ออกแบบส่วนต่างๆของระบบ
- 4) พัฒนาโปรแกรม
- 5) ทดสอบโปรแกรมและแก้ไขข้อบกพร่องที่เกิดขึ้น เพื่อให้ได้โปรแกรมที่มีความถูกต้องและ  
สมบูรณ์ที่สุด

### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้โปรแกรมที่สามารถที่สามารช่วยให้อผู้ใช้สามารถจัดการกับเอกสาร ในเรื่องของการ  
ค้นหา ความสะดวกสบายในการเข้าถึงและเรียกใช้เอกสาร รวมไปถึงการเผยแพร่แจกจ่าย  
เอกสาร
- 2) โปรแกรมที่พัฒนาสามารถอำนวยความสะดวก และ เป็นที่พึงพอใจของผู้ใช้งาน
- 3) สามารถนำโปรแกรมไปประยุกต์ใช้ได้มีโอกาสต่อไป

### 1.6 ข้อจำกัดของโปรแกรม

- 1) ไม่สามารถรองรับภาษาอื่นนอกจากภาษาไทยและภาษาอังกฤษได้
- 2) เกิดความยากลำบากในการใส่ keyword ถ้าหาก keyword มีจำนวนมาก
- 3) query และ keyword ไม่สามารถเป็น + - \* ได้
- 4) ถ้าหาก query ไม่ใช่ keyword จะไม่สามารถหาผลลัพธ์ในส่วนของการจัดอันดับตามความตรง  
ประเด็นของเอกสารได้

### 1.7 ส่วนประกอบวิทยานิพนธ์

ปริญญานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ  
 บทที่ 1 กล่าวถึงที่มาและความสำคัญของโครงการ วัตถุประสงค์ของโครงการ ขอบเขตของ  
 โครงการ วิธีการเนิการ ประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบของปริญญานิพนธ์  
 บทที่ 2 กล่าวถึง ทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้องกับการจัดทำ โครงการ ดังนั้น  
 บทที่ 3 กล่าวถึงการออกแบบและการทำงานของระบบ  
 บทที่ 4 กล่าวถึงการทดลองและผลของการทดลองของโครงการ  
 บทที่ 5 กล่าวถึง บทสรุปและข้อเสนอแนะของโครงการ บรรยายสรุปโครงการ ปัญหา  
 อุปสรรค แนวทางการแก้ไข และ แนวทางในการพัฒนาต่อ

## บทที่ 2

# ทฤษฎีที่เกี่ยวข้อง

### 2.1 FUZZY LOGIC

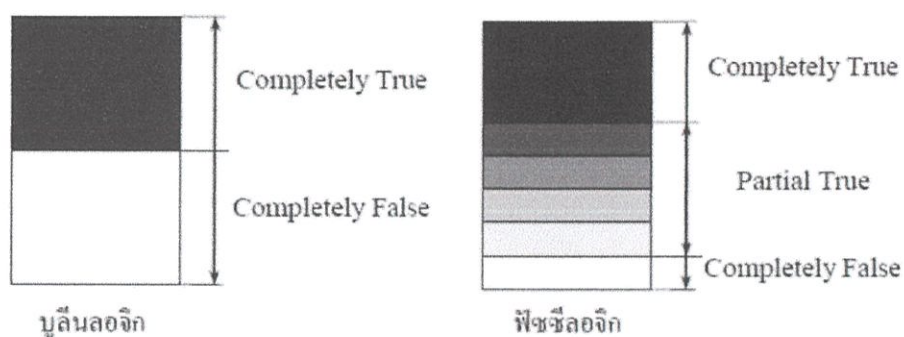
ตรรกศาสตร์คลุมเครือ หรือ ฟัซซีลอจิก (fuzzy logic) พัฒนาจาก ทฤษฎีเซตวิภังชน โดยเป็นการใช้เหตุผลแบบประมาณ ซึ่งแตกต่างจากการใช้เหตุผลแบบเด็ดขาดในลักษณะ ถูก/ผิด ใช่/ไม่ใช่ ของ ตรรกศาสตร์แบบฉบับ (classical logic) ตรรกศาสตร์คลุมเครือนั้นสามารถถือเป็นการประยุกต์ใช้งานเซตวิภังชน เพื่อจำลองการตัดสินใจของผู้เชี่ยวชาญ ต่อปัญหาที่ซับซ้อน

ค่าระดับความจริง ในตรรกศาสตร์คลุมเครือนั้นมักจะสับสนกับ ค่าความน่าจะเป็น ซึ่งมีแนวความคิดที่แตกต่างกัน ค่าระดับความจริงคลุมเครือนั้นใช้ในการระบุ ค่าความเป็นสมาชิก ของเซต แต่ค่าความน่าจะเป็นนั้นระบุความเป็นไปได้ของสภาพการณ์แต่ละรูปแบบที่อาจจะเกิดขึ้น ตัวอย่างเช่น สมมติว่า นาย ก กำลังเดินเข้าบ้าน สถานะของนาย ก ตามตรรกศาสตร์แบบฉบับ คือ "อยู่ในบ้าน" หรือ "อยู่นอกบ้าน" แต่หากเขากำลังยืนอยู่ระหว่างช่องประตู เราอาจพิจารณาได้ว่าเขา "อยู่ในบ้านบางส่วน" ระดับของสถานะกึ่งนี้ จะระบุด้วยค่าความเป็นสมาชิกของเซตวิภังชน สมมุติเขาเพิ่งจะก้าวปลายนิ้วเท้าผ่านข้ามธรณีประตูเข้าบ้าน เราอาจกล่าวว่า นาย ก นั้น 0.99 "อยู่นอกบ้าน" ซึ่งต่างจากความน่าจะเป็นของเหตุการณ์สุ่ม (เช่น ความน่าจะเป็นระบุผลลัพธ์ของการโยนเหรียญ แต่ผลลัพธ์จะออก หัว หรือ ก้อย) หากพิจารณาความน่าจะเป็นที่นาย ก "อยู่นอกบ้าน" และ "อยู่ในบ้าน" จะออกผลลัพธ์เป็น นาย ก อยู่นอกบ้าน หรือ ในบ้าน ไม่ได้จำลองสถานะกึ่ง คือ กำลังยืนอยู่ที่ประตู เซตวิภังชนนี้มีหลักการพื้นฐานจากเซตที่มีขอบเขตคลุมเครือ ไม่ชัดเจน ไม่ได้มีพื้นฐานจากการสุ่ม

ตรรกศาสตร์คลุมเครือนั้น สามารถระบุค่าความเป็นสมาชิกของเซต (set membership values) ด้วยค่าระหว่าง 0 และ 1 ทำให้เกิดระดับกึ่งในลักษณะของ สีเทา นอกจาก ขาว และ ดำ ซึ่งมีประโยชน์ในการจำลองระดับซึ่งสามารถระบุด้วยคำพูด "เล็กน้อย" "ค่อนข้าง" "มาก" โดยใช้ค่าความเป็นสมาชิกของเซตบางส่วน

### 2.1.1 แนวคิดพื้นฐานเกี่ยวกับฟัซซีลอจิก

ตรรกะแบบฟัซซี(fuzzy logic) เป็นเครื่องมือที่ช่วยในการตัดสินใจภายในได้ความไม่แน่นอนของข้อมูลโดยยอมให้มีความยืดหยุ่นได้ ใช้หลักเหตุผลที่คล้ายการเลียนแบบวิถีความคิดที่ ชับซ้อนของมนุษย์ฟัซซีลอจิกมีลักษณะที่พิเศษกว่าตรรกะแบบจริงเท็จ (Boolean logic) เป็นแนวคิด ที่มีการต่อขยายในส่วนของความจริง(partial true) โดยค่าความจริงจะอยู่ในช่วงระหว่างจริง (completely true) กับเท็จ(completely false) ส่วนตรรกศาสตร์เดิมจะมีค่าเป็นจริงกับเท็จเท่านั้นแสดงดังภาพที่ 2.1



บูลีนลอจิก

ฟัซซีลอจิก

รูปที่ 2.1 ตรรกะแบบจริงเท็จ (บูลีนลอจิก) และตรรกะแบบฟัซซี (ฟัซซีลอจิก)

ความเป็นฟัซซี (fuzziness) มีชื่อเรียกว่า มัลติวาลานซ์(multivalance) ซึ่งมีค่าที่ความเป็นสมาชิก มากกว่า 2 ค่าและแตกต่างกับ ไบวาลานซ์ (bivalance) ที่มีความเป็นสมาชิกเพียง 2 ค่า ฟัซซีเซต (Fuzzy set) เป็นเครื่องมือทางคณิตศาสตร์ที่สื่อถึง “ความไม่แน่นอน (uncertainty)” สามารถที่ ไม่ใช่ เพียง 2 กรณีซึ่งหากกำหนดว่าคนที่อ้วนคือคนที่มีน้ำหนักมากกว่า 75 กิโลกรัม คอมพิวเตอร์จะ ให้ผลว่าคนที่มีน้ำหนัก 74.50 กิโลกรัม ไม่จัดเป็นคนที่อ้วน จะสร้างและ กำหนดรูปแบบ (modeling) ของลักษณะความไม่แน่นอนที่เป็นความคลุมเครือความไม่ตายตัว รวมถึงความขาดข้อมูลบางส่วน โดย ทฤษฎีของฟัซซีเซตจะใช้ลักษณะความหมายตัวแปร (linguistic) มากกว่าปริมาณ (quantitative) ของ ตัวแปรเช่น การหาความหมายของ “คนที่อ้วน” เราไม่สามารถนิยามค่าความอ้วนที่ตรงกันและระบุเป็น หนึ่งเดียว(identical) สำหรับคนที่อ้วน นาย ก. จะให้ความหมายของ “คนอ้วน” หมายถึงคนที่มี น้ำหนักมากกว่า 70 กิโลกรัม นาย ข. ให้ความหมายว่าเป็นคนที่มีน้ำหนักมากกว่า 75 กิโลกรัม ซึ่งทั้ง สองคนต่างแสดงความหมาย ของคำว่าคนที่อ้วน โดยเปรียบเทียบและในมุมมองของตัวเองความน้ำหนัก ของคน ในการท างานในมุมมองแบบฐานสอง (Binary sense) จะได้ผลเป็น ใช่หรือแต่จะเห็นว่า บุคคลนี้เป็นคน อ้วนน้ำหนักเกือบจะ 75 กิโลกรัม และถึงแม้ว่าบุคคลนี้จะมีน้ำหนัก 75 กิโลกรัม แต่ 4 หาก พิจารณาจากกลุ่มคนที่น้ำหนักเฉลี่ย 90 กิโลกรัม บุคคลนี้ก็จะไม่จัดอยู่ในกลุ่มคนที่อ้วน แสดง

ให้เห็นว่าความอ้วนไม่ได้มีลักษณะความไม่แน่นอนแบบสุ่ม จากการศึกษาปัญหาทั่ว ๆ ไปจะแสดงถึง รูปแบบลักษณะการกระจายของปัญหา

### 2.1.2 การดำเนินการ Fuzzy logic และ Information storage and retrieval

ในการนำ Fuzzy logic มาใช้งานในโครงการจำเป็นต้องมีการผ่านขั้นตอนหลายขั้นตอน เพื่อที่จะหาความสัมพันธ์ระหว่าง keyword แต่ละตัวได้อย่างถูกต้องและแม่นยำ

โดยเราจะใช้สมการ Index Term Relationship ซึ่งเป็นสมการสำคัญที่ใช้ในการคำนวณค่าความสัมพันธ์ของ keyword แต่ละตัวดังรูป ค่าผลลัพธ์ที่ได้จากสมการจะอยู่ระหว่าง 0 ถึง 1 ตามทฤษฎีที่กล่าวไว้ข้างต้น

$$c_{i,j} = \frac{n_{i,j}}{n_i + n_j - n_{i,j}}$$

$c_{i,j}$  คือ ความสัมพันธ์ของศัพท์เวิร์ด  $i$  กับ ศัพท์เวิร์ด  $j$   
 $n_{i,j}$  คือ จำนวนเอกสารที่มีทั้งศัพท์เวิร์ด  $i$  และศัพท์เวิร์ด  $j$   
 $n_i$  คือ จำนวนเอกสารที่มีศัพท์เวิร์ด  $i$   
 $n_j$  คือ จำนวนเอกสารที่มีศัพท์เวิร์ด  $j$

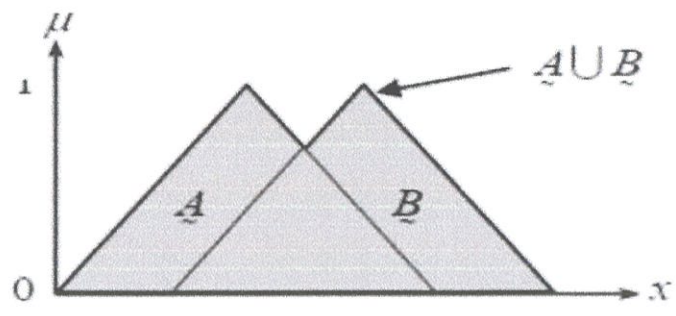
รูปที่ 2.2 สมการ Index Term Relationship

### 2.1.3 การดำเนินการทางพีชคณิตเซต

การดำเนินการของพีชคณิตเซตมีคุณสมบัติเหมือนกับเซตโดยทั่วไป มีการดำเนินการ (operation) คือ Union Intersection และ Complement

- 1) ยูเนียน (Union) ของพีชคณิตเซต จะเป็น OR operation ในสมการที่ 2.1 และภาพที่ 2.3

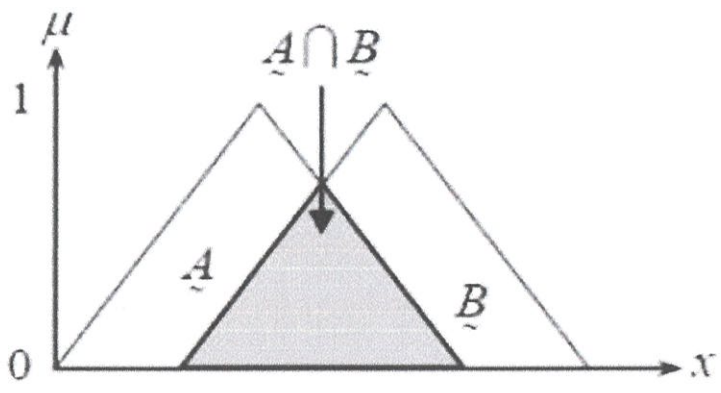
$$\mu(A \cup B) = \mu A \text{ OR } \mu B = \max(\mu A, \mu B) \quad (2.1)$$



รูปที่ 2.3 ยูเนียนของฟัซซีเซต A และ B

2) อินเตอร์เซกชัน (Intersection) ของฟัซซีเซต จะเป็น AND operation ในสมการที่ 2.2 และภาพที่ 2.4

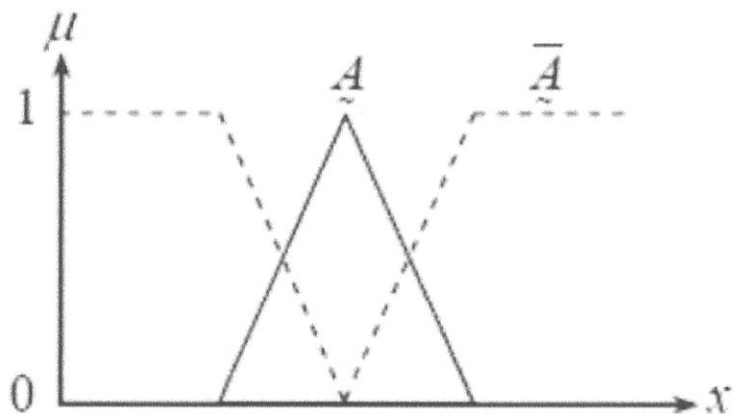
$$\mu(A \cap B) = \mu A \text{ AND } \mu B = \min(\mu A, \mu B) \tag{2.2}$$



รูปที่ 2.4 อินเตอร์เซกชันของฟัซซีเซต A และ B

3) คอมพลิเมนต์ (Complement) ของฟัซซีเซต จะเป็น NOT operation ในสมการที่ 2.3 และภาพที่ 2.5

$$\mu \bar{A} = \text{NOT } A = 1 - \mu A \tag{2.3}$$



รูปที่ 2.5 คอมพลิเมนต์ของฟัซซีเซต A

### 2.1.4 Fuzzy Example

D1: {bird, cat, bird, cat, tiger, kitty, fish, fish}

D2: {cat, dog, tiger, zoo}

D3: {house, kitchen, bird, bird, cat, cat, kitty}

D4: {dog, rubber, house, dog}

D5: {tiger, forest, fish}

Input keyword: cat and dog

ค่าความสัมพันธ์ของ keyword คำนวณจาก  $\frac{n_{ij}}{n_i+n_j-n_{ij}}$  เช่น bird, cat =  $\frac{2}{2+3-2} = 0.67$

คำนวณค่าความสัมพันธ์ของแต่ละ keyword จะได้

ตาราง 2.1 keyword relationship

	Bird	Cat	Dog	Fish	Forest	House	Kitchen	Kitty	Rubber	Tiger	Zoo
Bird	1.00	0.67	0.00	0.33	0.00	0.33	0.50	1.00	0.00	0.25	0.00
Cat	0.67	1.00	0.25	0.25	0.00	0.25	0.33	0.67	0.00	0.50	0.33
Dog	0.00	0.25	1.00	0.00	0.00	0.33	0.00	0.00	0.50	0.25	0.50
Fish	0.33	0.25	0.00	1.00	0.50	0.00	0.00	0.33	0.00	0.67	0.00
Forest	0.00	0.00	0.00	0.50	1.00	0.00	0.00	0.00	0.00	0.33	0.00
House	0.33	0.25	0.33	0.00	0.00	1.00	0.50	0.33	0.50	0.00	0.00
Kitchen	0.50	0.33	0.00	0.00	0.00	0.50	1.00	0.50	0.00	0.00	0.00
Kitty	1.00	0.67	0.00	0.33	0.00	0.33	0.50	1.00	0.00	0.25	0.00

Rubber	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00	1.00	0.00	0.00
Tiger	0.25	0.50	0.25	0.67	0.33	0.00	0.00	0.50	0.00	1.00	0.33
Zoo	0.00	0.33	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.33	1.00

Ranking =  $\mu_{\text{Cat}}$  and  $\mu_{\text{Dog}} = \min(\mu_{\text{Cat}}, \mu_{\text{Dog}})$

- Doc1 = 1 and 0.25 (ไม่มี dog แต่มี cat) =  $\min(1, 0.25) = 0.25$
- Doc2 = 1 and 1 =  $\min(1, 1) = 1$
- Doc3 = 1 and 0.33 (ไม่มี dog แต่มี cat (0.25) และ house (0.33) =  $\max(0.25, 0.33) = \min(1, 0.33) = 0.33$
- Doc4 = 0.25 (ไม่มี cat แต่มี dog (0.25) และ house (0.25) =  $\max(0.25, 0.25)$ ) and 1 =  $\min(0.25, 1) = 0.25$
- Doc5 = 0.5 (ไม่มี cat แต่มี tiger (0.5) และ fish (0.25) =  $\max(0.5, 0.25)$ ) and 0 =  $\min(0.5, 0) = 0$

ดังนั้นจึงเรียงลำดับความตรงประเด็นของเอกสารได้เป็น Doc2>Doc3>Doc1>Doc4>Doc5

## 2.2 ตารางเก็บข้อมูล

ฐานข้อมูลประกอบด้วยกลุ่มการจัดการข้อมูลสำหรับผู้ใช้หนึ่งคนหรือหลายๆ คน โดยทั่วไปมักอยู่ในรูปแบบดิจิทัล วิธีการแบ่งชนิดของฐานข้อมูลได้รูปแบบหนึ่งคือแบ่งตามชนิดของเนื้อหา เช่น บรรณานุกรม, เอกสารตัวอักษร, สถิติ โดยฐานข้อมูลดิจิทัลจะถูกจัดการโดยใช้ระบบจัดการฐานข้อมูลซึ่งเก็บเนื้อหาฐานข้อมูล โดยอนุญาตให้สร้าง, ดูแลรักษา, ค้นหา และการเข้าถึงในรูปแบบอื่นๆ

### 2.2.1 ระบบจัดการฐานข้อมูล

ระบบจัดการฐานข้อมูล (DBMS) ประกอบด้วยซอฟต์แวร์ที่ใช้ในการจัดการฐานข้อมูล, จัดเตรียมพื้นที่ในการเก็บ, การเข้าถึง, ระบบรักษาความปลอดภัย, สำรองข้อมูล และสิ่งอำนวยความสะดวกอื่นๆ ระบบจัดการฐานข้อมูลสามารถแบ่งหมวดหมู่ได้ตามแบบจำลองฐานข้อมูล (Database model) ที่สนับสนุน อาทิเช่น แบบจำลองเชิงสัมพันธ์ (Relational model) หรือ ฐานข้อมูล XML เป็นต้น แบ่งตามประเภทของคอมพิวเตอร์ที่สนับสนุน อาทิเช่น server cluster หรือ โทรศัพท์พกพา เป็นต้น แบ่งตามประเภทของภาษาสอบถามที่ใช้ในการเข้าถึงฐานข้อมูล อาทิเช่น ภาษาสอบถามเชิงโครงสร้าง หรือ XQuery แบ่งตามประสิทธิภาพในการ trade-offs อาทิเช่น ขนาดที่ใหญ่ที่สุด หรือ ความเร็วสูงสุด หรือ อื่นๆ เป็นต้น ในบาง DBMS จะครอบคลุมมากกว่าหนึ่งหมวดหมู่ เช่น สนับสนุนภาษาสอบถามได้หลายๆ ภาษา

## 2.2.2 การออกแบบฐานข้อมูล

การออกแบบฐานข้อมูล (Designing Databases) มีความสำคัญต่อการจัดการระบบฐานข้อมูล (DBMS) ทั้งนี้เนื่องจากข้อมูลที่อยู่ภายในฐานข้อมูลจะต้องศึกษาถึงความสัมพันธ์ของข้อมูล โครงสร้างของข้อมูลการเข้าถึงข้อมูลและกระบวนการที่โปรแกรมประยุกต์จะเรียกใช้ฐานข้อมูล ดังนั้น เราจึงสามารถแบ่งวิธีการสร้างฐานข้อมูลได้ 3 ประเภท แต่ประเภทที่เราจะใช้งานคือ

รูปแบบความสัมพันธ์ข้อมูล (Relational model) เป็นลักษณะการออกแบบฐานข้อมูลโดยจัดข้อมูลให้อยู่ในรูปของตารางที่มีระบบคล้ายแฟ้ม โดยที่ข้อมูลแต่ละแถว (Row) ของตารางจะแทนเรคอร์ด (Record) ส่วน ข้อมูลแนวตั้งจะแทนคอลัมน์ (Column) ซึ่งเป็นขอบเขตของข้อมูล (Field) โดยที่ตารางแต่ละตารางที่สร้างขึ้นจะเป็นอิสระ ดังนั้นผู้ออกแบบฐานข้อมูลจะต้องมีการวางแผนถึงตารางข้อมูลที่เป็นต้องใช้ เช่นระบบฐานข้อมูลบริษัทแห่งหนึ่ง ประกอบด้วย ตารางประวัติพนักงาน ตารางแผนกและตารางข้อมูลโครงการ แสดงประวัติพนักงาน ตารางแผนก และตารางข้อมูลโครงการ

## 2.2.3 การออกแบบฐานข้อมูลเชิงสัมพันธ์

การออกแบบฐานข้อมูลในองค์กรขนาดเล็กเพื่อตอบสนองความต้องการของผู้ใช้งานอาจเป็นเรื่องที่ไม่ยุ่งยากนัก เนื่องจากระบบและขั้นตอนการทำงานภายในองค์กรไม่ซับซ้อน ปริมาณข้อมูลที่มีก็ไม่มาก และจำนวนผู้ใช้งานฐานข้อมูลก็มีเพียงไม่กี่คน หากทว่าในองค์กรขนาดใหญ่ ซึ่งมีระบบและขั้นตอนการทำงานที่ซับซ้อน รวมทั้งมีปริมาณข้อมูลและผู้ใช้งานจำนวนมาก การออกแบบฐานข้อมูลจะเป็นเรื่องที่มีความละเอียดซับซ้อน และต้องใช้เวลาในการดำเนินการนานพอสมควรทีเดียว ทั้งนี้ ฐานข้อมูลที่ได้รับการออกแบบอย่างเหมาะสมจะสามารถตอบสนองต่อความต้องการของผู้ใช้งานภายในหน่วยงานต่าง ๆ ขององค์กรได้ ซึ่งจะทำให้การดำเนินงานขององค์กรมีประสิทธิภาพดียิ่งขึ้น เป็นผลตอบแทนที่คุ้มค่าต่อการลงทุนเพื่อพัฒนาระบบฐานข้อมูลภายในองค์กร ทั้งนี้ การออกแบบฐานข้อมูลที่นำซอฟต์แวร์ระบบจัดการฐานข้อมูลมาช่วยในการดำเนินการ สามารถจำแนกหลักในการดำเนินการได้ 6 ขั้นตอน คือ

- 1) การรวบรวมและวิเคราะห์ความต้องการในการใช้ข้อมูล
- 2) การเลือกระบบจัดการฐานข้อมูล
- 3) การออกแบบฐานข้อมูลในระดับแนวคิด
- 4) การนำฐานข้อมูลที่ออกแบบในระดับแนวคิดเข้าสู่ระบบจัดการฐานข้อมูล
- 5) การออกแบบฐานข้อมูลในระดับกายภาพ
- 6) การนำฐานข้อมูลไปใช้และการประเมินผล

## 2.3 Python

ภาษาไพทอน (Python programming language) เป็นภาษาโปรแกรมระดับสูง เพื่อใช้งานทั่วไป เป็นแบบอินเทอร์พรีเตอร์ (ประมวลผลไปที่ละบรรทัดและปฏิบัติ) ที่สร้างโดย กิโด ฟาน รอสซัม (Guido van Rossum) ในพ.ศ. 2533 ซึ่งถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux, Windows NT, Windows 2000, Windows XP หรือแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาไพทอนนี้เป็น OpenSource เหมือนอย่าง PHP ทำให้ทุกคนสามารถที่จะนำ Python มาพัฒนาโปรแกรมของเราได้ฟรีๆ โดยไม่ต้องเสียค่าใช้จ่าย และความเป็น Open Source ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ Python มีความสามารถสูงขึ้น และใช้งานได้ครอบคลุมกับทุกลักษณะงาน

### 2.3.1 คุณลักษณะเด่นของภาษา Python

- 1) สนับสนุนแนวแบบคิดออบเจกต์โอเรียนเตด หรือ OOP (Object Oriented Programming)
- 2) เป็น Open Source
- 3) โค้ดที่เขียนด้วย Python สามารถนำไปรันบนระบบปฏิบัติการได้หลากหลาย
- 4) สนับสนุนเทคโนโลยี COM ของ Ms-windows
- 5) Python รวมมาตรฐานการอินเทอร์เฟซ Tkinter ซึ่งสนับสนุนบนระบบ X windows, Ms-windows และ Macintosh การใช้คำสั่ง Tkinter API ช่วยให้โปรแกรมเมอร์ไม่ต้องแก้ไขโค้ดเมื่อนำไปรันบนระบบปฏิบัติการอื่นๆ
- 6) เป็น Dynamic typing คือ สามารถเปลี่ยนชนิดข้อมูลได้ง่ายและสะดวก
- 7) มี Built-in Object Types คือ โครงสร้างของข้อมูลที่สามารถใช้ได้ภายใน Python ประกอบด้วย ลิสต์, ดิกชันนารี, สตริง ที่ง่ายต่อการใช้งานและมีประสิทธิภาพสูง
- 8) มีเครื่องมือต่างๆ มากมาย เช่น การประมวลผลเท็กซ์ไฟล์ การเรียงข้อมูล การเชื่อมต่อสตริง การตรวจสอบเงื่อนไขของข้อความ การแทนค่า เป็นต้น
- 9) มีโมดูลสำหรับการจัดการ Regular Expression
- 10) มีโมดูลที่สร้างขึ้นจากนักพัฒนาสนับสนุนมากมาย ได้แก่ COM, Image, CORBA, ORBs, XML เป็นต้น
- 11) จัดการหน่วยความจำอย่างอัตโนมัติ สามารถจัดการพื้นที่หน่วยความจำที่ไม่ต่อเนื่องให้ทำงานได้อย่างมีประสิทธิภาพ
- 12) อนุญาตให้ฝั่งชุดคำสั่งของ Python เอาไว้ภายในโค้ดภาษา C/C++ ได้
- 13) อนุญาตให้โปรแกรมเมอร์สร้าง Dynamic Link Library (DLL) เพื่อใช้ร่วมกับ Python

- 14) มีโมดูลสนับสนุนเกี่ยวกับเน็ตเวิร์ก โปรเซส เทรด regular, expression, xml, GUI และอื่นๆ
- 15) ประกอบด้วยโมดูลสำหรับสร้าง Internet Script และติดต่อกับอินเทอร์เน็ตผ่าน Sockets, และทำหน้าที่เป็น CGI Script ตลอดจนใช้งานคำสั่ง FTP , Gopher, XML และอื่นๆอีกมาก
- 16) สามารถประมวลผลทางด้านวิทยาศาสตร์ และ วิศวกรรมศาสตร์ได้อย่างมีประสิทธิภาพ
- 17) มีฟังก์ชันสนับสนุนฐานข้อมูล เช่น MySQL, Sybase, Oracle , Informix, ODBC และอื่นๆ
- 18) มีไลบรารีสนับสนุนด้านการสร้างภาพกราฟฟิก เช่น ทำภาพเบลอ หรือภาพชัด หรือเขียนข้อความบนภาพ ตลอดจนบันทึกไฟล์ในรูปแบบต่างๆ ได้อย่างสะดวก และมีประสิทธิภาพ
- 19) มีไลบรารีสนับสนุนด้านปัญญาประดิษฐ์
- 20) มีไลบรารีสำหรับสร้างเอกสาร PDF โดยไม่ต้องติดตั้ง Acrobat Writer
- 21) มีไลบรารีสำหรับสร้าง Shockwaves Flash (SWF) โดยไม่ต้องติดตั้ง Macromedia Flash
- 22) เนื่องจากไพทอนเป็นภาษากาว (Glue Language) สามารถเรียกใช้ภาษาโปรแกรมอื่นๆ ได้หลายภาษา ทำให้เหมาะที่จะใช้เขียนเพื่อประสานงานโปรแกรมที่เขียนในภาษาต่างกันได้
- 23) ไวยากรณ์ของไพทอนได้กำจัดการใช้สัญลักษณ์ที่ใช้ในการแบ่งบล็อกของโปรแกรม และใช้การย่อหน้าแทน ทำให้สามารถอ่านโปรแกรมที่เขียนได้ง่าย นอกจากนี้ยังมีการสนับสนุนการเขียน docstring ซึ่งเป็นข้อความสั้นๆ ที่ใช้อธิบายการทำงานของฟังก์ชัน, คลาส, และโมดูลอีกด้วย
- 24) เนื่องจากไพทอนเป็นภาษาสคริปต์ ทำให้ใช้เวลาในการเขียนและคอมไพล์ไม่มาก ทำให้เหมาะกับงานด้านการดูแลระบบ (System administration) เป็นอย่างยิ่ง ได้มีการสนับสนุนภาษาไพทอนโดยเป็นส่วนหนึ่งของระบบปฏิบัติการยูนิกซ์, ลินุกซ์ และสามารถติดตั้งให้ทำงานเป็นภาษาสคริปต์ของวินโดวส์ ผ่านระบบ en:Windows Script Host ได้อีกด้วย

### 2.3.2 องค์การสำคัญที่ใช้ไพทอน

google ใช้ภาษาไพทอนในบริการหลายอย่าง เช่น gmail google map

## บทที่ 3

### การออกแบบและพัฒนา

#### 3.1 สิ่งที่ต้องการจากระบบ

##### 3.1.1 การอัปโหลดเอกสาร

###### 3.1.1.1 Input / Output Specification

- 1) Input Specification
  - เอกสารที่ผู้ใช้งานต้องการอัปโหลดเข้าไปในระบบ
  - keyword ของเอกสาร
- 2) Output Specification
  - แสดงข้อความตอบกลับเมื่อผู้ใช้งานอัปโหลดเอกสารเสร็จสิ้น

###### 3.1.1.2 Functional

- 1) สามารถอัปโหลดเอกสารขึ้นไปบนระบบ พร้อมทั้งสามารถระบุ keyword ของเอกสาร
- 2) เก็บ keyword และเอกสารลงในฐานข้อมูล

##### 3.1.2 การค้นหาเอกสาร

###### 3.1.2.1 Input / Output Specification

- 1) Input Specification
  - Keyword ที่ใช้ในการค้นหา
  - รูปแบบการจัดเรียงผลลัพธ์ ซึ่งแบ่งได้เป็น จัดเรียงตามวันที่, จัดเรียงตามความเกี่ยวข้อง
- 2) Output Specification
  - แสดงรายการเอกสารที่มีความเกี่ยวข้องกับ Keyword

###### 3.1.2.2 Functional

- 1) สามารถค้นหาและแสดงเอกสารที่เกี่ยวข้องต้องการด้วยการใส่ Keyword
- 2) แสดงผลลัพธ์การค้นหา โดยการจัดเรียงตามวันที่หรือจัดเรียงตามความเกี่ยวข้อง

### 3.1.2 การลบเอกสาร

#### 3.1.2.1 Input / Output Specification

- 1) Input Specification
  - เอกสารที่ต้องการลบ(เฉพาะที่ user เป็นผู้ upload เท่านั้น)
- 2) Output Specification
  - ลบเอกสารในฐานข้อมูล

#### 3.1.2.2 Functional

- 1) สามารถลบเอกสารที่ user เป็นผู้ upload ได้

### 3.2 ภาพรวมของระบบ

โครงการนี้ได้ทำการออกแบบโปรแกรมให้ทำงานบนเครื่องคอมพิวเตอร์มีหน้าที่ในการติดต่อกับผู้ใช้และฐานข้อมูล โดยระบบจะแบ่งการทำงานออกเป็น 3 ส่วนหลักๆ คือการอัปโหลดเอกสาร การค้นหาเอกสาร และ การลบเอกสารที่ user เป็นผู้ upload

#### 1) การอัปโหลดเอกสาร

ผู้ใช้งานเลือกเอกสารที่ต้องการจะอัปโหลดเข้าสู่ระบบ โปรแกรมจะแสดง keyword ที่มีอยู่เดิมในระบบพร้อมกับให้ผู้ใช้งานเลือก keyword ที่ต้องการจะเพิ่มลงในฐานข้อมูล เมื่อเสร็จโปรแกรมจะใส่เอกสารและ keyword ใหม่ลงไปในฐานข้อมูล

#### 2) การค้นหาเอกสาร

ผู้ใช้งานพิมพ์ keyword ที่ต้องการลงในโปรแกรม โปรแกรมจะนำ keyword ไปประมวลผล และจากนั้นจะตอบกลับผู้ใช้งานเป็นผลลัพธ์เอกสารที่เกี่ยวข้อง ซึ่งจะจัดเรียงตามรูปแบบที่ผู้ใช้ได้กำหนดไว้

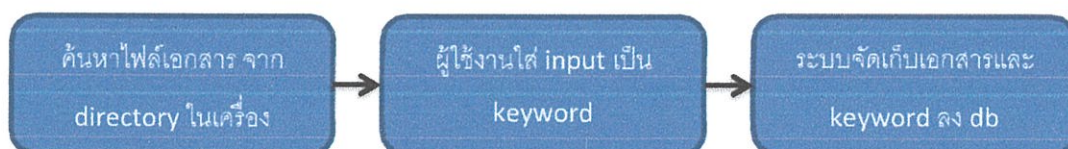
#### 3) การลบเอกสาร

ผู้ใช้งานสามารถลบเอกสารที่ตนเองเป็นผู้ upload ได้

### 3.3 รายละเอียดการทำงานของระบบ

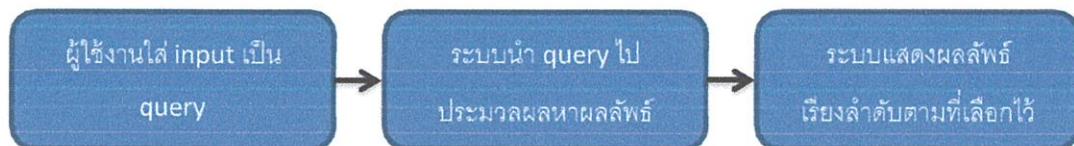
#### 3.3.1 Data Flow Diagram

##### 3.3.1.1 ส่วนของการอัปโหลดเอกสาร



รูปที่ 3.1 โค้ดแกรมแสดงขั้นตอนการอัปโหลดเอกสาร

### 3.3.1.2 ส่วนของการค้นหาเอกสาร



รูปที่ 3.2 ไดอะแกรมแสดงขั้นตอนการค้นหาเอกสาร

### 3.4 ตารางเก็บข้อมูล

ในการสร้างตารางเก็บข้อมูลจะใช้หลักการ 5NF โดย 5NF มีหลักการอยู่ว่า ทุกๆ JD (JD = Join Dependency คือ ความสามารถในการ split ได้ของตาราง) ใน R จะต้องเป็น consequence ของ candidate key ของ R หรือก็คือ ถ้าหาก R split ได้จะต้องมี candidate key ของ R อยู่ในตารางย่อยเสมอ

ตารางที่เป็น 5NF แล้วจะไม่มีปัญหาในการ insert update delete

ตาราง keyword มี keyid เป็น primary key มี attribute keyname บอกว่าเป็นคำว่าอะไร

ตาราง 3.1 ตาราง Keyword

keyid	keyname
001	Keyword1
002	Keyword2

ตาราง document มี DocumentID (docid) เป็น primary key มี attribute docname บอกชื่อเอกสาร date บอกวันที่อัปโหลด uname บอกว่าใครเป็นคนอัปโหลด confidential บอกถึงระดับการปกปิดเป็นความลับของเอกสาร

ตาราง 3.2 ตาราง Document

docid	docname	uname	date	confidential
001	Document1	Uname1	Date1	3
002	Document2	Uname2	Date2	4

ตาราง document & keyword มี docid และ keyid ประกอบกันเป็น combine key ตารางนี้บอก  
ว่าในเอกสารไหนมี keyword อะไรอยู่บ้าง และ f บอกความถี่

**ตาราง 3.3 ตาราง Document & Keyword**

docid	keyid	f
001	001	1
001	002	2
002	002	3

จากตาราง document & keyword ด้านบนหมายความว่าใน document 001 มีทั้ง keyword 001  
และ 002 ส่วน document 002 มีแค่ keyword 002 เท่านั้น

ตาราง account มี Username เป็น primary key มี attribute Password บอกถึงรหัสผ่านของ  
Username นี้ Name-Lastname บอกชื่อ-นามสกุลของผู้ใช้งาน Privilege บอกระดับสิทธิ์ในการ  
เข้าถึงเอกสาร (สามารถเข้าถึงเอกสารที่มี Confidential  $\leq$  Privilege)

**ตาราง 3.4 ตาราง Account**

Username	Passwd	Name	Lastname	Privilege
User001	Passwd1	Name1	Lastname1	5
User002	Passwd2	Name2	Lastname2	4

ตาราง Fuzzy เก็บค่าความสัมพันธ์ของแต่ละ keyword มี keyid1 และ keyid2 ประกอบกันเป็น  
combine key

**ตาราง 3.5 ตาราง Fuzzy**

Keyid1	Keyid2	c
K001	K002	0.5
K001	K003	0.25

## บทที่ 4

# วิธีการดำเนินงาน

### 4.1 การพัฒนาโปรแกรม

หลังจากที่ได้ออกแบบระบบมาแล้ว ผู้จัดทำได้เริ่มการพัฒนาโปรแกรมโดย มีรายละเอียดดังนี้

#### 4.1.1 Class

1) Main

คลาสนี้เป็นคลาสหลักที่สร้าง GUI และเรียกใช้งานคลาสอื่นๆ

2) Login

เป็นคลาสที่มีฟังก์ชันในการจัดการเกี่ยวกับการเข้าระบบ Check Username และ Password ของ User ว่าตรงกับที่มีอยู่ใน Database หรือไม่

3) Search

เป็นคลาสที่มีฟังก์ชันในการจัดการเกี่ยวกับการค้นหาเอกสาร โดยจัดเรียงลำดับตามวันที่หรือตามความตรงประเด็นของเอกสาร

4) MyAccount

เป็นคลาสที่มีฟังก์ชันในการจัดการเกี่ยวกับ account ของ user สามารถลบเอกสารที่เป็นผู้ upload ออกจากระบบได้

5) Database

เป็นคลาสที่มีฟังก์ชันในการจัดการและการเชื่อมต่อกับ database

6) PDFprocess

เป็นคลาสที่ใช้สำหรับการจัดการเอกสารนามสกุลไฟล์ PDF ในการแปลงไฟล์ PDF เป็น String และค้นหา Keyword ภายในเอกสาร โดยคลาสนี้จะประกอบไปด้วย 2 คลาสย่อยคือ

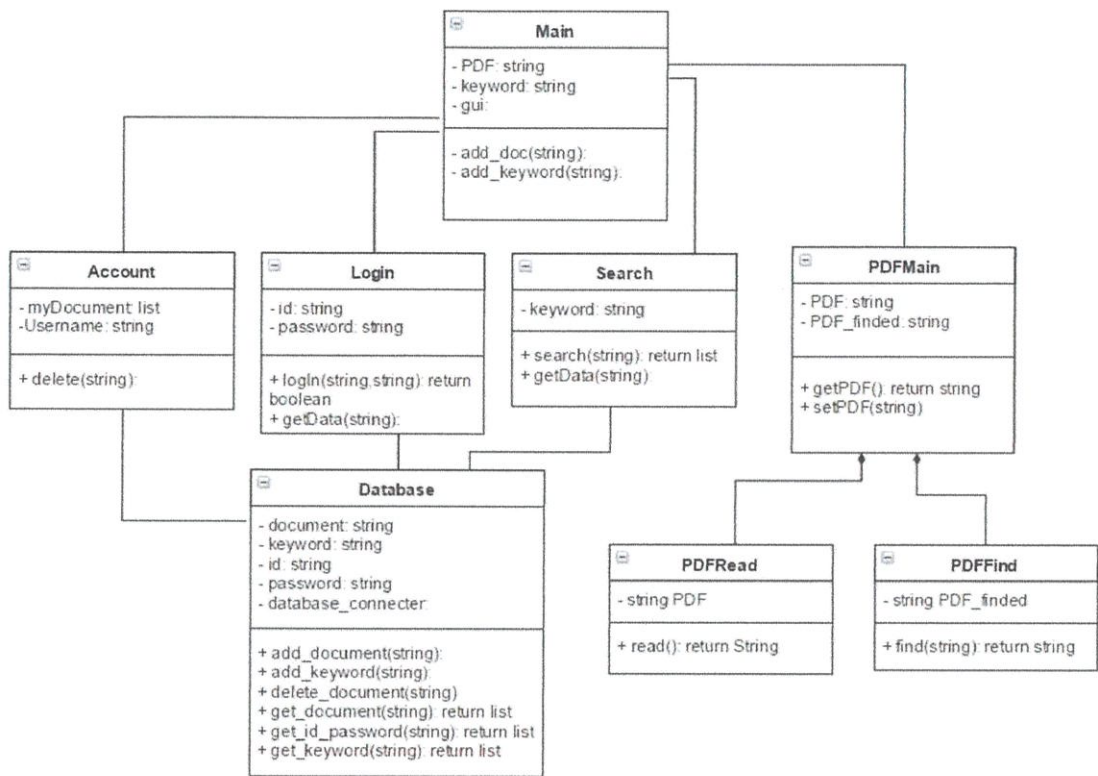
7) PDFreader

ใช้ในการแปลงข้อความในเอกสารเป็น string

8) PDFfinder

ใช้ในการค้นหา keyword และแสดงตำแหน่งของ keyword ภายในเอกสาร

## 4.1.2 Class Diagram

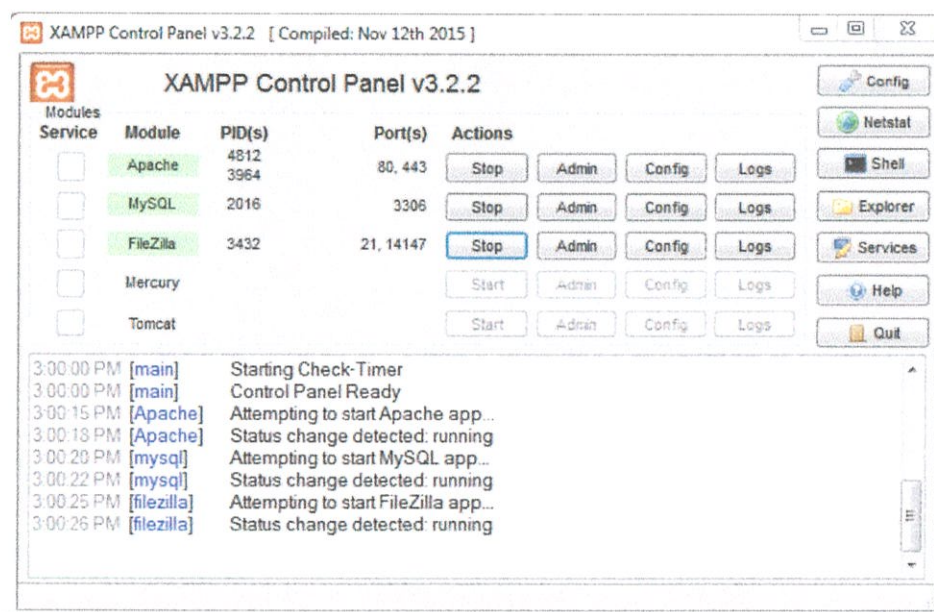


រូបថត 4.1 Class Diagram

## 4.2 เทคโนโลยีที่ใช้

### 4.2.1 XAMPP

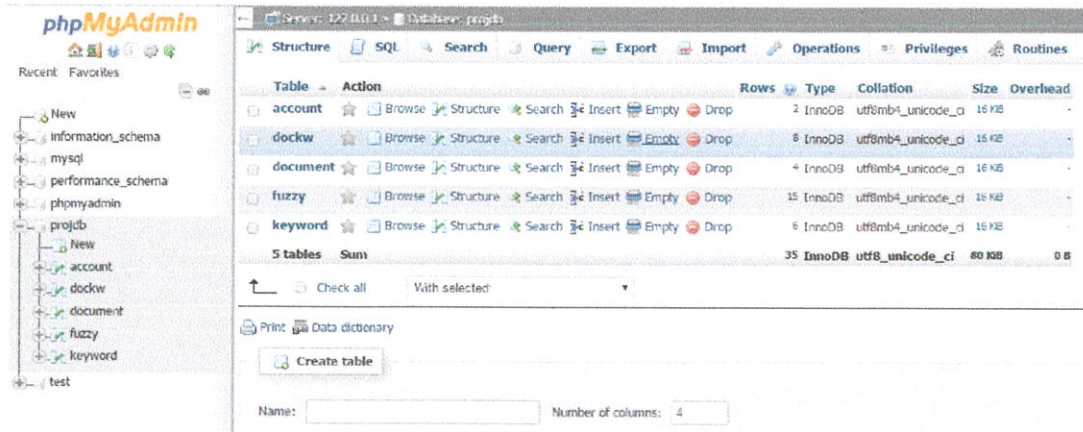
Xampp คือโปรแกรม Apache web server ไว้จำลอง web server เพื่อไว้ทดสอบ สคริปต์ หรือเว็บไซต์ในเครื่องของเรา โดยที่ไม่ต้องเชื่อมต่ออินเทอร์เน็ตและไม่ต้องมีค่าใช้จ่ายใดๆ ง่ายต่อการติดตั้งและใช้งาน โปรแกรม Xampp จะมาพร้อมกับ PHP ภาษาสำหรับพัฒนาเว็บแอปพลิเคชันที่เป็นที่นิยม , MySQL ฐานข้อมูล, Apache จะทำหน้าที่เป็นเซิร์ฟเวอร์, Perl อีกทั้งยังมาพร้อมกับ OpenSSL , phpMyadmin (ระบบบริหารฐานข้อมูลที่พัฒนาโดย PHP เพื่อใช้เชื่อมต่อไปยังฐานข้อมูล สนับสนุนฐานข้อมูล MySQL และ SQLite โปรแกรม Xampp จะอยู่ในรูปแบบของไฟล์ Zip, tar, 7z หรือ exe โปรแกรม Xampp อยู่ภายใต้ใบอนุญาตของ GNU General Public License แต่บางครั้งอาจจะมีการเปลี่ยนแปลงเรื่องของลิขสิทธิ์ในการใช้งาน



รูปที่ 4.2 XAMPP

### 4.2.2 phpMyAdmin

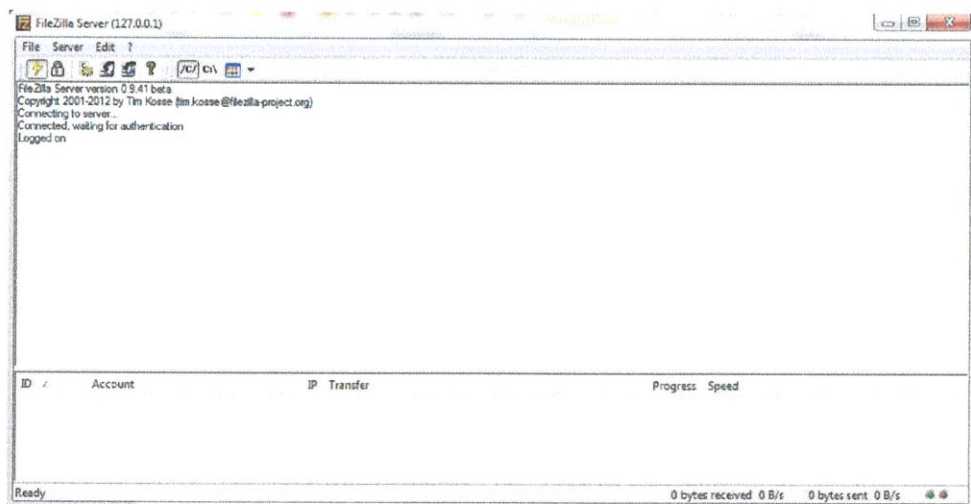
phpMyAdmin คือ โปรแกรมที่ถูกพัฒนาโดยใช้ภาษา PHP เพื่อใช้ในการบริหารจัดการฐานข้อมูล Mysql แทนการคีย์คำสั่ง เนื่องจากถ้าเราจะใช้ฐานข้อมูลที่เป็น MySQL บางครั้งจะมีความลำบากและยุ่งยากในการใช้งาน ดังนั้นจึงมีเครื่องมือในการจัดการฐานข้อมูล MySQL ขึ้นมาเพื่อให้สามารถจัดการ ตัวDBMS ที่เป็น MySQL ได้ง่ายและสะดวกยิ่งขึ้น โดย phpMyAdmin ก็ถือเป็นเครื่องมือชนิดหนึ่งในการจัดการนั่นเอง



รูปที่ 4.3 phpMyAdmin

#### 4.2.3 FileZilla

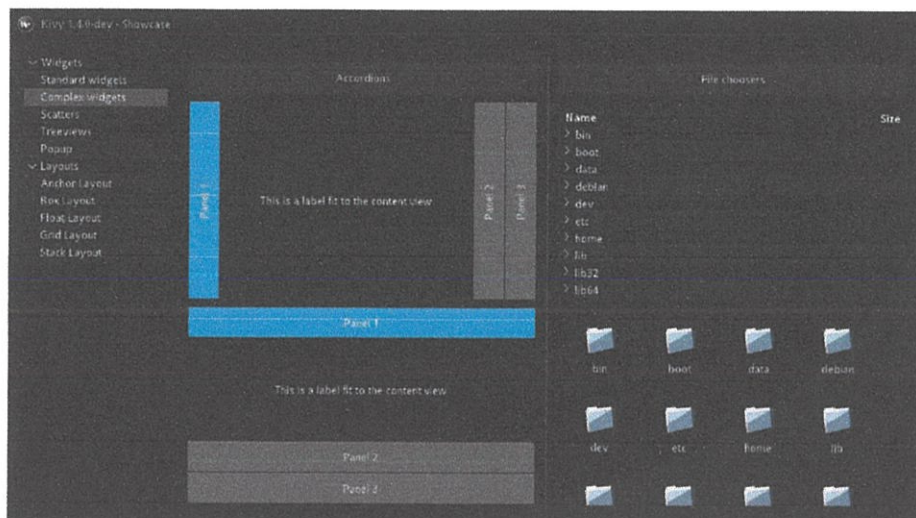
โปรแกรม FileZilla Server คือ โปรแกรมที่ไว้สำหรับจำลองเครื่องของเราให้เป็นเครื่อง ftp server ไว้สำหรับเก็บไฟล์หรือทำงานเป็นเครื่อง server



รูปที่ 4.4 FileZilla

#### 4.2.4 Kivy

Kivy เป็น framework สำหรับใช้สร้าง GUI โดยทำงานได้ทั้ง Windows, Linux, OS X, Android และ iOS (ครอบจักรวาลเลย) และมี Kivy language (เหมือนกับ QML ใน PyQt/PySide) ให้ใช้งานออกแบบโดยใช้ MIT License และยังรองรับทั้ง Python 2 และ Python 3 โดยจุดเด่นของ Kivy คือความสะดวกในการใช้งานและสามารถนำไปใช้งานได้หลาย Platform



รูปที่ 4.5 kivy

#### 4.2.4 PyPDF2

PyPDF2 เป็น Library ตัวหนึ่งของ Python ที่ใช้ในการแปลงข้อความที่ได้จากเอกสาร PDF ให้อยู่ในรูปแบบของ string เพื่อที่สามารถนำค่า string ที่ได้ไปใช้งานต่อ

#### 4.2.6 MySQL

MySQL คือ โปรแกรมระบบจัดการฐานข้อมูล ที่พัฒนาโดยบริษัท MySQL AB มีหน้าที่เก็บข้อมูลอย่างเป็นระบบ รองรับคำสั่ง SQL เป็นเครื่องมือสำหรับเก็บข้อมูล ที่ต้องใช้ร่วมกับเครื่องมือหรือโปรแกรมอื่นอย่างบูรณาการ เพื่อให้ได้ระบบงานที่รองรับ ความต้องการของผู้ใช้ เช่นทำงานร่วมกับเครื่องบริการเว็บ (Web Server) เพื่อให้บริการแก่ภาษาสคริปต์ที่ทำงานฝั่งเครื่องบริการ (Server-Side Script) เช่น ภาษา php ภาษา asp.net หรือภาษาเจเอสพี เป็นต้น หรือทำงานร่วมกับโปรแกรมประยุกต์ (Application Program) เช่น ภาษาวิซวลเบสิก คอทเน็ต ภาษาจาวา หรือภาษาซีชาร์ป เป็นต้น โปรแกรมถูกออกแบบให้สามารถทำงานได้บนระบบปฏิบัติการที่หลากหลาย และเป็นระบบฐานข้อมูล โอเพนซอร์ซ (Open Source) ที่ถูกนำไปใช้งานมากที่สุด

MySQL : มาเอสคิวแอล เป็นระบบจัดการฐานข้อมูลโดยใช้ภาษา SQL. แม้ว่า MySQL เป็นซอฟต์แวร์โอเพนซอร์ส แต่แตกต่างจากซอฟต์แวร์โอเพนซอร์สทั่วไป โดยมีการพัฒนาภายใต้บริษัท MySQL AB ในประเทศสวีเดน โดยจัดการ MySQL ทั้งในแบบที่ให้ใช้ฟรี และแบบที่ใช้ในเชิงธุรกิจ

MySQL สร้างขึ้นโดยชาวสวีเดน 2 คน และชาวฟินแลนด์ ชื่อ David Axmark, Allan Larsson และ Michael "Monty" Widenius. ปัจจุบันบริษัทซันไมโครซิสเต็มส์ (Sun Microsystems, Inc.) เข้าซื้อกิจการของ MySQL AB เรียบร้อยแล้ว ฉะนั้นผลิตภัณฑ์ภายใต้ MySQL AB ทั้งหมดจะตกเป็นของซัน

ชื่อ "MySQL" อ่านออกเสียงว่า "มายเอสคิวเอล" หรือ "มายเอสคิวเอล" (ในการอ่านอักษร L ในภาษาไทย) ซึ่งทางซอฟต์แวร์ไม่ได้อ่าน มายซีเควล หรือ มายซีควอล เหมือนกับซอฟต์แวร์จัดการฐานข้อมูลตัวอื่น

ความสามารถและการทำงานของโปรแกรม MySQL มีดังต่อไปนี้

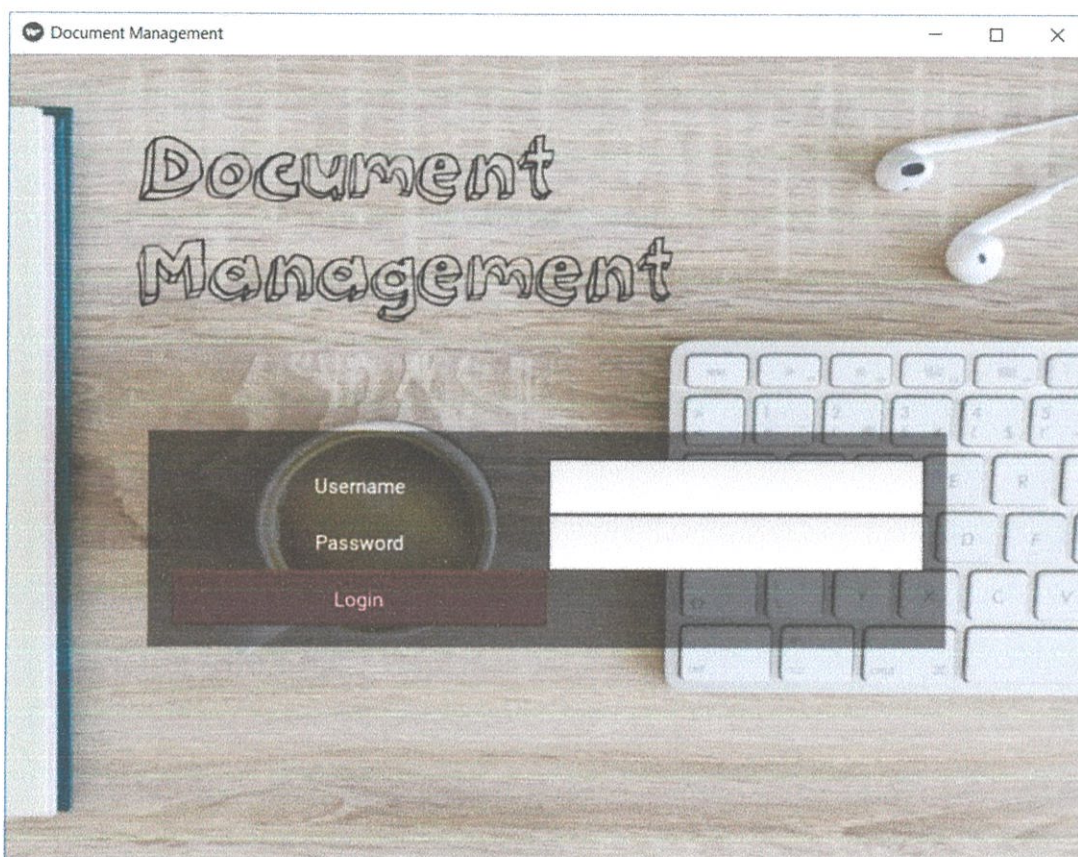
- 1) MySQL ถือเป็นระบบจัดการฐานข้อมูล (DataBase Management System (DBMS) ฐานข้อมูลมีลักษณะเป็นโครงสร้างของการเก็บรวบรวมข้อมูล การที่จะเพิ่มเติมเข้าถึงหรือประมวลผลข้อมูลที่เก็บในฐานข้อมูลจำเป็นต้องอาศัยระบบจัดการฐานข้อมูล ซึ่งจะทำหน้าที่เป็นตัวกลางในการจัดการกับข้อมูลในฐานข้อมูลทั้งสำหรับการ ใช้งานเฉพาะ และรองรับการทำงานของแอปพลิเคชันอื่นๆ ที่ต้องการใช้งานข้อมูลในฐานข้อมูล เพื่อให้ได้รับความสะดวกในการจัดการกับข้อมูลจำนวนมาก MySQL ทำหน้าที่เป็นทั้งตัวฐานข้อมูลและระบบจัดการฐานข้อมูล
- 2) MySQL เป็นระบบจัดการฐานข้อมูลแบบ relational ฐานข้อมูลแบบ relational จะทำการเก็บข้อมูลทั้งหมดในรูปแบบของตารางแทน การเก็บข้อมูลทั้งหมดลงในไฟล์ เพียงไฟล์เดียว ทำให้ทำงานได้รวดเร็วและมีความยืดหยุ่น นอกจากนั้น แต่ละตารางที่เก็บข้อมูลสามารถเชื่อมโยงเข้าหากันทำให้สามารถรวมหรือจัด กลุ่มข้อมูลได้ตามต้องการ โดยอาศัยภาษา SQL ที่เป็นส่วนหนึ่งของโปรแกรม MySQL ซึ่งเป็นภาษามาตรฐานในการเข้าถึงฐานข้อมูล
- 3) MySQL แจกจ่ายให้ใช้งานแบบ Open Source นั่นคือ ผู้ใช้งาน MySQL ทุกคนสามารถใช้งานและปรับแต่งการทำงานได้ตามต้องการ สามารถดาวน์โหลดโปรแกรม MySQL ได้จากอินเทอร์เน็ตและนำมาใช้งาน โดยไม่มีค่าใช้จ่ายใดๆ

ในระบบปฏิบัติการ Red Hat Linux นั้น มีโปรแกรมที่สามารถใช้งานเป็นฐานข้อมูลให้ผู้ใช้และระบบสามารถเลือกใช้งานได้ หลายโปรแกรม เช่น MySQL และ PostgreSQL ผู้ดูแลระบบสามารถเลือกติดตั้งได้ทั้งในขณะติดตั้งระบบปฏิบัติการ Red Hat Linux หรือจะติดตั้งภายหลังจากที่ติดตั้งระบบปฏิบัติการก็ได้ อย่างไรก็ตาม สาเหตุที่ผู้ใช้งานจำนวนมากนิยมใช้งานโปรแกรม MySQL คือ MySQL สามารถทำงานได้อย่างรวดเร็ว น่าเชื่อถือและใช้งานได้ง่าย เมื่อเปรียบเทียบประสิทธิภาพในการทำงานระหว่างโปรแกรม MySQL และ PostgreSQL

โดยพิจารณาจากการประมวลผลแต่ละคำสั่งได้ผลลัพธ์ดังรูปที่ 1 นอกจากนั้น MySQL ถูกออกแบบและพัฒนาขึ้นมาเพื่อทำหน้าที่เป็นเครื่องให้บริการรองรับการจัดการกับฐานข้อมูลขนาดใหญ่ ซึ่งการพัฒนายังคงดำเนินอยู่อย่างต่อเนื่อง ส่งผลให้มีฟังก์ชันการทำงานใหม่ๆ ที่อำนวยความสะดวกแก่ผู้ใช้งานเพิ่มขึ้นอยู่ตลอดเวลา รวมไปถึงการปรับปรุงด้านความต่อเนื่อง ความเร็วในการทำงาน และความปลอดภัย ทำให้ MySQL เหมาะสมต่อการนำไปใช้งานเพื่อเข้าถึงฐานข้อมูลบนเครือข่ายอินเทอร์เน็ต

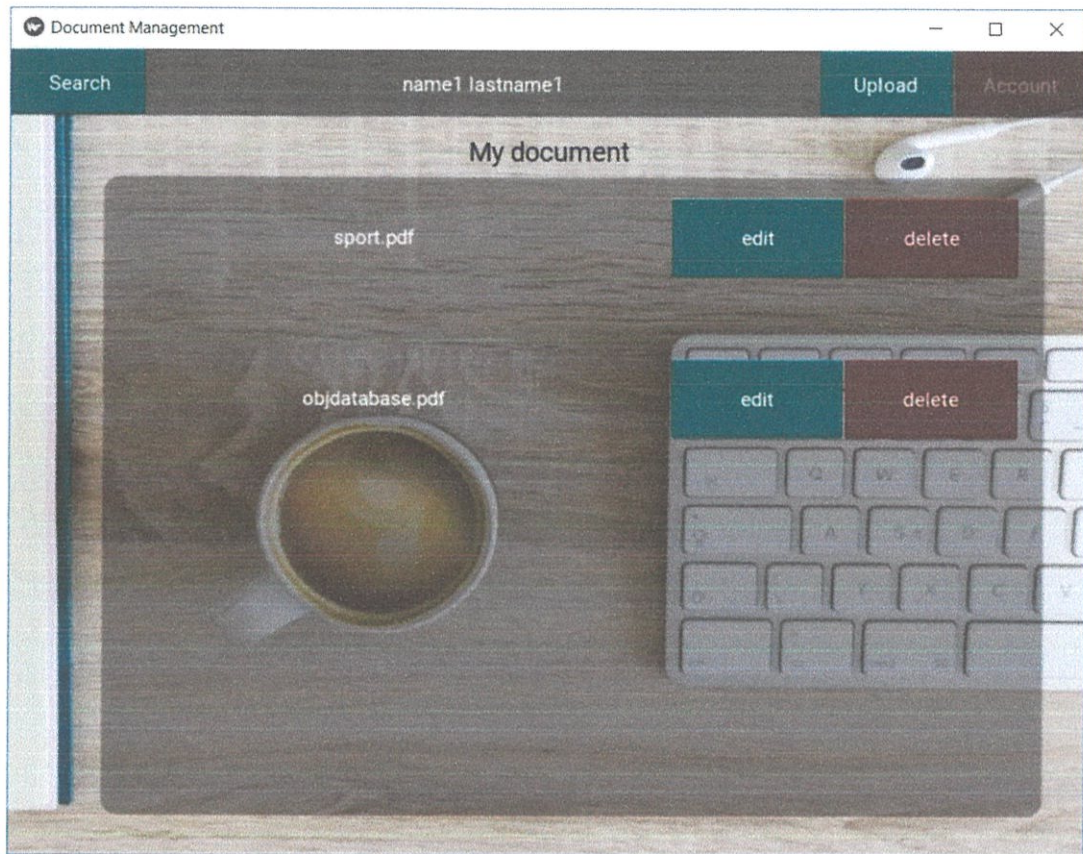
### 4.3 ผลการดำเนินงาน

#### 1) Login



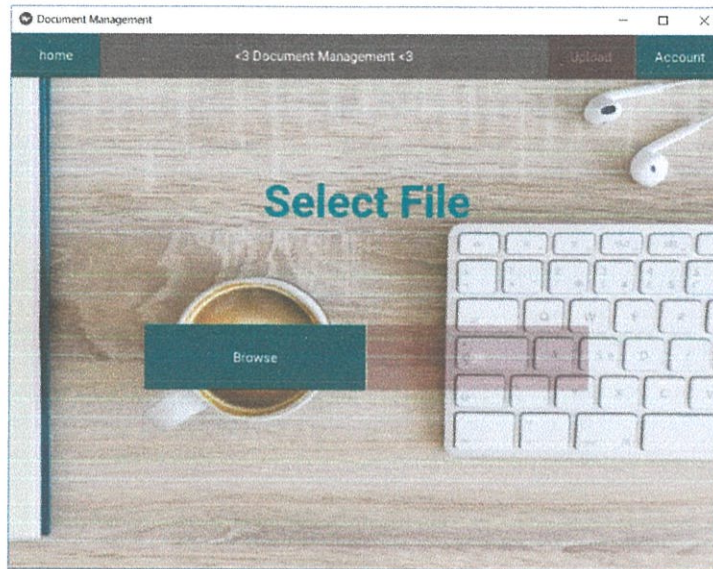
รูปที่ 4.6 Login

## 2) MyAccount



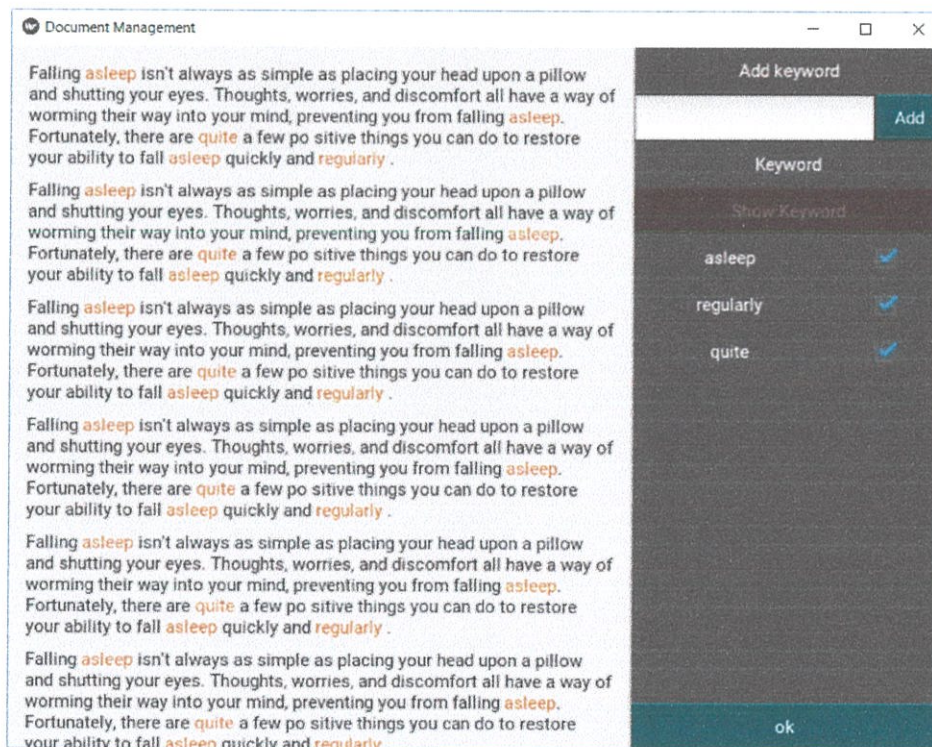
รูปที่ 4.7 MyAccount

## 3) Browse



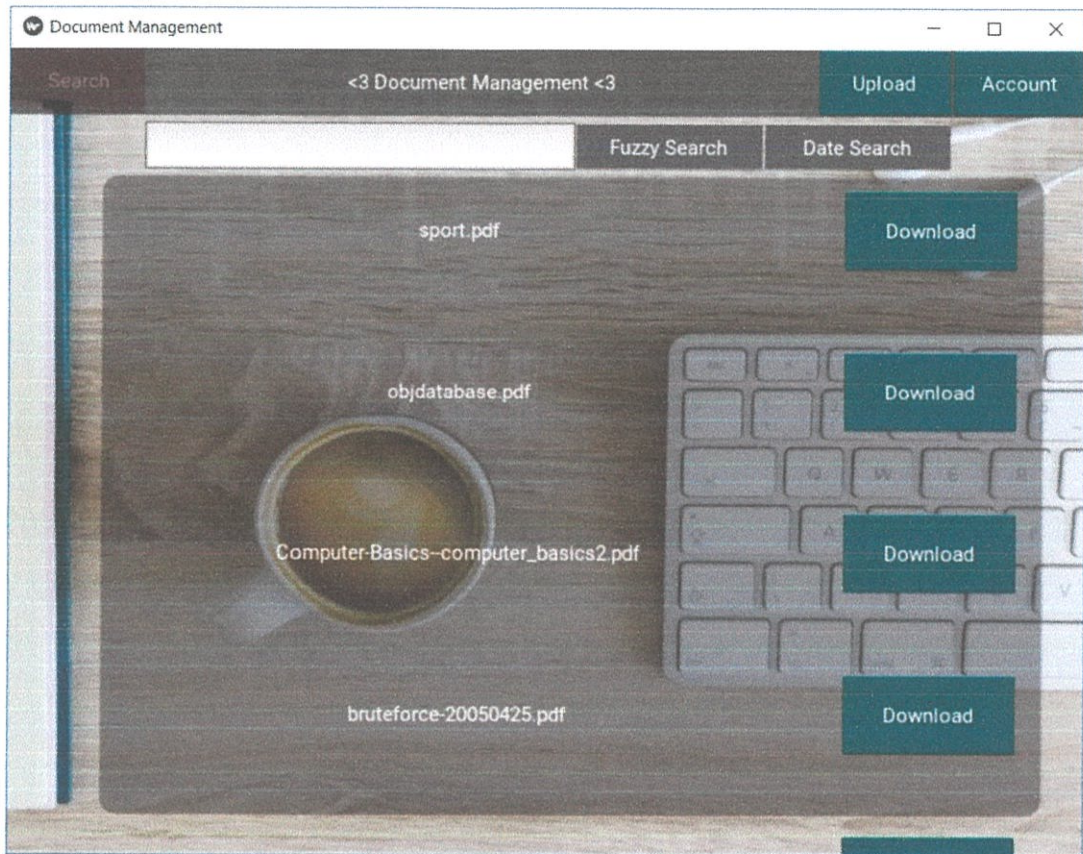
รูปที่ 4.8 Browse

## 4) Upload



รูปที่ 4.9 Upload

## 5) Search



รูปที่ 4.10 Search

# บทที่ 5

## บทสรุป

### 1.สรุปผล

โปรแกรมจัดการเอกสาร เป็นโปรแกรมที่ถูกพัฒนาขึ้นด้วยภาษาไพธอน ส่วนของโปรแกรมช่วยการจัดการกับเอกสารเพื่อให้สะดวกต่อการจัดเก็บ เผยแพร่ และค้นหา มีการใช้แนวคิด Fuzzy logic มาคำนวณหาค่าความตรงประเด็นของเอกสารในการค้นหาเพื่อให้ตรงกับความต้องการของผู้ใช้งาน มีการยืนยันตัวตนและมอบสิทธิ์ผู้ใช้งาน

### 2.ปัญหาที่เกิดขึ้น

- 1) ในการพัฒนาโปรแกรม ผู้จัดทำต้องศึกษาเทคโนโลยีที่ไม่เคยใช้เพิ่มเติมทำให้เกิดปัญหาและใช้ เวลาในการทำความเข้าใจ แต่เป็นผลดีที่ผู้จัดทำได้รับความรู้ความสามารถและประสบการณ์เพิ่ม
- 2) library ของภาษา python ใช้งานยาก

## บรรณานุกรม

อาจารย์ บัณฑิต พัสชา. **Information Storage and Retrieval**. [สไลด์]. กรุงเทพฯ :  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

Th Wikipedia. ภาษาไพทอน. [Online].

Available: <https://th.wikipedia.org/wiki/ภาษาไพทอน>

mindphp. Python คืออะไร. [Online].

Available: <http://www.mindphp.com/คู่มือ/73-คืออะไร/2417-python-คืออะไร.html>

รศ.ดร.ศุภมิตร จิตตะยโสธร. “**Database Design and SQL.**” กรุงเทพฯ : สถาบันเทคโนโลยีพระ  
จอมเกล้าเจ้าคุณทหาร ลาดกระบัง

stackoverflow. **stackoverflow**. [Online].

Available: <https://stackoverflow.com/>

## ภาคผนวก

### 1. การอ่านไฟล์โดยใช้ภาษา python

ในการอ่านไฟล์ PDF เราใช้ Library PDFminer ในการอ่านและแปลงตัวอักษรในไฟล์ PDF มาเป็นข้อมูลประเภท String เพื่อนำไปใช้ในงานในฟังก์ชันต่อไป

#### โปรแกรม 1 การอ่านไฟล์โดยใช้ภาษา python

```

from pdfminer.pdfinterp import PDFResourceManager,
PDFPageInterpreter
from pdfminer.converter import TextConverter
from pdfminer.layout import LAParams
from pdfminer.pdfpage import PDFPage
from io import StringIO

rsrcmgr = PDFResourceManager()
retstr = StringIO()
codec = 'utf-8'
laparams = LAParams()
device = TextConverter(rsrcmgr, retstr, codec=codec,
laparams=laparams)
fp = open(self.locations, 'rb')
interpreter = PDFPageInterpreter(rsrcmgr, device)
password = ""
maxpages = max
caching = True
pagenos=set()

for page in PDFPage.get_pages(fp, pagenos,
maxpages=maxpages, password=password,caching=caching,
check_extractable=True):
    interpreter.process_page(page)

text = retstr.getvalue()
fp.close()
device.close()
retstr.close()

```

จากโค้ดข้างต้นมีการ import คลาสที่จำเป็นจาก Library PDFminer และ io พร้อมกับสร้างตัวแปรมารองรับ

ใช้ฟังก์ชัน open เพื่อเปิดไฟล์ PDF โดย self.location คือ Directory ของไฟล์และกำหนดประเภทของการเปิดไฟล์เป็น 'rb' จากนั้นกำหนดค่าต่างๆที่จำเป็น ใช้ for loop ในการอ่านเอกสารแต่ละหน้าและใช้ฟังก์ชัน getvalue() เพื่อรับค่าที่อ่านได้ลงในตัวแปร text  
 ขั้นตอนสุดท้ายคือการ close() เพื่อปิดการทำงาน

## 2.การอัปโหลดไฟล์โดยใช้ภาษา python

ในการอัปโหลดไฟล์ขึ้นบน Server ก่อนอื่นเราจำเป็นต้องมี File Server ของเราเสียก่อน จากนั้นใช้ Library FTP ของ python ในการเชื่อมต่อโปรแกรมกับ Server

### โปรแกรม 2 การอัปโหลดไฟล์โดยใช้ภาษา python

```
from ftplib import FTP
import os

def uploadfile(self,directory,filename):

    ftp = FTP('IP_FLIESERVER')
    ftp.login(user='Admin')
    fh = open(directory,'rb')
    ftp.storbinary('STOR '+filename, fh)
    ftp.quit()
```

ขั้นแรก Import FTP และสร้างตัวแปร ftp โดยใช้ IP fileserver ที่เรามี จากนั้น login ด้วย ID และ Password (กรณีนี้ File Server ไม่ได้กำหนด Password จึงไม่จำเป็นต้องใส่)

ขั้นตอนต่อมา open ไฟล์ที่ต้องการจะอัปโหลดขึ้นมา และใช้ฟังก์ชัน storbinary ในการอัปโหลดขึ้น File Server

ขั้นตอนสุดท้ายใช้คำสั่ง quit() เพื่อหยุดการเชื่อมต่อกับ File Server

### 3. การค้นหา keyword ในเอกสารโดยใช้ภาษา python

จะใช้ฟังก์ชัน find ของ python ฟังก์ชันดังกล่าวในที่นี้จะใช้สองพารามิเตอร์คือ(keyword,m)

โดยที่ keyword คือ keyword ที่เราต้องการค้นหาในเอกสาร m คือ ตำแหน่ง index ที่ต้องการจะเริ่มค้นหา

#### โปรแกรม 3 การค้นหา keyword ในเอกสารโดยใช้ภาษา python

```
stringa = 'hello hello myname ispython'
arraykw = ["hello", "my", "name", "is", "python"]
for i in range(len(arraykw)):
    m=0
    f=0
    while (stringa.find(arraykw[i],m) != -1):
        print (arraykw[i])
        n=stringa.find(arraykw[i],m)
        print ("n = ",n)
        m=n+len(arraykw[i])
        print ("m = ",m)
        f+=1
    if(stringa.find(arraykw[i],m) == -1):
        print ("f = ",f)
        f=0
```

จากโค้ดข้างต้น stringa คือเนื้อความของเอกสาร arraykw คือ keyword ทั้งหมดที่เราจะต้องการค้นหา

ขั้นแรกใน for loop เราจะทำจำนวนครั้งตามจำนวน keyword ที่เราต้องการจะค้นหาในเอกสาร

m คือ ตำแหน่ง index ที่ต้องการจะเริ่มค้นหา โปรแกรมจะทำการค้นหาจนกว่าจะไม่เจอ keyword นั้นๆในเอกสาร โปรแกรมได้มีการใส่คำสั่งให้ print keyword, ตำแหน่ง( n = ตำแหน่งเริ่ม m = ตำแหน่งจบ ) และ ความถี่ของ keyword ที่ปรากฏในเอกสารนั้นๆ