

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง  
โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน  
IMAGE SEARCH ENGINE



ปริญญาบัตรนี้เป็นส่วนหนึ่งของการศึกษาศาสนาหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต  
สาขาวิชาวิศวกรรมคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ปีการศึกษา 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน**  
**IMAGE SEARCH ENGINE**



**ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาคามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต**  
**สาขาวิชาวิศวกรรมคอมพิวเตอร์**  
**คณะวิศวกรรมศาสตร์**  
**สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง**  
**ปีการศึกษา 2549**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2549

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน

Image Search Engine

ผู้จัดทำ

1.นายภาณุพงศ์ พันระไชย รหัสนักศึกษา 47015688

2.นายอิทธิพงษ์ นพฤทธิ์ รหัสนักศึกษา 47015700



Ornat Jinn.

อาจารย์ที่ปรึกษา

(ผศ.ดร.อรฉัตร จิตต์โสภักตร์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน

นายภาณุพงศ์ พันระไชย	47015688
นายอิทธิพงษ์ นพฤทธิ	47015700
ศศ.ดร.อรณัตร์ จิตต์โสภักตร์	อาจารย์ที่ปรึกษา
ปีการศึกษา 2549	

## บทคัดย่อ

เนื่องจากในปัจจุบันในระบบเครือข่ายอินเทอร์เน็ตมีเว็บเพจอยู่เป็นจำนวนมาก และในแต่ละเว็บเพจนั้นก็มียูเออาร์อยู่เป็นจำนวนมากเช่นกัน การค้นหารูปภาพในระบบเครือข่ายอินเทอร์เน็ตจึงมีความยากลำบากที่จะค้นหารูปภาพมาให้ได้ตามความต้องการ ดังนั้นในโครงการนี้ จึงพัฒนาเสิร์จเอนจินซึ่งทำหน้าที่ในการค้นหาและเก็บข้อมูลภาพที่มีอยู่ในอินเทอร์เน็ต พร้อมทั้งพัฒนาเว็บเพจเพื่อให้ผู้ใช้สามารถเข้ามาทำการค้นหาภาพและข้อมูลเว็บเพจที่อยู่ของภาพนั้น จากฐานข้อมูลที่เก็บไว้ในเซิร์ฟเวอร์ได้อีกด้วย

การทำงานของเสิร์จเอนจินแบ่งออกเป็น 3 ส่วน คือ โปรแกรมโรบอท โปรแกรมอินเด็กซ์เซอร์ และเสิร์จเอนจินซอฟต์แวร์ ส่วนของโรบอทจะมีหน้าที่ค้นหาลิงค์รูปภาพจากเว็บเพจภายในเครือข่ายอินเทอร์เน็ตแล้วนำมาเก็บในฐานข้อมูล โดยโรบอทจะมีการเช็คเวลาในการเก็บลิงค์แต่ละลิงค์ด้วยเพื่อแก้ปัญหาการวนลูปของโปรแกรม และเพื่อป้องกันปัญหาการวิ่งเก็บลิงค์ที่ซ้ำจึงมีการตั้งระยะเวลาของการเก็บลิงค์เพื่อให้โรบอท ทำงาน ได้อย่างมีประสิทธิภาพในขณะที่เดียวกันก็มีการโหลดรูปจากลิงค์ที่ได้มาจนครบและนำรูปมาเก็บในฐานข้อมูล ส่วนโปรแกรมอินเด็กซ์เซอร์เป็นส่วนที่ทำหน้าที่สร้างดัชนีค้นหาจากไฟล์ที่โรบอทหามา ส่วนสุดท้ายคือเว็บเพจการค้นหาภาพซึ่งจะทำการเปรียบเทียบรูปภาพโดยใช้หลักการหาค่าฮิสโตแกรมเพื่อหาลักษณะเฉพาะของรูปที่ได้มา แล้วค้นหารูปที่ตรงกับความต้องการของผู้ใช้ให้ได้มากที่สุด

ในการใช้งานเว็บเสิร์จเอนจินนั้นแบ่งออกเป็น 2 ส่วนคือ ส่วนของผู้ดูแลระบบ และส่วนของผู้ใช้งาน ส่วนของผู้ดูแลระบบมีการทำงานโดยสามารถกำหนดวันเวลาในการทำงานของโปรแกรมว่าต้องการระยะเวลาในการ โหลดเท่าใด สามารถแสดงข้อมูลในตารางฐานข้อมูล สามารถเพิ่มลิงค์ลงในฐานข้อมูล สามารถค้นหาลิงค์ของรูปภายในฐานข้อมูล และสามารถลบข้อมูลในฐานข้อมูลได้ ส่วนของผู้ใช้สามารถนำภาพที่มีอยู่ในเครื่องส่งให้กับเซิร์ฟเวอร์เพื่อค้นหารูปที่ผู้ใช้ต้องการ และสามารถเลือกได้ว่าต้องการรูปที่ใช่ หรือรูปที่คล้ายได้จากโปรแกรม จากนั้นเซิร์ฟเวอร์จะค้นหารูปและแสดงรูปทางหน้าเว็บที่ตรงตามความต้องการของผู้ใช้ให้ได้มากที่สุด

# Image Search Engine

Panupong Puntachai	47015688
Ithiphong Nopparit	47015700
Asst.Prof.Dr. Orachat Chitsobhuk	Advisor
Academic Year 2006	

## ABSTRACT

Nowadays, there are numerous web pages in the internet which each of them contains several pictures. As a result, it becomes difficult to find and seek that large amount of pictures in the internet as required by users. Therefore, this project develops a search engine which explores and collects picture information from the internet. In addition, search engine web page is developed in order to allow user to look for his/her desired picture and picture's information from the database in the search engine server.

The search engine system can be separated into three parts which are robot program, indexer program and search engine software. The Robot is responsible for searching and linking the pictures from the web pages in the internet. However, the Robot will check the result date and time in order to prevent looping of the program and running the duplicate links. This helps robot to effectively search all links and pictures and store in the database. The indexer program is responsible for creating the index from the file obtained from the Robot. The last part is the search engine web page which will compare the pictures using their histograms to find the most relevant pictures from user's request.

The search engine web can be separated into two parts which are the webmaster and the user. The webmaster takes care of controlling the Robot and indexer programs by setting running date and time. Moreover, the webmaster can search for the link and delete the data from the database if it is needed. The user is allowed to send the pictures to the search engine server in order to search for his/her desired pictures. Then, the server will find the pictures which are relevant or exact matched to the pictures the user demand and present the searching results on the web page.

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้อย่างดีอันเนื่องมาจากคำแนะนำ และคำปรึกษาจากอาจารย์ที่ปรึกษา ศศ.ดร.อรฉัตร จิตต์โสภักดิ์ ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์ ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านอาจารย์และขอขอบพระคุณเป็นอย่างสูงที่ทำให้ข้าพเจ้าได้สำเร็จการศึกษา

ขอกราบพระคุณคณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้าจนได้สำเร็จการศึกษา



กาญจกั พันระไชย  
อิทธิพงษ์ นพฤทธิ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	I
1.1 ความสำคัญและที่มาของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 ขอบเขตของโครงการ.....	2
1.4 วิธีการดำเนินการ.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.6 ส่วนประกอบของปริญาานิพนธ์.....	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	4
2.1 กลไกการทำงานของเสิร์จเอนจิน.....	4
2.2 องค์ประกอบของเสิร์จเอนจิน.....	5
2.2.1 ไรบอท (robots).....	5
2.2.2 อินเด็กเซอร์ (Indexer).....	5
2.2.3 ซอฟต์แวร์เสิร์จเอนจิน ( Search engine software ).....	5
2.3 โปรแกรมไรบอท (Web Crawler).....	6
2.3.1 การทำงานของไรบอท.....	7
2.4 โปรแกรมอินเด็กเซอร์.....	7
2.5 ซอฟต์แวร์เสิร์จเอนจิน.....	7

## สารบัญ (ต่อ)

	หน้า
2.6 ความรู้เบื้องต้นเกี่ยวกับภาษา Python.....	8
2.6.1 หลักการทำงานของภาษา Python.....	8
2.6.2 คอมไพเลอร์ (Compiler).....	8
2.6.3 อินเตอร์พรีเตอร์ (Interpreter) .....	9
2.6.4 ข้อดีของภาษา Python.....	10
2.7 ภาษา PHP.....	11
2.7.1 สาเหตุที่เลือกพัฒนาเว็บด้วย PHP.....	11
2.7.2 ความสามารถของ PHP.....	12
2.7.3 การทำงานของ PHP.....	12
2.7.4 ข้อดีของภาษา PHP.....	13
2.7.5 ข้อเสียของภาษา PHP.....	13
2.8 ความรู้เบื้องต้นเกี่ยวกับระดับของสี.....	13
2.8.1 สีระดับที่ 1(primary) .....	13
2.8.2 สีระดับที่ 2 (secondary).....	14
2.8.3 สีระดับที่ 3 (Tertiary) .....	14
2.9 ข้อมูลภาพ (Image Data).....	14
2.9.1 คุณสมบัติของกลุ่มสีที่แสดงผล.....	15
2.9.1.1 Uniformity.....	15
2.9.1.2 Completeness.....	15
2.9.1.3 Compactness.....	15
2.9.1.4 Naturalness.....	15
2.9.2 ระบบสีแบบ RGB .....	16
2.9.3 กลุ่มสีแสดงผลแบบ RGB.....	16
2.10 ฮิสโตแกรม (Histogram).....	17
2.10.1 การ Quantization.....	18
2.10.2 การ Normalization.....	18
2.10.3 ขั้นตอนการทำฮิสโตแกรม.....	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

	หน้า
บทที่ 3 การค้นหารูปภาพโดยใช้เสิร์จเอนจิน.....	20
3.1 การออกแบบและพัฒนาระบบเสิร์จเอนจิน.....	20
3.2 การออกแบบฐานข้อมูลที่ใช้ในระบบเสิร์จเอนจิน.....	21
3.2.1 ตารางเก็บข้อมูล table_link_web .....	21
3.2.2 ตารางเก็บข้อมูล table_link_image.....	22
3.3 การทำงานของเสิร์จเอนจินเพื่อทำการเก็บลิงค์ของเว็บ.....	23
3.3.1 การทำงานของ Web crawler ที่ใช้เก็บลิงค์ของเว็บและลิงค์ในระดับถัดไป.....	25
3.4 การทำงานของเสิร์จเอนจินเพื่อทำการ โหลดรูปภาพมาคำนวณหาค่าฮิสโตแกรม.....	31
3.4.1 หลักการทำงาน.....	31
3.4.2 ตัวอย่างค่าฮิสโตแกรมของภาพที่เก็บอยู่ในฐานข้อมูล.....	32
3.5 การค้นหารูปภาพของผู้ใช้.....	36
บทที่ 4 การทดลองและผลการทดลอง .....	39
4.1 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้เก็บลิงค์ของเว็บ.....	39
4.2 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้เก็บลิงค์ของภาพ.....	40
4.3 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้คำนวณค่าฮิสโตแกรม.....	41
4.4 การทดลองเพื่อทดสอบค่าพารามิเตอร์ที่ใช้ค้นหาภาพที่เหมือน คล้ายและใกล้เคียง.....	42
4.4.1 การค้นหาข้อมูล โดยใช้ค่า Differential = 0.....	42
4.4.2 การค้นหาข้อมูล โดยใช้ค่า Differential = 1.....	44
4.4.3 การค้นหาข้อมูล โดยใช้ค่า Differential = 2.....	45
4.4.4 การค้นหาข้อมูล โดยใช้ค่า Differential = 3.....	47
4.4.5 การค้นหาข้อมูล โดยใช้ค่า Differential = 4.....	50
4.4.6 การค้นหาข้อมูล โดยใช้ค่า Differential = 5.....	54
4.4.7 การวิเคราะห์และสรุปผลการค้นหาด้วยค่า differential 0-5.....	59
4.5 การทดลองหาภาพที่เหมือน คล้ายและใกล้เคียง.....	60
4.5.1 การทดลองหาภาพที่เหมือน.....	60
4.5.2 การทดลองหาภาพที่คล้าย.....	61
4.5.3 การทดลองหาภาพที่ใกล้เคียง.....	62

# สารบัญ (ต่อ)

หน้า

บทที่ 5 บทวิจารณ์และสรุป.....	65
5.1 บทสรุป.....	65
5.2 ปัญหาอุปสรรคและแนวทางในการแก้ไข.....	66
5.3 แนวทางการพัฒนาต่อ.....	66
บรรณานุกรม.....	67



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

ตารางที่	หน้า
2.1 ฐานข้อมูลที่ภาษา PHP รองรับ.....	12
2.2 ระดับของสีที่ใช้ในการทำ Quantization.....	18
3.1 ตาราง table_link_web.....	21
3.2 ตาราง table_link_image.....	22
4.1 ทดสอบการทำงานของโปรแกรม Web Crawler ที่ใช้เก็บลิงค์ของเว็บ.....	39
4.2 ทดสอบการทำงานของโปรแกรม Web Crawler ที่ใช้เก็บลิงค์ของภาพ.....	40
4.3 ทดสอบการทำงานของโปรแกรมที่คำนวณหาค่าฮิสโตแกรม.....	41
4.4 ตารางเปรียบเทียบการทำงานของ Differential ในแต่ละระดับ.....	59



# สารบัญรูป

รูปที่	หน้า
2.1 โครงสร้างของระบบ Search Engine.....	6
2.2 การทำงานของคอมพิวเตอร์.....	9
2.3 การทำงานของอินเทอร์เน็ต.....	9
2.4 แสดงกราฟของฮิสโตแกรม.....	17
2.5 ขั้นตอนการทำฮิสโตแกรม.....	19
3.1 ภาพรวมของระบบอิมเมจเสิร์จเอนจินที่ออกแบบ.....	20
3.2 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์.....	24
3.3 แสดงค่า empty เพื่อหาลิงค์ที่จะนำมาโหลด.....	25
3.4 แสดงลิงค์ของระดับถัดไปว่ามาจากลิงค์ที่ตัว Id_Web ใดในตาราง table_link_web.....	29
3.5 แสดงลิงค์ของระดับถัดไปว่ามาจากลิงค์ของเว็บเพจใด.....	29
3.6 แสดงลิงค์ของภาพว่ามาจากลิงค์ของเว็บเพจใดในตาราง table_link_web.....	30
3.7 ขั้นตอนการทำงานของเสิร์จเอนจินที่ใช้ในการคำนวณค่าฮิสโตแกรม.....	31
3.8 แสดงค่าฮิสโตแกรมของภาพต้นแบบที่มีค่า Differential = 0.....	32
3.9 แสดงค่าฮิสโตแกรมของภาพต้นแบบที่มีค่า Differential = 1.....	32
3.10 แสดงค่าฮิสโตแกรมของภาพต้นแบบที่มีค่า Differential = 2.....	33
3.11 แสดงค่าฮิสโตแกรมของภาพต้นแบบที่มีค่า Differential = 3.....	34
3.12 แสดงค่าฮิสโตแกรมของภาพต้นแบบที่มีค่า Differential = 4.....	34
3.13 ขั้นตอนการทำงานผู้ใช้ในการค้นหารูปภาพ.....	36
3.14 หน้าจอของผู้ใช้งาน.....	37
3.15 ผู้ใช้เลือกไฟล์รูปที่ต้องการค้นหาจากเครื่อง.....	37
3.16 แสดงรูปภาพที่ผู้ใช้ต้องการค้นหา.....	38
4.1 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 0 ในภาพชุดทดลองที่ 1.....	42
4.2 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 0 ในภาพชุดทดลองที่ 2.....	43
4.3 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 0 ในภาพชุดทดลองที่ 3.....	43
4.4 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 0 ในภาพชุดทดลองที่ 4.....	43
4.5 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 0 ในภาพชุดทดลองที่ 5.....	43
4.6 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 1.....	44
4.7 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 2.....	44

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป(ต่อ)

รูปที่	หน้า
4.8 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 3.....	44
4.9 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 4.....	45
4.10 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 5.....	45
4.11 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 1.....	45
4.12 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 2.....	46
4.13 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 3.....	46
4.14 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 4.....	46
4.15 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 5.....	47
4.16 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 1.....	47
4.17 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 2.....	48
4.18 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 3.....	48
4.19 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 4.....	49
4.20 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 5.....	49
4.21 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 1.....	50
4.22 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 2.....	50
4.23 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 3.....	51
4.24 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 4.....	52
4.25 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 5.....	53
4.26 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 1.....	54
4.27 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 2.....	55
4.28 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 3.....	56
4.29 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 4.....	56
4.30 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 5.....	58
4.31 ภาพชุดทดลองการหาภาพที่เหมือนชุดที่ 1.....	60
4.32 ภาพชุดทดลองการหาภาพที่เหมือนชุดที่ 2.....	60
4.33 ภาพชุดทดลองการหาภาพที่เหมือนชุดที่ 3.....	60
4.34 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 1.....	61
4.35 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 2.....	61
4.36 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 3.....	62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป(ต่อ)

รูปที่	หน้า
4.37 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 1.....	62
4.38 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 2.....	63
4.39 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 3.....	63



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของโครงการ

ในปัจจุบันเครือข่ายอินเทอร์เน็ตได้มีความทันสมัยและมีขนาดที่กว้างขวางมาก ภายในเครือข่ายอินเทอร์เน็ตก็จะประกอบไปด้วยเว็บไซต์จำนวนมากเช่นกันเพราะฉะนั้นข้อมูลรูปภาพจึงมีจำนวนมากตามไปด้วยเช่นกัน ข้อมูลรูปภาพเป็นข้อมูลที่สำคัญต่อการใช้ชีวิตประจำวันไม่ว่าจะนำไปใช้ในรูปแบบต่างๆ ซึ่งปัจจุบันเทคโนโลยีทางด้านเว็บไซต์ได้มีการพัฒนาอย่างรวดเร็วและมีความก้าวหน้าอย่างรวดเร็ว จึงก่อให้เกิดเว็บไซต์ที่ต่างๆมากขึ้น ข้อมูลข่าวสารในเครือข่ายอินเทอร์เน็ตจึงเพิ่มมากขึ้น เป็นสาเหตุทำให้การค้นหาข้อมูลในเครือข่ายอินเทอร์เน็ตมีความยากลำบากดังนั้นข้อมูลรูปภาพที่อยู่ภายในเครือข่ายอินเทอร์เน็ตก็ย่อมมีจำนวนมากเช่นกัน การที่จะค้นหารูปภาพภายในเครือข่ายอินเทอร์เน็ตนั้นจึงมีความยากลำบากมาก

ข้อมูลประเภทรูปภาพถือว่าเป็นข้อมูลที่มีความสำคัญต่อผู้ใช้งานในเครือข่ายอินเทอร์เน็ตมาก และยังเป็นข้อมูลสำคัญในด้านต่างๆ มากมาย ซึ่งข้อมูลรูปภาพนี้อาจก่อให้เกิดการละเมิดสิทธิส่วนบุคคล หรือนำรูปภาพไปใช้ในทางผิดกฎหมายซึ่งปัจจุบันรูปภาพยังสามารถสร้างขึ้นได้ง่าย เช่น การถ่ายรูป (Photograph) การสแกนรูปภาพด้วยสแกนเนอร์ (Scanner) เป็นต้น ดังนั้นการที่จะได้มาซึ่งข้อมูลรูปภาพที่เราต้องการจึงจำเป็นที่จะต้องหาเครื่องมือมาช่วยในการค้นหารูปภาพที่เราต้องการให้ได้ถูกต้อง และสามารถค้นหาว่าอยู่ที่ใดในเครือข่ายอินเทอร์เน็ต โดยใช้รูปภาพเป็นอินพุตในการค้นหา ซึ่งนั่นเป็นที่มาของ โครงการอิมเมจเสิร์จเอนจิน Image Search Engine และ อิมเมจเสิร์จเอนจิน Image Search Engine จะสามารถบอกถึงที่อยู่ของรูปภาพได้ดีกว่ารูปที่ได้มานั้นได้มาจากลิงค์ใดภายในเครือข่ายอินเทอร์เน็ต

### 1.2 วัตถุประสงค์ของโครงการ

ปรัชญาพันธกิจฉบับนี้ได้นำเสนอวิธีการศึกษาเกี่ยวกับข้อมูลที่เป็นรูปภาพในรูปแบบต่างๆ นำเสนอวิธีการศึกษาด้วยตนเองว่ามีการทำงานในรูปแบบใด และศึกษาวิธีการออกแบบตัวอิมเมจเสิร์จเอนจิน ในส่วนของเค้าเบสก็ได้มีการนำเสนอการออกแบบเกี่ยวกับตารางต่างๆ ที่เกี่ยวข้องกับการใช้งาน และยังได้นำเสนอการศึกษาภาษาที่ใช้ในการพัฒนาโปรแกรมในส่วนต่างๆ การใช้งานโดยใช้ภาพจากเครื่องผู้ใช้เป็นตัวค้นหารูปภาพที่อยู่ในเครือข่ายอินเทอร์เน็ตที่ได้ทำการโหลดเพื่อนำมาเก็บไว้ในฐานข้อมูล โดยที่ส่วนของเครื่องเซิร์ฟเวอร์จะทำการรวบรวมรูปภาพจากเว็บไซต์ต่างๆ มาเก็บไว้ในฐานข้อมูลของตัวเครื่องเซิร์ฟเวอร์ และทำการหาลักษณะเฉพาะของรูปภาพแต่ละรูปด้วยกระบวนการหาค่าฮิสโตแกรม เมื่อได้ค่าดังกล่าวมาแล้วจะนำไปเก็บไว้ในฐานข้อมูลเพื่อทำการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เปรียบเทียบในแต่ละรูปว่ามีความคล้าย หรือเหมือนมากน้อยเพียงใด และจะนำผลที่ได้แสดงผลออกทางหน้าเว็บเพจตามที่ผู้ใช้ต้องการ แล้วนำรูปภาพที่ได้มาเก็บไว้ในฐานข้อมูล ส่วนที่ใช้ในการประมวลผลภาพเพื่อหาค่าสีสโตแกรม ของภาพเพื่อบอกลักษณะเฉพาะของรูปภาพแต่ละรูปโดยค่าที่แสดงออกมานั้นจะแสดงออกมาเป็นค่าสีสโตแกรม และส่วนสุดท้าย คือส่วนที่ทำหน้าที่ในการติดต่อกันระหว่างผู้ใช้กับเซิร์ฟเวอร์เพื่อทำการค้นหารูปภาพตามที่ผู้ใช้ต้องการ

### 1.3 ขอบเขตของโครงการ

ปริญญาบัตรฉบับนี้ได้นำเสนอวิธีการศึกษา และออกแบบอิมเมจเสิร์จเอนจิน โดยแบ่งออกเป็น ส่วนของโปรแกรมที่ใช้ในการรวบรวมรูปภาพมาจากเว็บไซต์ต่างๆ (Web Crawler) แล้วทำการเก็บลิงค์ของรูปภาพต่างๆ ไว้ในฐานข้อมูล และส่วนของโปรแกรมที่ทำหน้าที่ในการโหลดรูปหลังจากที่ได้ทำการเก็บลิงค์รูปภาพนั้นมาเก็บไว้ ส่วนของโปรแกรมที่ทำการเปรียบเทียบรูปเพื่อค้นหารูปที่ผู้ใช้ต้องการ และส่วนของการจัดการกับฐานข้อมูล ภายในระบบ

### 1.4 วิธีการดำเนินการ

1. ศึกษาการทำงานของระบบอิมเมจเสิร์จเอนจิน โดยรวมว่ามีการทำงานโดยรวมเป็นอย่างไร แต่ละส่วนทำหน้าที่อย่างไร
2. ศึกษาภาษาที่ใช้ในการพัฒนาโปรแกรมในส่วนต่างๆ ของระบบทั้งหมด
3. ศึกษาวิธีการจัดการระบบฐานข้อมูลว่าแต่ละส่วนในระบบมีการทำงานอย่างไร
4. ศึกษาวิธีการหาลักษณะเฉพาะของภาพรูปว่ามีการทำงานอย่างไร
5. ศึกษาข้อมูลที่เกี่ยวข้องกับองค์ประกอบของรูปภาพ
6. ศึกษาข้อมูลของกลุ่มสีที่แสดงผลแบบ RGB
7. วิเคราะห์ และออกแบบระบบ
  - 7.1 ออกแบบระบบ โดยรวมของการทำงาน ในแต่ละส่วน
  - 7.2 ออกแบบระบบ Database ในการจัดการเกี่ยวกับฐานข้อมูลภายในระบบ
  - 7.3 ออกแบบส่วนของโปรแกรมที่ทำหน้าที่ในการเก็บลิงค์ของรูปภาพ
  - 7.4 ออกแบบส่วนของโปรแกรมที่ทำหน้าที่ในการโหลดรูปภาพ
  - 7.5 ออกแบบส่วนของโปรแกรมที่ทำการหาค่าสีสโตแกรมเพื่อหาลักษณะเฉพาะของรูป
  - 7.6 ออกแบบส่วนของหน้าเว็บที่จะให้ผู้ใช้ค้นหารูปภาพ
8. พัฒนาส่วนประกอบของระบบ ในแต่ละส่วนและนำส่วนประกอบต่างๆของระบบมารวมกัน
9. ทดสอบในแต่ละส่วนของระบบว่าการทำงานเป็นไปตามความต้องการหรือไม่
10. แก้ไขจุดบกพร่องที่เกิดจากปัญหาต่างๆ ที่เกิดขึ้นในระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้รับความรู้ความเข้าใจเกี่ยวกับการออกแบบระบบเสิร์จเอนจินเพื่อนำไปใช้งานได้จริง
2. ได้รับความรู้ความเข้าใจเกี่ยวกับระบบเสิร์จเอนจิน และวิธีการทำงานของเสิร์จเอนจินว่าในแต่ละส่วนของระบบเสิร์จเอนจินมีการทำงานอย่างไร
3. ได้รับความรู้ความเข้าใจเกี่ยวกับวิธีการทำงานของโปรแกรมเก็บลิงค์ของรูปภาพและโปรแกรมโหลดรูปภาพจากเว็บไซต์ต่างๆ เพื่อนำมาเก็บในฐานข้อมูล
4. ได้รับความรู้ความเข้าใจเกี่ยวกับระบบฐานข้อมูลและวิธีการจัดการกับระบบฐานข้อมูล
5. ได้รับความรู้ความเข้าใจเกี่ยวกับข้อมูลรูปภาพในรูปแบบต่างๆ
6. ได้รับความรู้ความเข้าใจเกี่ยวกับการหาลักษณะเฉพาะของภาพโดยอาศัยหลักการหาค่าฮิสโตแกรม
7. ได้รับความรู้ความเข้าใจเกี่ยวกับการสร้างเว็บเพจเพื่อทำหน้าที่ในการติดต่อกับผู้ใช้

## 1.6 ส่วนประกอบของปฏิญานิพนธ์

ปฏิญานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

- บทที่ 1 กล่าวถึงความสำคัญและที่มาของโครงการ วัตถุประสงค์ของโครงการ ขอบเขตของโครงการ วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ
- บทที่ 2 กล่าวถึงทฤษฎีที่เกี่ยวข้องที่ใช้ในการทำงานของเสิร์จเอนจิน
- บทที่ 3 กล่าวถึงกระบวนการค้นหาภาพโดยใช้เสิร์จเอนจิน
- บทที่ 4 กล่าวถึงการทดลองและผลการทดลอง
- บทที่ 5 เป็นบทวิจารณ์และสรุป ซึ่งกล่าวถึงบทสรุปของโครงการ วิจารณ์สิ่งที่ได้รับจากโครงการ และข้อเสนอแนะสำหรับเป็นแนวทางในการพัฒนาต่อ

## บทที่ 2

# ทฤษฎีที่เกี่ยวข้อง

### 2.1 กลไกการทำงานของเสิร์จเอนจิน

เสิร์จเอนจินทำงาน โดยการเก็บข้อมูลของแต่ละหน้าเว็บเพจในเครือข่ายอินเทอร์เน็ต ซึ่งเสิร์จเอนจินสามารถดึงเอาข้อมูลออกมาจากตัวเว็บ ไซค์ได้เอง โปรแกรมที่ทำหน้าที่ดึงข้อมูลออกมาจากเว็บต่างๆ คือ โปรแกรมโรบอท โดยโรบอทหรือ Web crawler จะทำหน้าที่ในการดึงลิงค์ข้อมูลของหน้านั้นมาเก็บไว้ในฐานข้อมูล ซึ่งข้อมูลที่โรบอท Web crawler ได้หามาจะเป็นลิงค์ของรูปภาพในเว็บไซค์ต่างๆ หลังจากนั้นจะนำมาเก็บไว้ยังฐานข้อมูลเพื่อไว้ทำการ ตรวจสอบว่าลิงค์ใดสามารถที่จะโหลดรูปมาได้บ้าง

การทำงานของเสิร์จเอนจิน จะแบ่งออกเป็น 3 ส่วน คือ

1) โปรแกรมรวบรวมเอกสารเว็บ (Spider หรือ Crawler) จะทำหน้าที่ออกไปสำรวจเว็บจากหน้าเว็บที่เราต้องการจะค้นหารูป โดยตัวโรบอทจะเข้าไปสำรวจและเก็บลิงค์ที่เกี่ยวข้องกับรูปภาพ เมื่อโรบอทพบลิงค์ที่ต้องการ จากนั้นจึงนำมาเก็บไว้ในฐานข้อมูล และเมื่อพบว่าลิงค์นั้นไม่สามารถเข้าไปดึงข้อมูลออกมาได้ (dead link) ก็จะมีการจับเวลาเพื่อป้องกันการวนลูบและการทำงานแบบซ้ำๆ เมื่อไม่สามารถเก็บข้อมูลได้ก็จะหยุดการทำงานและจะเริ่มทำการเก็บข้อมูลที่ลิงค์อื่นแทนต่อไป

2) การสร้างดัชนีค้นหาโดยก่อนที่จะนำมาทำการค้นหานั้นต้องนำมาทำการสร้างดัชนีค้นหาก่อน โดยจะนำเอาเว็บที่โรบอทหามาทำโดยจะเปรียบเสมือนเป็นสมุดที่ใช้ในการจัดเก็บสำเนาของเว็บเพจที่ Spider พบ การทำงานในลักษณะนี้อาจเรียกได้ว่า (อินเด็กเซอร์)

3) โปรแกรมการสืบค้น (เสิร์จเอนจินซอฟต์แวร์) ทำหน้าที่ค้นหาข้อมูลจากฐานข้อมูลของเสิร์จเอนจิน จะเริ่มต้นการทำงานเมื่อผู้ใช้ทำการป้อนข้อมูลที่ต้องการค้นหา โดยโปรแกรมจะทำหน้าที่นำข้อมูลการค้นหาของผู้ใช้มาจับคู่กับบรรณในฐานข้อมูลแล้วทำการดึงข้อมูลที่ตรงกับคำค้นหาออกมาและทำการเปรียบเทียบความเหมือนโดยใช้หลักการฮีสโตแกรมเพื่อค้นหาข้อมูลที่เกี่ยวข้องที่โปรแกรมประเมิน ได้ออกมาซึ่งเสิร์จเอนจินแต่ละตัวจะใช้คำหรือตรรกะในการค้นหาที่แตกต่างกันออกไปหรือ Regular Expression (หรือ REs) ที่แตกต่างกันออกไป

## 2.2 องค์ประกอบของเสิร์จเอนจิน

เสิร์จเอนจินประกอบด้วย 3 ส่วนหลักได้แก่

- โรบอท
- อินเด็กซ์เซอร์
- เสิร์จเอนจินซอฟต์แวร์

### 2.2.1 โรบอท (robots)

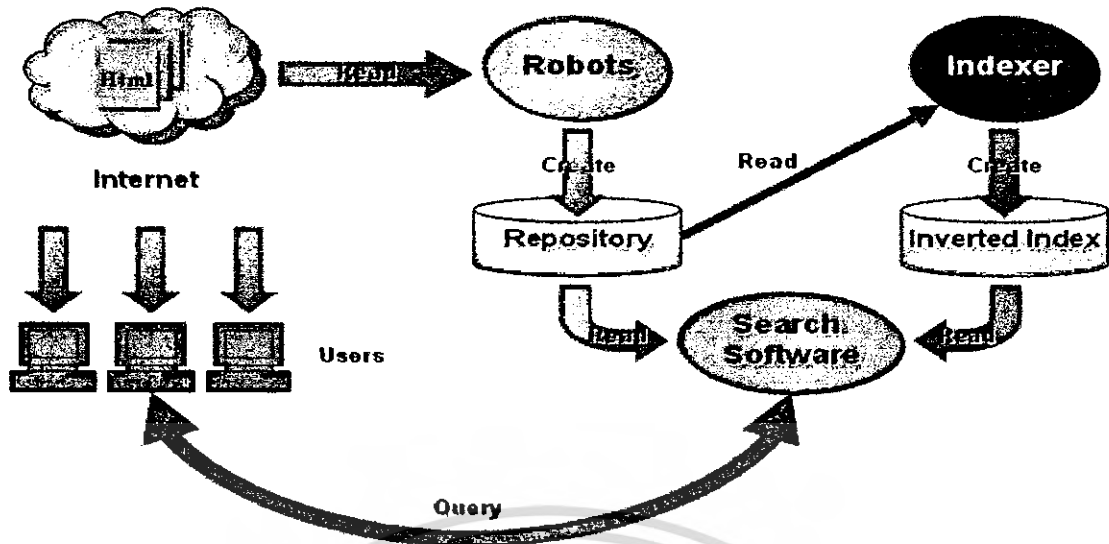
ทำหน้าที่ไปยังเว็บไซต์ต่างๆเพื่อสะสมไฟล์ HTML หรือไฟล์ที่เกี่ยวข้องกับรูปภาพ หรืออาจเป็นลิงค์ของรูปภาพ เพื่อนำข้อมูลเหล่านั้นมาสร้างดัชนีค้นหาและเมื่อโรบอทรวบรวมข้อมูลต่างๆ แล้ว โรบอทยังสามารถที่จะเช็คเวลาในการเข้าไปยังลิงค์ของข้อมูลในแต่ละเว็บเพจ ได้อีกด้วยเพื่อป้องกันปัญหาการเก็บลิงค์ที่ซ้ำกันหรือป้องกันการวนลูปเพื่อที่จะได้ไม่เป็นการเสียเวลา

### 2.2.2 อินเด็กซ์เซอร์ (Indexer)

เป็นส่วนที่ทำหน้าที่สร้างดัชนีค้นหาจากไฟล์ที่โรบอทหามา วิธีการสร้างดัชนีค้นหานั้นมีอยู่ด้วยกันหลายวิธี ซึ่งแต่ละวิธีก็มีข้อดีข้อเสียต่างกัน แต่วิธีที่เป็นที่นิยมมากสำหรับเสิร์จเอนจินก็คือ Inverted index เว็บเพจที่จะทำการค้นหาได้นั้นจะต้องได้รับการทำดัชนีค้นหามาก่อนจึงจะทำการค้นหาได้

### 2.2.3 ซอฟต์แวร์เสิร์จเอนจิน ( Search engine software )

เป็น โปรแกรมทำหน้าที่รับคำศัพท์ที่ผู้ใช้งานต้องการค้นหาหรือข้อมูลที่ผู้ใช้งานต้องการค้นหามาค้นหาในดัชนีค้นหาที่ถูกสร้างขึ้น โดยอินเด็กซ์เซอร์เพื่อทำการค้นหาแล้วจากนั้นตัวซอฟต์แวร์ให้ตรงกับความต้องการของผู้ใช้ให้ได้มากที่สุด



รูปที่ 2.1 โครงสร้างการทำงานของระบบ Search Engine

## 2.3 โปรแกรมโรบอท (Web Crawler)

คือโปรแกรมซึ่งจะทำการดึงทุกๆข้อมูลที่อยู่ภายในเว็บไซต์ หรือระบบเครือข่ายอินเทอร์เน็ต โดยการกระจายสู่ Hyperlink ที่อ้างถึงจากหน้าแรกแล้วทำการดึงข้อมูลมา โดยที่การทำงานของเว็บโรบอท จะต้องขึ้นกับ robots.txt เป็นโปรโตคอล ซึ่งแต่ละเว็บไซต์จะมีเป็นของตัวเองเพื่อที่จะบอกว่าสามารถเข้าไปยังไฟล์ใดได้บ้างแต่โรบอทก็มีข้อเสีย อาทิเช่น

- โรบอททำให้เครือข่ายทำงานหนักมากขึ้น ซึ่งส่วนมากจะเกิดกับคนที่เริ่มเขียน โรบอท แต่ปัจจุบันมีข้อมูลเพียงพอที่จะป้องกันความผิดพลาดนี้แล้ว
- โรบอทถูกสร้างขึ้น โดยมนุษย์ ซึ่งถูกกระทำอย่างผิดพลาดในการตั้งองค์ประกอบ หรือไม่ได้พิจารณาการกระทำที่แฝงมาภายหลัง

### 2.3.1 การทำงานของโรบอท

การทำงานของโรบอททั่วไปจะมีลักษณะใกล้เคียงกันคือ โรบอทจะอ่านเว็บเพจเริ่มต้นที่กำหนดไว้ให้ หลังจากได้เว็บเพจเริ่มต้นมาแล้ว โรบอทจะไปอ่านเว็บเพจเหล่านั้น มาแล้วตรวจดูว่าภายในเว็บเพจนั้นมีไฮเปอร์ลิงค์ชี้ไปยัง URL อะไรบ้างและตรวจสอบดูว่าโรบอทเคยพบ URL เหล่านั้นมาก่อนหรือไม่ หากเคยพบ URL มาก่อนก็จะทิ้ง URL นั้นไป แต่หากยังไม่เคยพบมาก่อนก็จะมาวิเคราะห์ดูว่าสมควรที่จะอ่านเว็บเพจนั้นหรือไม่ โรบอททั่วไปจะสามารถกำหนดกฎได้ โดยกฎที่ตั้งไว้จะใช้สำหรับกำหนดคุณสมบัติของ URL เช่น จะอ่านเว็บไซต์ใดและไม่อ่านเว็บไซต์ใดบ้าง

จากกฎที่ตั้งไว้หากเว็บพบเว็บเพจที่ผ่านกฎที่ตั้งไว้ มันก็จะเก็บ URL นั้นไว้ในคิวเพื่อรอที่จะอ่านเว็บเพจนั้นต่อไป แต่หากเว็บเพจใดไม่ผ่านกฎเว็บก็ข้าม URL นั้นไปโดยไม่สนใจโปรแกรมเว็บจะวนซ้ำอ่านเว็บเพจเช่นนี้ไปเรื่อยๆ จนกว่า URL ที่อยู่ในคิวจะหมดหรือจนกว่าจะได้เว็บเพจครบตามจำนวนที่ต้องการ

## 2.4 โปรแกรมอินเด็กเซอร์

การทำงานของโปรแกรม Indexer นั้นมีวิธีการสร้างดัชนีค้นหาอยู่หลายวิธีเช่น แบบอินเวิร์ตอินเด็กซ์ , แบบซิกเนเจอร์ไฟล์ และแบบเวกเตอร์สเปซโมเดล เป็นต้น ซึ่งในโครงงานนี้จะอธิบายถึงการใช้แบบอินเวิร์ตอินเด็กซ์ ก่อนที่จะอธิบายถึงวิธีการจัดเก็บข้อมูลของอินเวิร์ตอินเด็กซ์จะขอยกตัวอย่างเพื่ออธิบายหลักการทำงาน หากเรามีหนังสือเล่มหนึ่งและต้องการค้นหาคำว่า “network” อยู่ที่ใด หากแต่เรามักใช้วิธีเปิดไปที่ท้ายเล่มเพื่อดูดัชนีคำศัพท์ของหนังสือเล่มนั้นว่ามีคำว่า “network” อยู่ที่หน้าไหน รูปแบบของดัชนีค้นหาของอินเวิร์ตอินเด็กซ์ก็อาศัยหลักการเดียวกันกับดัชนีท้ายเล่มของหนังสือ อินเวิร์ตอินเด็กซ์ก็คือตารางของคำศัพท์ที่ถูกสร้างขึ้นมาไว้สำหรับค้นหา โดยที่ภายในของอินเวิร์ตอินเด็กซ์นั้นจัดเก็บคำศัพท์และตำแหน่งที่อยู่ของคำศัพท์เหล่านั้นว่าปรากฏอยู่ที่ใด เมื่อต้องการค้นหาก็ไม่จำเป็นต้องค้นหาคำศัพท์จากเอกสารทีละเอกสาร แต่สามารถตรวจดูที่อินเวิร์ตอินเด็กซ์ได้ทันทีว่าคำศัพท์คำที่ต้องการค้นหาอยู่ที่ใด

โปรแกรมอินเด็กเซอร์มีหน้าที่สร้างตารางอินเวิร์ตอินเด็กซ์ ข้อมูลในอินเวิร์ตอินเด็กซ์ นั้นนอกจากจะเก็บตำแหน่งที่อยู่ของคำศัพท์แล้วจะมีคำสำคัญอื่นๆ อีกหลายคำด้วยกัน ซึ่งคำเหล่านี้มีไว้สำหรับช่วยให้โปรแกรมเสิร์จเอนจินสามารถจัดเรียงลำดับของผลลัพธ์ได้ดียิ่งขึ้น อาทิ คำนำหน้าที่ใช้ในการจัดเรียงลำดับ เป็นต้น

## 2.5 ซอฟต์แวร์เสิร์จเอนจิน

โปรแกรมค้นหาที่มีหน้าที่รับคำศัพท์ที่ต้องการค้นหาจากผู้ใช้ มาค้นหาในดัชนีค้นหา และจัดเรียงลำดับผลลัพธ์ให้ได้ตรงกับความต้องการของผู้ใช้มากที่สุด โดยอาศัยข้อมูลที่ได้เตรียมไว้ก่อนแล้วจากโปรแกรมอินเด็กเซอร์

การค้นหาแต่ละครั้งนั้น โปรแกรมค้นหาจะนำคำแต่ละคำในชุดวลีค้นหา (query string) ที่ผู้ใช้ป้อนเข้ามา เพื่อเปรียบเทียบกับคำศัพท์ในตารางอินเวิร์ตอินเด็กซ์ เพื่อหาเอกสารที่มีคำศัพท์ตรงกัน หลังจากที่ได้รายการเอกสารที่ตรงกันกับคำศัพท์แต่ละคำแล้ว โปรแกรมค้นหาจะนำรายการเอกสารมาจัดเรียงลำดับของเอกสาร โดยเรียงลำดับจากเอกสารที่มีคำนำหน้าในการค้นหามากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไปน้อยที่สุด หลังจากนั้น โปรแกรมค้นหาจะส่งคืนรายการเอกสารที่ได้มีการจัดเรียงลำดับเรียบร้อยแล้ว แล้วกลับไปแสดงผล

## 2.6 ความรู้เบื้องต้นเกี่ยวกับภาษา Python

Python คืออะไร Python เป็นภาษาระดับสูงภาษาหนึ่ง ที่มีความสามารถสูงถูกสร้างขึ้นในปี 1989 โดย Guido van Rossum ซึ่งถูกพัฒนาขึ้นมาโดยไม่มีขีดจำกัดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux, Windows NT, Windows 2000, Windows XP อีกอย่างหนึ่งภาษาตัวนี้เป็นภาษาลักษณะ Open Source

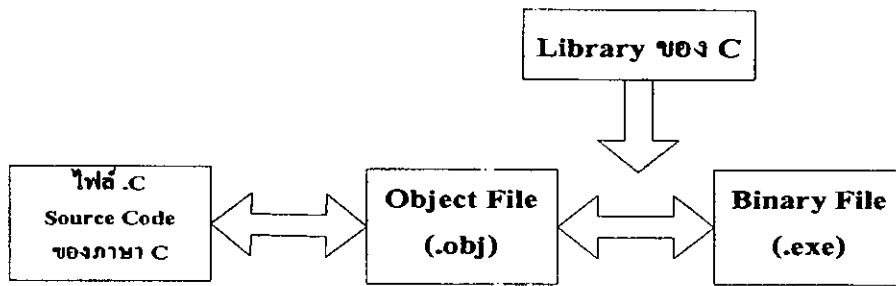
### 2.6.1 หลักการทำงานของภาษา Python

การใช้ภาษา Python ในการพัฒนา web application นี้เพราะ Python เป็นภาษาระดับสูงภาษาหนึ่งที่เป็น object-oriented ที่มีความสามารถสูงถูกสร้างขึ้นในปี 1989 โดย Guido van Rossum ที่รวมเอาความสามารถแบบ dynamic ซึ่งทำให้เป็นที่น่าสนใจในการพัฒนา application ได้อย่างรวดเร็ว เช่นเดียวกับการใช้ในรูปแบบสคริปต์เพื่อเชื่อมต่อกับ component ที่มีอยู่แล้วเข้าด้วยกัน ทั้งนี้ยังเป็นภาษาที่ง่ายต่อการเรียนรู้ และยังสนับสนุนโมดูลและแพ็คเกจต่างๆ ทำให้สามารถนำมาประกอบใช้ได้ใหม่ ซึ่งภาษานี้พัฒนาขึ้นมาโดยไม่มีขีดจำกัดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux, Windows NT/2000/XP/2003 หรือแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาตัวนี้เป็นภาษาลักษณะ Open Source ที่แจกจ่ายให้ใช้งานฟรี

เมื่อทำการเขียนโค้ดขึ้นมาตามโครงสร้างของโปรแกรมภาษาใดก็ตาม และการจะให้โค้ดคำสั่งเหล่านั้นทำงานได้ก็จะต้องมีตัวแปลภาษามาจัดการแปลโค้ดคำสั่ง เพื่อให้ทำงานตามที่ ต้องการ โดยลักษณะของตัวแปลภาษานั้นแบ่งได้ 2 ประเภทใหญ่ๆ คือ

### 2.6.2 คอมไพเลอร์ (Compiler)

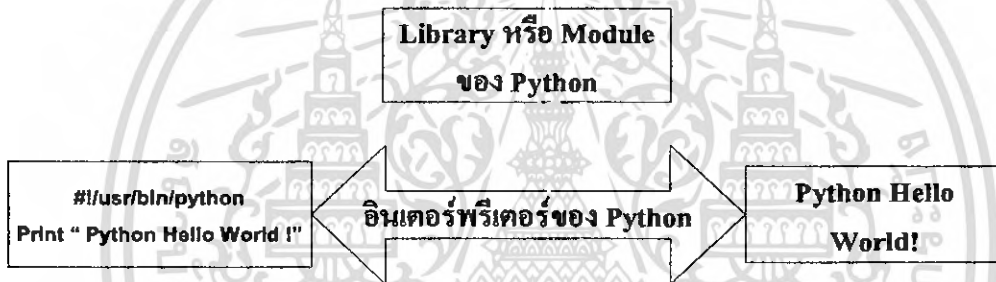
เป็นตัวแปลภาษาสำหรับภาษา C, C++, Pascal การทำงานก็คือจะตรวจสอบความผิดพลาดของโค้ดคำสั่งตั้งแต่ต้นจนจบก่อน หรือเรียกว่าการคอมไพล์ ถ้าไม่มีข้อผิดพลาดก็จะทำการแปลโค้ดคำสั่งให้เป็นไฟล์นามสกุล .obj (object file) จากนั้นก็ทำการแปลไฟล์ .obj ให้เป็นไบนารีไฟล์ .exe เพื่อทำงานต่อไป ดังตัวอย่างการทำงานของคอมไพเลอร์ภาษา C ดังรูปที่ 2.2



รูปที่ 2.2 การทำงานของคอมไพเลอร์

### 2.6.3 อินเตอร์พรีเตอร์ (Interpreter)

จะทำงานเป็นบรรทัดต่อบรรทัด คืออ่าน โค้ดคำสั่งมาบรรทัดหนึ่งแล้วก็ทำงานให้ผลออกมาเลย ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 การทำงานของอินเตอร์พรีเตอร์

จากรูปตัวอย่างในกรณีที่มีการเรียกใช้ฟังก์ชันจากไลบรารี (Library) หรือโมดูล (Module) ของภาษา Python อินเตอร์พรีเตอร์ของภาษา Python ก็จะไปทำการเรียกฟังก์ชันเหล่านั้นให้ทำงานแล้วจึงแสดงผลการทำงานออกมา

ในส่วนของประสิทธิภาพการทำงานนั้นตัวแปลภาษาแบบคอมไพเลอร์จะทำงานได้เร็วกว่าตัวแปลภาษาแบบอินเตอร์พรีเตอร์เพราะ โค้ดคำสั่งถูกคอมไพล์และลิงค์โดยตัวแปลภาษาแบบคอมไพเลอร์ผ่านแล้วได้เป็นไฟล์ .exe ออกมาต่อจากนั้นก็เป็นส่วนของการทำงานเพียงอย่างเดียว

อีกสิ่งหนึ่งที่ใช้ในการพัฒนา คือ Regular Expression (หรือ REs) นั้นเป็นภาษาโปรแกรมมิ่งชนิดพิเศษที่มีขนาดเล็ก โดยถูกฝังตัวไว้ภายใน Python โดยการใช้งาน Regular Expression นี้ต้องทำการระบุกฎเกณฑ์สำหรับกลุ่มของสตริงที่เป็นไปได้ที่ต้องการจะทำการตรวจจับ โดยที่กลุ่มสตริงเหล่านี้อาจจะเป็นประโยคภาษาอังกฤษ เช่น e-mail address เป็นต้น และยังสามารที่จะหาความเหมือนของรูปแบบสตริงที่กำหนดควอยู่ในตำแหน่งใดหรือใช้ REs ในการแก้ไขสตริง หรือแยกเป็นส่วนๆ ได้หลายวิธี โดยนำ Regular Expression มาใช้ในส่วนของการ application ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อต้องการที่จะดึงแท็ก Html ที่เกี่ยวข้องกับลิงค์ของรูปภาพภายในเว็บไซต์ต่างๆ โดย Regular Expression ที่ใช้นั้นภายในแต่ละเว็บนั้นอาจมีความแตกต่างกัน เช่น ตัวอย่างการดึงข้อมูลที่เป็นลิงค์ของรูปภาพ จะเขียนได้ดังนี้ `` สิ่งที่ได้ อาจเป็นลิงค์ของรูปภาพที่ไม่ครบตามต้องการเพราะยังมีปัญหาที่ต้องแก้ไขอีกต่อไปซึ่งเป็น Library ของภาษา Python ที่ไว้เรียกใช้งาน Regular Expression และเพื่อดึงส่วนของ URL ที่ต้องการ โดยเมื่อจะทำงานต้องใช้ฟังก์ชัน `match` ค่าของ Regular Expression จึงจะได้ลิงค์หรือ URL ที่ต้องการมาได้ และเมื่อต้องการค้นหาเว็บเพจที่เราต้องการอาจจะใช้รูปแบบหรืออาจเรียกเป็นฟังก์ชัน `Search` ค้นหา เมื่อทำการค้นหาแล้วผลที่ได้ออกมาที่เป็น URL หรือลิงค์ตามที่ต้องการในส่วนของการดึงแท็ก Html ที่เกี่ยวข้องกับ link ย่อยๆ หรือ Hyperlink ที่อยู่ลึกลงไปจากหน้าแรก ก็มีขั้นตอนที่คล้ายกัน

#### 2.6.4 ข้อดีของภาษา Python

- ง่ายต่อการเรียนรู้ โดยภาษา Python มีโครงสร้างของภาษาไม่ซับซ้อนเข้าใจง่าย ซึ่งโครงสร้าง ภาษา Python จะคล้ายกับภาษา C มาก เพราะภาษา Python สร้างขึ้นมาโดยใช้ภาษา C ทำให้ผู้ที่คุ้นเคยภาษา C อยู่แล้วใช้งานภาษา Python ได้ไม่ยาก นอกจากนี้โดยตัวภาษาเองมีความยืดหยุ่นสูงทำให้การจัดการกับงานด้านข้อความ และ Text File ได้เป็นอย่างดี
- ไม่ต้องเสียค่าใช้จ่ายใดๆ ทั้งสิ้น เพราะตัวแปลภาษา Python อยู่ภายใต้ลิขสิทธิ์ GNU
- ใช้ได้หลายแพลตฟอร์ม ในช่วงแรกภาษา Python ถูกออกแบบใช้งานกับระบบ Unix อยู่ก็จริง แต่ในปัจจุบันได้มีการพัฒนาตัวแปลภาษา Python ให้สามารถใช้กับระบบปฏิบัติการอื่นๆ อาทิเช่น Linux, Windows 95/98/ME, Windows NT, Windows 2000
- Python ถูกสร้างขึ้นโดยได้รวบรวมเอาส่วนดีของภาษาต่างๆเข้ามาไว้ด้วยกัน อาทิเช่น ภาษา C, C++, Java, Perl
- Python เป็นภาษาประเภท Server side Script คือการทำงานของภาษา Python จะทำงานด้านฝั่ง Server แล้วส่งผลลัพธ์กลับมายัง Client ทำให้มีความปลอดภัยสูง
- ใช้พัฒนา Web Service โดยที่ภาษา Python สามารถนำมาพัฒนาเว็บเซอร์วิส และสร้างเว็บสำเร็จรูปที่เรียกว่า Content Management Framework (CMF)

## 2.7 ภาษา PHP

PHP ย่อมาจาก "Hypertext Preprocessor" เป็นภาษา Server-Side Script อีกภาษาหนึ่งเช่นเดียวกับ ASP ที่มีการทำงานที่เครื่องคอมพิวเตอร์ฝั่ง Server ซึ่งรูปแบบในการเขียนคำสั่งการทำงานนั้นจะมีลักษณะคล้ายกับภาษา Perl หรือภาษา C และสามารถใช้ร่วมงานกันกับ ภาษา HTML ได้อย่างมีประสิทธิภาพเป็นภาษาที่ใช้เขียน Website ที่ได้พัฒนามาจาก Perl และ C ดังนั้น จึงมีความยืดหยุ่นในการเขียนเป็นอย่างมาก สามารถเรียนการใช้งาน library ต่างๆ ได้อย่างสะดวก และมีผู้ใช้งานเป็นจำนวนมาก ไม่น้อยไปกว่าภาษาอื่นเลย มีผู้นิยมใช้เป็นจำนวนมาก เป็นภาษาที่ฟรี เขียนง่าย มีผู้ให้บริการ Server อยู่มาก Compiler มีขนาดเล็ก มี library สำเร็จมากมายมีความเร็วสูงและมีการพัฒนาอย่างต่อเนื่อง ถูกออกแบบมาเพื่องาน Internet Programming โดยเฉพาะใช้ติดต่อกับ Database ได้เกือบทุกตระกูล

### 2.7.1 สาเหตุที่เลือกพัฒนาเว็บด้วย PHP

- ความรวดเร็วในการพัฒนาโปรแกรม

เพราะว่า PHP เป็นสคริปต์แบบ Embedded คือ สามารถแทรกร่วมกับภาษา HTML ได้อย่างอิสระ และหากเราพัฒนาโค้ดไว้ในรูปแบบของ Class ที่เขียนขึ้นเพียงครั้งเดียวแล้วเรียกใช้งานได้ตลอด ทำให้สะดวกและรวดเร็วต่อการพัฒนาโปรแกรมต่างๆ

- PHP เป็นโค้ดแบบเปิดเผย (Open Source)

คำว่า Open Source ทั่วไปแล้วก็มีความหมายเหมือนกับของฟรีนั่นเอง เนื่องจาก PHP มีกลุ่มของผู้ใช้งานอยู่เป็นจำนวนมากทั่วโลก และมีเว็บไซต์อยู่เป็นจำนวนมากที่เป็นแหล่งรวบรวมซอสโค้ด โปรแกรม หรือจะเป็นบทความต่างๆทำให้ผู้ใช้มือใหม่ๆ หรือผู้ที่ต้องการศึกษา สามารถค้นหาซอสโค้ดมาเป็นแนวทางในการพัฒนาโปรแกรมได้ง่ายขึ้น

- การบริหารหน่วยความจำ (Memory Usage)

มีการใช้งานหน่วยความจำที่ดี กล่าวคือ PHP จะไม่เรียกใช้หน่วยความจำตลอดเวลา ทำให้เซิร์ฟเวอร์ไม่จำเป็นต้องมีทรัพยากรมากนัก

- อิสระต่อระบบปฏิบัติการ

เว็บแอปพลิเคชันที่ถูกสร้างขึ้นมาสามารถที่จะรันได้หลายระบบปฏิบัติการไม่ว่าจะเป็น Unix, Linux หรือ Windows เป็นต้น

### 2.7.2 ความสามารถของ PHP

ความสามารถของ PHP นั้นสามารถที่จะทำงานเกี่ยวกับ Dynamic Web ได้ทุกรูปแบบ เหมือนกับการเขียนโปรแกรมแบบ CGI (Common Gateway Interface) หรือ ASP (Active Server Pages) ไม่ว่าจะเป็นด้านการดูแลจัดการระบบฐานข้อมูล ระบบรักษาความปลอดภัยของเว็บเพจ การรับ-ส่ง Cookies เป็นต้น

คุณสมบัติที่โดดเด่นที่สุดของ PHP น่าจะเป็นการติดต่อกับ โปรแกรมจัดการระบบฐานข้อมูลที่มีอยู่มากมาย ซึ่งฐานข้อมูลที่ PHP สนับสนุนมีดังนี้

Adabas D	InterBase Solid	Microsoft Access
dBase	mSQL	Sybase
Empress	MySQL	Velocis
FilePro	Oracle	Unix dbm
Informix	PostgreSQL	SQL Server

### ตารางที่ 2.1 ฐานข้อมูลที่ภาษา PHP รองรับ

แต่ความสามารถที่พิเศษกว่านี้คือ PHP สามารถที่จะติดต่อกับบริการต่างๆ ผ่านทางโปรโตคอล (Protocol) เช่น IMAP, SNMP, NNTP, POP3, HTTP และยังสามารถติดต่อกับ Socket ได้อีกด้วย

### 2.7.3 การทำงานของ PHP

ซึ่งเราสามารถแบ่งออกเป็น 6 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ฟังก์ชัน Client จะทำการร้องขอหรือเรียกใช้งานไฟล์ PHP ที่เก็บในเครื่อง Server

ขั้นตอนที่ 2 ฟังก์ชัน Server จะทำการค้นหาไฟล์ PHP ตัวที่ถูกร้องขอแล้วทำการประมวลผลไฟล์ PHP ตามที่ Client ทำการร้องขอมา

ขั้นตอนที่ 3 ทำการประมวลผลไฟล์ PHP

ขั้นตอนที่ 4 และ 5 เป็นการติดต่อกับฐานข้อมูลและนำข้อมูลในฐานข้อมูลมาใช้ร่วมกับการประมวลผล

ขั้นตอนที่ 6 ส่งผลลัพธ์จากการประมวลผลไปให้เครื่อง Client

#### 2.7.4 ข้อดีของภาษา PHP

- เป็น Opensource ดังนั้น จึงสามารถใช้งานฟังก์ชันต่างๆ ได้โดยไม่ต้องจ่ายเงินแต่ประการใด อีกทั้งยังการแพร่กระจายของ Free Source code ต่างๆ เป็นจำนวนมาก
- ฐานข้อมูลที่รองรับคือ (mySql) ซึ่งสามารถเรียนรู้ได้ง่าย และมีใช้เป็นจำนวนมาก
- รองรับการทำงานบน Server ทั้ง Windows และ linux เพราะการ Run เพียงแต่ต้องการ Program PHP บน server เท่านั้น

#### 2.7.5 ข้อเสียของภาษา PHP

- ไม่สามารถเรียกใช้งาน Webservice แบบ .Net ได้ ดังนั้น การใช้งานฟังก์ชันเพื่อดึงข้อมูลจาก Web อื่น จำต้องเขียน Code เองเอง Free Code ที่ให้ มักจะมีฟังก์ชันต่างๆ เกินความต้องการ ทำให้เกิดการหน่วงในการดึงข้อมูลจากฐานข้อมูลขึ้น อีกทั้งยังปรับแต่งตามความต้องการได้ยากมาก

### 2.8 ความรู้เบื้องต้นเกี่ยวกับระดับของสี

สี หมายถึง ลักษณะความเข้มของคลื่นแสงและมีความยาวคลื่นต่างกันที่ปรากฏต่อสายตา หรือสิ่งที่นำไปข้อมหรือนำวัตถุอื่นให้เป็นสีเช่นเดียวกับตัวมันเอง สีมียุทธศิลป์ในเรื่องของอารมณ์ การสื่อความหมายที่เด่นชัด กระตุ้นต่อการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีทำให้ความรู้สึก ความสัมพันธ์และอารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้น รุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการสร้างงานศิลปะและการออกแบบ เพื่อความสวยงาม และเพื่อสื่อความหมาย ฉะนั้นงานจะสมบูรณ์สวยงามดึงดูดใจผู้ดูได้ดี จะต้องมิสีเข้าไปเป็น องค์ประกอบสำคัญในการสร้างงาน ดังนั้นเราจึงควรศึกษาทฤษฎีสีและหลักการใช้สีอย่างละเอียด เพื่อเลือกใช้สีให้เหมาะสมกับงานแต่ละชนิด

สีเป็นองค์ประกอบหลักสำหรับการใช้งานในรูปแบบต่างๆ การเลือกใช้สีให้เหมาะสมกับ สภาพแวดล้อมของเว็บนั้น จำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจในหลักพื้นฐานของสีก่อนขั้นสี ซึ่งมีดังต่อไปนี้

#### 2.8.1 สีระดับที่ 1 (primary)

แม่สีที่เป็นหัวใจหลักของทฤษฎีสีประกอบด้วย 3 สีด้วยกัน คือ แดง เหลือง และน้ำเงิน ทั้ง 3 สีนี้ถือว่าเป็นแม่สีหลักระดับที่ 1 ที่ไม่ได้เกิดจากการผสมสีจากสีอื่น แต่เป็นสีต้นกำเนิดสีอื่น ๆ แต่ละสีเหล่านี้สามารถเปลี่ยนแปลงหรือปรับระดับค่าสีได้อย่างไม่มีที่สิ้นสุด

### 2.8.2 สีระดับที่ 2 (secondary)

สีในระดับที่ 2 นี้เป็นการผสมสีจากแม่สีระดับที่ 1 ที่อยู่ติดกันเพื่อให้เกิดสีใหม่ขึ้นมา โดยการนำเอา สีแดงผสมน้ำเงินเป็นสีม่วง สีแดงผสมเหลืองเป็นสีส้ม และสีเหลืองผสมน้ำเงินเป็นสีเขียว ภายในแต่ละสีระดับที่ 1 เหล่านี้ สามารถปรับเปลี่ยนหรือการปรับค่าของสีเพื่อจะได้สีแบบไม่มีที่สิ้นสุด

### 2.8.3 สีระดับที่ 3 (Tertiary)

สีระดับที่ 3 ซึ่งเกิดจากการนำเอาสีระดับที่ 1 และสีระดับที่ 2 ที่อยู่ติดกันทั้งสองด้านนำผสมกันเพื่อให้เกิดสีขึ้นมาอีกหนึ่งระดับ ทั้งหมด 6 สี ประกอบด้วยการผสมสี ดังนี้ คือ สีแดง-ส้ม แดง-ม่วง น้ำเงิน-ม่วง น้ำเงิน-เขียว เหลือง-เขียวและเหลือง-ส้ม ซึ่งมีสีรวมกันทั้งหมดจากการผสมสี ตั้งแต่ระดับที่ 1 ระดับที่ 2 และระดับที่ 3 มีทั้งหมด 12 สี แต่ยังมีสีที่นอกเหนือจากผสมของแม่สี คือ สีขาวและสีดำ

## 2.9 ข้อมูลรูปภาพ (Image Data)

ในการศึกษาเรื่องรูปภาพแบบต่างๆเราจะต้องรู้ก่อนว่าภาพแต่ละภาพมีความแตกต่างกันอย่างไร องค์ประกอบของภาพที่ใช้หรือแสดงออกมาทางหน้าจอเป็นอย่างไรเพราะฉะนั้นจึงต้องศึกษาข้อมูลที่เกี่ยวข้องกับรูปภาพ โดยในทำโครงการนี้ จะกล่าวถึงข้อมูลที่เกี่ยวข้องกับรูปแบบของภาพ การทำการประมวลผลรูปภาพต้องศึกษาพื้นฐานของรูปภาพก่อนว่ามีการสร้างขึ้นได้อย่างไร ซึ่งในรูปภาพแต่ละรูปจะมีองค์ประกอบต่างๆมากมายแต่องค์ประกอบที่จะนำมาประมวลผลนั้นเรียกว่า “พิกเซล (Pixel)” พิกเซลนั้นจะมีลักษณะเป็นจุดสี่เหลี่ยม ซึ่งเป็นองค์ประกอบเล็กๆที่มีอยู่มากมายบนภาพการประมวลผลภาพต้องนำพิกเซลแต่ละรูปภาพมาทำการประมวลผลโดยที่แต่ละพิกเซลจะประกอบด้วยรายละเอียดสีต่างๆมารวมกันเป็นพิกเซล ซึ่งเรียกกลุ่มของสีที่ใช้ว่า “Color Space” ซึ่งรูปแบบการบอกรายละเอียดของสีบนรูปภาพ โดยกลุ่มสีที่แสดงผลนั้นมีอยู่หลายแบบ

การเกิดภาพต่างๆ บนจอคอมพิวเตอร์จะเกิดจากการเรียงตัวของจุดสี่เหลี่ยมสีต่างๆที่เรียกว่า “พิกเซล” ซึ่ง “พิกเซล” เป็นตัวกำหนดความคมชัดของภาพในจอคอมพิวเตอร์ ถ้ากำหนดให้ภาพแต่ละภาพมีจำนวน “พิกเซล” มากภาพก็จะมีรายละเอียดมากขึ้นตามไปด้วย แต่จะทำให้ไฟล์มีขนาดใหญ่ขึ้นด้วยตามความละเอียดของภาพ

พิกเซล (Pixel มาจากคำว่า "Picture Element") เป็นหน่วยพื้นฐานของสีที่สามารถโปรแกรมได้บนจอแสดงผลหรือบนภาพ ขนาดทางกายภาพของพิกเซลจะขึ้นอยู่กับกำหนัดความละเอียดของจอภาพ ดังนั้นถ้าต้องการแสดงผลให้มีความละเอียดสูงสุดขนาดทางกายภาพของพิกเซลก็จะเท่ากับขนาดทางกายภาพของจุด (Dot Pitch หรือ Dot Size) บนจอภาพแต่ถ้าตั้งความละเอียดไว้ต่ำกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าสูงสุด พิกเซลจะมีขนาดใหญ่กว่าขนาดทางกายภาพของจุดบนจอภาพ นั่นคือแต่ละพิกเซลจะใช้มากกว่าหนึ่งจุดในการแสดงผล

ในปัจจุบันนี้กลุ่มสีที่แสดงผลมีมากมายหลายแบบ เช่น กลุ่มสีที่แสดงผล RGB เป็นกลุ่มสีที่แสดงผลที่บอกแสดงลักษณะของสี 3 สี คือ สีแดง สีเขียว และสีน้ำเงิน, กลุ่มสีที่แสดงผล HSV เป็นกลุ่มสีที่แสดงผลที่บอกลักษณะของภาพโดยใช้ ระดับสี, ความเข้มของสี, ความสว่างของสี เป็นต้น แต่ในงานวิจัยนี้เลือกใช้กลุ่มสีที่แสดงผล RGB ในการแสดงรายละเอียดของภาพ

### 2.9.1 คุณสมบัติของกลุ่มสีที่แสดงผล

โดยทั่วไปแล้วภาพจะอยู่ในรูปแบบของกลุ่มสีที่แสดงผล RGB ซึ่งแทนสีทั้งหมดในกลุ่มสีที่แสดงผลด้วยเวกเตอร์ 3 มิติ จึงใช้กลุ่มสีที่แสดงผลนี้เป็นจุดเริ่มต้นของลักษณะเฉพาะของภาพได้โดยใช้ลักษณะเฉพาะของสี (Color Feature) จากนั้นค่อยทำการลดระดับของสีให้เหมาะสมต่อการทำฮิสโตแกรม คุณสมบัติดังกล่าวคือ

#### 2.9.1.1 Uniformity

Uniformity เป็นคุณสมบัติที่ใช้ในการเปรียบเทียบความใกล้เคียงของเมตริกสีโดยกลุ่มสีที่แสดงผลจะมีคุณสมบัตินี้จะต้องคำนวณได้ง่าย อย่างเช่น การหาความใกล้เคียงของเมตริกสีโดยใช้ฟังก์ชันที่ไม่ขึ้นกับตำแหน่งในกลุ่มสีนั้น การทำการแปลงสีจะเป็นวิธีที่ทำให้กลุ่มสีมีคุณสมบัตินี้

#### 2.9.1.2 Completeness

Completeness เป็นคุณสมบัติสำคัญที่กลุ่มสีที่แสดงผลจะต้องมีสีที่สามารถแยกแยะได้อย่างชัดเจน สามารถทำได้โดยการลดระดับสี

#### 2.9.1.3 Compactness

Compactness เป็นคุณสมบัติที่สีในกลุ่มสีที่แสดงผลจะต้องแตกต่างจากสีอื่นๆ ในกลุ่มสีที่แสดงผลเดียวกัน คุณสมบัตินี้จะเป็นตัวกำหนดมิติของภาพเพื่อให้แน่ใจว่า เมื่อผ่านการเปรียบเทียบจากกลุ่มสีที่แสดงผล RGB แล้วจะไม่มีสี 2 สีที่เหมือนกัน

#### 2.9.1.4 Naturalness

Naturalness เป็นคุณสมบัติของผู้ใช้สามารถเข้าใจได้ง่าย ทำให้ผู้ใช้สามารถทำการดึงข้อมูลจากฐานข้อมูลโดยใช้สีเป็นดัชนีได้ ซึ่งสีดังกล่าวจะถูกแยกแยะออกเป็น 3 ส่วน คือ ความสว่าง (Brightness), ระดับสี (Hue), ความเข้มของสี (Saturation)

### 2.9.2 ระบบสีแบบ RGB

ระบบสี RGB เป็นระบบสีของแสง ซึ่งเกิดจากการหักเหของแสงผ่านแท่งแก้วปริซึม จะเกิดแถบสีที่เรียกว่า สเปกตรัม (Spectrum) ซึ่งแยกสีตามทิวสายตามองเห็นได้ 7 สี คือ แดง แสด เหลือง เขียว น้ำเงิน คราม ม่วง ซึ่งเป็นพลังงานอยู่ในรูปของรังสี ที่มีช่วงคลื่นที่สายตา สามารถมองเห็นได้ แสงสีม่วงมีความถี่คลื่นสูงสุด คลื่นแสงที่มีความถี่สูงกว่าแสงสีม่วง เรียกว่า อัลตราไวโอเล็ต (Ultra Violet) และคลื่นแสงสีแดง มีความถี่คลื่นต่ำที่สุด คลื่นแสง ที่ต่ำกว่าแสงสีแดงเรียกว่า อินฟราเรด (InfraRed) คลื่นแสงที่มีความถี่สูงกว่าสีม่วง และต่ำ กว่าสีแดงนั้น สายตาของมนุษย์ไม่สามารถรับได้ และเมื่อศึกษาดูแล้วแสงสีทั้งหมดเกิดจาก แสงสี 3 สี คือ สีแดง (Red) สีน้ำเงิน (Blue) และสีเขียว (Green) ทั้งสามสีถือเป็นแม่สีของแสง เมื่อนำมาฉายรวมกันจะทำให้เกิดสีใหม่ อีก 3 สี คือ สีแดงมาเจนน้ำ สีฟ้า ไซแอน และสีเหลือง และถ้าฉายแสงสีทั้งหมดรวมกันจะได้แสงสีขาว จากคุณสมบัติของแสงนี้ เรา ได้นำมาใช้ประโยชน์ทั่วไป ในการฉายภาพยนตร์การบันทึกภาพวิดีโอภาพโทรทัศน์ การสร้างภาพเพื่อการนำเสนอทางจอคอมพิวเตอร์ และการจัดแสงสีในการแสดง เป็นต้น

### 2.9.3 กลุ่มสีแสดงผลแบบ RGB

ภาพสีที่อยู่ในรูปแบบของ RGB จะเป็นรูปแบบพื้นฐานของภาพที่ใช้แสดงบนจอคอมพิวเตอร์ โดยแต่ละ พิกเซลของภาพจะประกอบด้วยองค์ประกอบของสี 3 ค่า คือสีแดง, สีเขียว, สีน้ำเงิน โดยแสดงองค์ประกอบสีสีละ  $n$  บิต ดังนั้นจะสามารถแสดงเฉดสีได้องค์ประกอบสีละ  $2^n$  เฉดสีที่แตกต่างกัน การผสมสีบนคอมพิวเตอร์นั้นอาศัยระดับสีหลัก 3 สีคือ สีแดง สีเขียว และ น้ำเงิน หรือที่ เรียกว่า “RGB Colors” (Red-Green-Blue) ความเหมือนจริงของสีคอมพิวเตอร์นั้น ขึ้นอยู่กับในหนึ่งจุดพิกเซลของการแสดงผลนั้น ใช้ระดับของสี (Color depths) ที่มีค่าเท่าไร เช่นถ้าสี RGB มีระดับของสีเป็น 8 planes นั่นคือ ใช้ 8 บิตเก็บข้อมูลหนึ่งสี หมายความว่า เฉพาะระดับสีหลัก เช่น สีแดง สีเขียวก็จะมี ความเป็นแดง อยู่ถึง  $2^8 = 256$  ระดับ อธิบายง่ายๆก็ คือ ถ้าจะผสมสี หนึ่งสี จาก แดง-เขียว-น้ำเงิน (RGB) จะใช้สีแดง ก็ส่วนจาก 0 ถึง 255 ส่วน ใช้เขียว ก็ส่วนจาก 0 ถึง 255 และเช่นเดียวกัน สีน้ำเงินก็ ส่วน จาก 0 ถึง 255 สีที่เกิดขึ้นก็จะเกิดจากการผสมของสีทั้งสาม ในอัตราส่วนต่างๆ กัน เช่น สีเหลืองธรรมดา เกิดจากการผสมสี แดง 255 ส่วน สีเขียว 255 ส่วน และสีน้ำเงิน 0 ส่วน จากข้างต้น ระดับของสีแดง มีถึง 256 ระดับ สีเขียว 256 ระดับ สีน้ำเงิน 256 ระดับ ดังนั้น RGB ทั้งหมดใช้ 24 บิต (8+8+8) ในการแสดงสี RGB ของหนึ่งจุดพิกเซลซึ่งสามารถแสดงสีได้มากถึง  $256 \times 256 \times 256 = 16.7$  ล้านสี

2.10 ฮิสโตแกรม (Histogram)



รูปที่ 2.4 แสดงกราฟของฮิสโตแกรม

ฮิสโตแกรม เป็นกราฟแสดงจำนวนพิกเซลที่ความสว่างต่างๆ ของภาพ สังเกตได้จากรูปที่ 2.4 แกนนอนเป็นระดับความสว่างที่แบ่งระดับเป็น 256 ระดับ (มักเรียกว่าระดับสีเทา หรือ gray level) โดยมีค่าตั้งแต่ 0-255 เมื่อระดับสีเทามีค่าต่ำ (ด้านซ้ายมือ) หมายถึง มีความสว่างน้อย จะมองเห็นเป็นสีดำ ค่าระดับสีเทาสูง (ด้านขวามือ) หมายถึง มีความสว่างมากจะมองเห็นเป็นสีขาว แกนตั้งของกราฟแสดงจำนวนพิกเซลในแต่ละความระดับสีเทาซึ่งเป็นค่าสัมพัทธ์

ฮิสโตแกรมของสีประกอบด้วยระดับต่างๆของสี ซึ่งระดับต่างๆนี้จะมีจำนวนของสีในแต่ละระดับที่ปรากฏในภาพ ฮิสโตแกรมของภาพสามารถอธิบายความมืด ความสว่าง ความคมชัดสูงหรือต่ำ ฮิสโตแกรมของสีจะมีความทนทานต่อการเปลี่ยนมุมมองของภาพเนื่องจากการเปลี่ยนมุมมองจะทำให้จำนวนพิกเซลรวมของสีไม่เกิดการเปลี่ยนแปลงมากนัก แต่ถ้ารูปร่างของภาพเมื่อมีการเปลี่ยนแปลงมุมมองแล้วจะเกิดการเปลี่ยนแปลงไปมาก ฮิสโตแกรมจะเป็นเสมือนฟังก์ชันในการกำหนดความน่าจะเป็นของสีที่ปรากฏในภาพ ซึ่งค่าของสีเดียวกันจะต้องอยู่ในระดับของสีเดียวกัน แต่ฮิสโตแกรมของสีอย่างเดียวยังไม่สามารถใช้ในการค้นหารูปภาพได้อย่างมีประสิทธิภาพ เนื่องจากว่าฮิสโตแกรมของสีไม่สามารถบอกได้ว่าระดับสีในแต่ละระดับอยู่ในตำแหน่งใดของภาพ อย่างไรก็ตาม ฮิสโตแกรมกลับได้รับความนิยมในการเป็นดัชนีของภาพ เพราะส่วนใหญ่แล้วพบว่าถ้าสีที่ปรากฏในรูปภาพมีความมืด เนื่องจากปัจจัยแวดล้อมต่างๆ ทำให้กลุ่มของสีที่ปรากฏในภาพเปลี่ยนตำแหน่งไป แต่ผลกระทบที่มีต่อ ฮิสโตแกรมของสีจะมีเพียงเล็กน้อยเท่านั้น ซึ่งทำให้การค้นหาภาพยังคงมีประสิทธิภาพที่ดีได้

72948

### 2.10.1 การ Quantization

ในงานวิจัยที่ได้จัดทำนี้การสร้างสีสโตแกรมของสีเพื่อเก็บสีที่ปรากฏอยู่ในภาพโดยได้แบ่งค่า RGB ของรูปภาพออกเป็น 216 ระดับ ( $R = 6$  ระดับ,  $G = 6$  ระดับ,  $B = 6$  ระดับ) เพื่อลดระดับความซับซ้อนของสีและเป็นการประหยัดเนื้อที่ในการจัดเก็บข้อมูลสีสโตแกรมของแต่ละภาพและทำให้ง่ายต่อการทำสีสโตแกรม

ช่วงของสี	ค่าที่ใช้แทน
0 – 42	0
43 – 85	1
86 – 128	2
129 – 171	3
172 – 213	4
214 - 255	5

ตารางที่ 2.2 ระดับของสีที่ใช้ในการทำ Quantization

เริ่มด้วยต้องการหาว่าในพิกเซลที่สนใจนั้นประกอบไปด้วยค่าของสีอะไรบ้าง(RGB)เท่าใดแล้วนำค่าที่ได้ไปเปรียบเทียบกับตารางว่าค่าของพิกเซลที่สนใจนั้นอยู่ในช่วงสีไหน แล้วใช้ค่า 0-5 แทนการแสดงด้วยค่าระดับสีที่แท้จริง ในสีแต่ละสีจะแบ่งค่าออกเป็นทั้งหมด 6 ช่วงสี ซึ่งเดิมจะถูกแบ่งไว้ทั้งหมด 255 สี เพราะฉะนั้นสามารถแสดงระดับสีจริงๆ ที่ผสมสีทั้ง 3 แล้วได้ทั้งหมด 216 ระดับ ตั้งแต่ค่า  $R=0, G=0, B=0$  จนถึง  $R=6, G=6, B=6$  เช่น ถ้า  $R = 130, G = 79, B = 127$  ก็จะถูก quantized เป็น  $R = 3, G = 1, B = 2$

### 2.10.2 การ Normalization

เนื่องจากขนาดของรูปภาพจะมีผลทำให้จำนวนพิกเซลในแต่ละรูปภาพไม่เท่ากัน ส่งผลให้ค่าสีสโตแกรมของรูปภาพ 2 รูปภาพที่มีขนาดไม่เท่ากันจึงไม่สามารถจะนำมาเปรียบเทียบกันได้ ดังนั้นจึงจำเป็นต้องทำการ Normalization เพื่อให้สามารถนำมาเปรียบเทียบกันได้ โดยเมื่อหาค่าสีสโตแกรมของรูปภาพได้แล้วจะต้องนำความถี่ของแต่ละระดับสีมาหารด้วยจำนวนพิกเซลทั้งหมดที่มีในภาพ ผลที่ออกมาจะได้ค่าของความน่าจะเป็น (Probability) ของแต่ละระดับสีที่มีในรูปภาพ

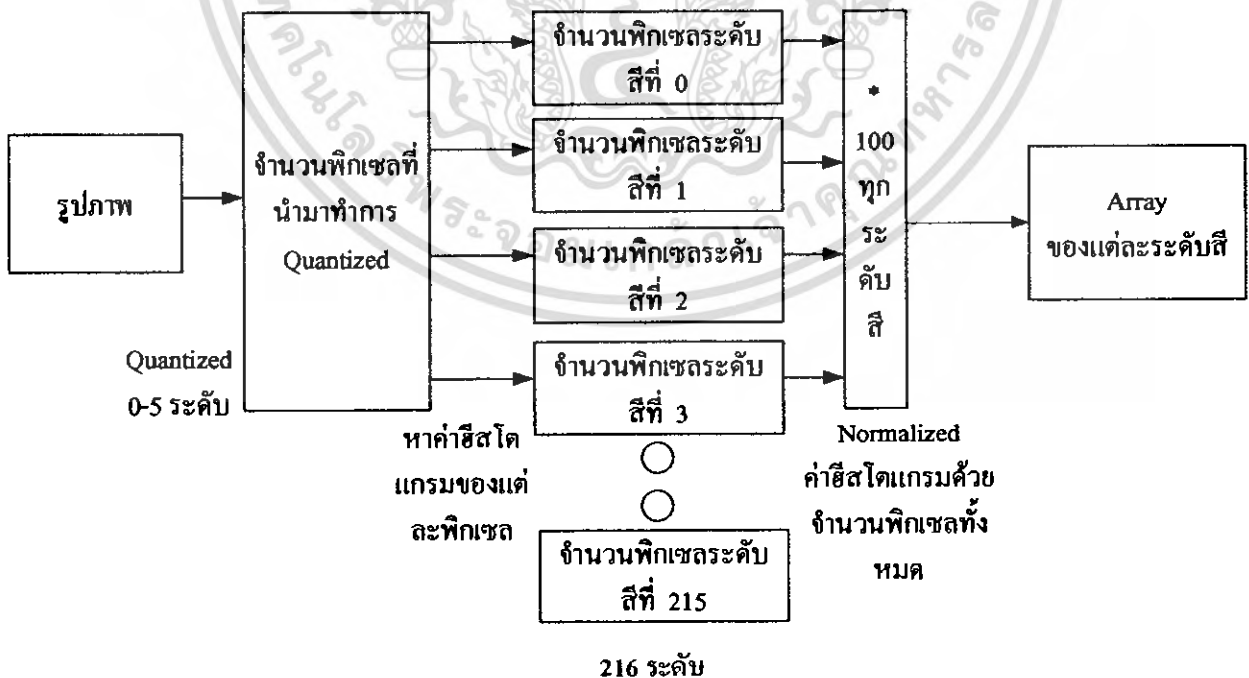
$$p(n) = \frac{f(n)}{P \times Q}$$

โดย

- p(n) - ค่าความน่าจะเป็นของจำนวนพิกเซลที่มีระดับสีที่ n
- f(n) - ค่าความถี่ของพิกเซลที่มีในระดับสีที่ n
- PxQ - จำนวนพิกเซลทั้งหมดที่มีในรูปภาพได้จากการนำจำนวนพิกเซลตามความกว้างคูณกับจำนวนพิกเซลตามความยาว

### 2.10.3 ขั้นตอนการทำฮิสโตแกรม

หลังจากที่เลือกรูปภาพแล้วก็นำแต่ละภาพไปทำการ Quantization ซึ่งเป็นการลดระดับของสีลงเพื่อให้ง่ายต่อการคำนวณ โดยในงานวิจัยนี้จะแบ่งเป็น 6 ระดับของแต่ละสี คือ R 6 ระดับ G 6 ระดับ และ B 6 ระดับ รูปภาพที่ถูกทำ Quantization แล้วก็นำมาหาฮิสโตแกรมโดยนำแต่ละพิกเซลในรูปภาพมาดูว่าอยู่ในระดับสีที่เท่าไร แล้วเก็บเป็นความถี่สะสมในแต่ละระดับสี เมื่อสแกนดูทุกพิกเซลในรูปภาพเสร็จแล้วก็จะนำความถี่สะสมของแต่ละระดับสีมาทำการ Normalized โดยการนำความถี่สะสมของแต่ละระดับสีมาหารด้วยจำนวนพิกเซลทั้งหมด แล้วจะได้ผลลัพธ์ออกมาเป็น ความน่าจะเป็นของแต่ละระดับสี และนำค่าความน่าจะเป็นของแต่ละระดับสีคูณกับ 100 เพื่อให้ได้เป็นค่าเปอร์เซ็นต์ (%) หลังจากนั้นก็ปัดให้เป็นเลขจำนวนเต็มไปเก็บไว้ในอาร์เรย์เพื่อใช้ในการค้นหารูปภาพต่อไป



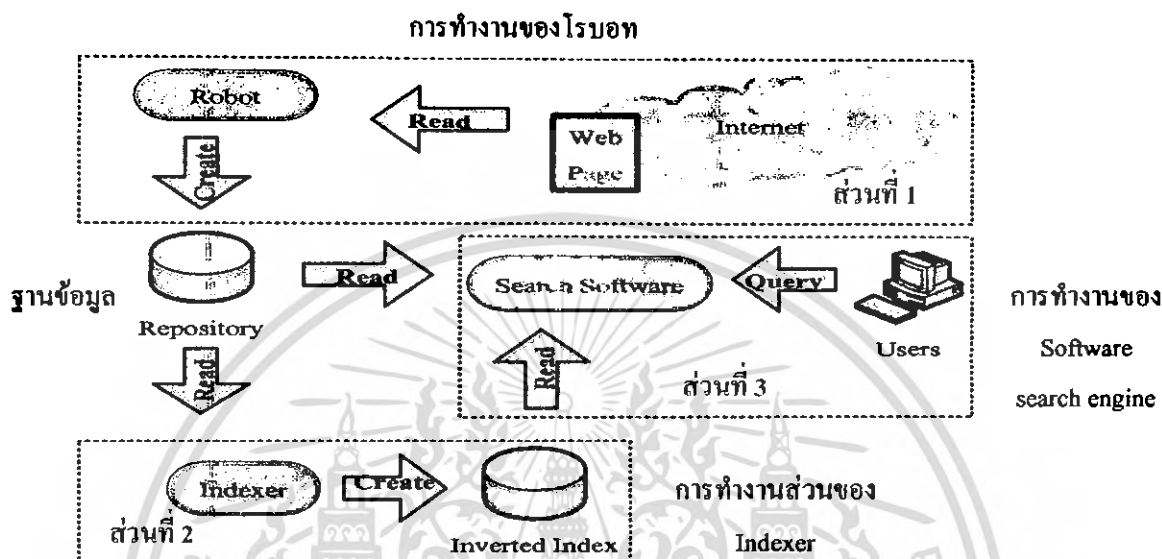
รูปที่ 2.5 ขั้นตอนการทำฮิสโตแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### การค้นหารูปภาพโดยใช้เสิร์จเอนจิน

#### 3.1 การออกแบบและพัฒนาระบบเสิร์จเอนจิน



รูปที่ 3.1 ภาพรวมของระบบเสิร์จเอนจินที่ออกแบบ

การทำงานของเสิร์จเอนจินแบ่งได้ 3 ส่วน คือ

ส่วนที่ 1 โปรแกรมไหลคลิกของรูป และไหลครูปโดยมีการทำงานแบบเทรค (โรบอท)

โรบอท มีหน้าที่ค้นหาลิงค์จากเว็บเพจที่อยู่ในเครือข่ายอินเทอร์เน็ตและนำมาเก็บไว้ในฐานข้อมูลเพื่อจะนำไปทำดัชนีการค้นหาต่อไป

ส่วนที่ 2 โปรแกรมสร้างดัชนีการค้นหา (อินเด็กเซอร์) มีหน้าที่สร้างดัชนีค้นหาจากไฟล์ที่

โรบอทหามาโดยเลือกวิธี Inverted index ซึ่งเป็นวิธีที่ได้รับความนิยมมากที่สุด

ส่วนที่ 3 เสิร์จเอนจินซอฟต์แวร์ ทำหน้าที่ค้นหาข้อมูลจากฐานข้อมูล จะเริ่มต้นการทำงานเมื่อผู้ใช้ป้อนข้อมูลที่ต้องการค้นหาโดยโปรแกรมจะทำหน้าที่นำข้อมูลการค้นหาของผู้ใช้มาจับคู่กับดัชนีในฐานข้อมูลแล้วดึงข้อมูลที่ตรงกับคำค้นหาออกมาและทำการเปรียบเทียบความเหมือนโดยใช้หลักการฮีสโตแกรม จากนั้นจะแสดงผลให้ผู้ใช้ทราบ

### 3.2 การออกแบบฐานข้อมูลในระบบเสิร์จเอนจิน

การออกแบบฐานข้อมูล แบ่งเป็น 2 ตาราง คือ table\_link\_web และ table\_link\_image

#### 3.2.1 ตารางการเก็บข้อมูล table\_link\_web

Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
1	http://www.kapook.com	0000-00-00	0	Load
2	http://movie.kapook.com/	2007-03-06	1	Load
3	http://game.kapook.com/	2007-03-06	1	Load
4	http://music.kapook.com/	2007-03-06	1	Load
5	http://mobile.kapook.com/	2007-03-06	1	Load
6	http://dookdik.kapook.com/	2007-03-06	2	Load
7	http://kookkik.kapook.com/	2007-03-06	2	Load
8	http://planet.kapook.com	2007-03-06	2	Load
9	http://guide.kapook.com/	2007-03-06	3	Load
10	http://football.kapook.com/	2007-03-06	3	empty
11	http://gamecenter.kapook.com	2007-03-06	3	empty
12	https://sonic2.tac.co.th/bp3/bin/Index	2007-03-06	4	empty

ตารางที่ 3.1 ตาราง table\_link\_web

โดยตาราง table\_link\_web เป็นตารางที่ใช้เก็บลิงค์ของเว็บที่โรบอทมาจากเครือข่ายอินเทอร์เน็ต โดยจะนำลิงค์ที่ได้เก็บลงฟิลด์ Url\_Web และเก็บข้อมูลของวันที่ได้ทำการเก็บลิงค์ โดยจะนำมาแสดงที่ฟิลด์ Date\_Web ส่วนฟิลด์ Id\_Chis เป็นฟิลด์ที่ใช้บอกถึงลิงค์ย่อยที่ได้มาจากฟิลด์ Id\_Web ใด และฟิลด์ Load\_Web เป็นส่วนที่ใช้บอกสถานะว่าลิงค์ของเว็บเพจนั้นได้โหลดแล้วหรือยัง ไม่ได้โหลดถ้าโหลดแล้วมีค่าเท่ากับ Load ถ้ายังไม่มีการโหลดจะมีค่าเท่ากับ empty โดยจะเช็คว่าการที่จะนำลิงค์ใดมาโหลดนั้นฟิลด์ Load\_Web จะต้องมีค่าเท่ากับ empty และ Id\_Web จะต้องมีค่าน้อยที่สุดจึงจะนำค่าใน Url\_Web ที่มีค่า Id\_Web ที่น้อยที่สุดมาโหลด

สำหรับข้อมูลที่ได้จากรางนี้ ในส่วนของฟิลด์ Url\_Web เป็นลิงค์ที่ได้จากโรบอท ซึ่งโรบอทจะค้นหาจากเว็บเพจภายในเครือข่ายอินเทอร์เน็ตแล้วนำมาเก็บในฐานข้อมูล โดยลิงค์ที่เก็บมานั้นเป็นลิงค์ของรูปภาพ ซึ่งมีความจำเป็นต่อการนำไปใช้ในการโหลดรูปต่อไปโดยแต่ละฟิลด์ มีความสำคัญดังนี้

- Id\_Web - เป็นฟิลด์ตัวเลขที่มีหน้าที่เพื่อกำหนดหมายเลขของลิงค์
- Url\_Web - เป็นฟิลด์ที่ใช้แสดงชื่อลิงค์ของเว็บที่โหลดจากเว็บเพจ
- Date\_Web - เป็นฟิลด์ที่แสดงวันที่ ที่โรบอทได้เก็บตามลิงค์มาจากอินเทอร์เน็ต
- Id\_Chis - เป็นฟิลด์ที่บอกถึงลิงค์ย่อยที่ได้มาจากตัว Id\_Web
- Load\_Web - เป็นฟิลด์ที่ใช้บอกสถานะว่าลิงค์ของเว็บเพจนั้นได้โหลดแล้วหรือยังไม่ได้โหลดถ้าโหลดแล้วมีค่าเท่ากับ Load ถ้ายังไม่มีการ โหลดจะมีค่าเท่ากับ empty

ตาราง table\_link\_web ใช้ Id\_Web และ Url\_Web เป็น Primary Key ร่วมกัน

### 3.2.2 ตารางการเก็บข้อมูล table\_link\_image

Id_Image	Url_Image	Histogram	File_Name	Date_Image	Id_Web_Chis
1	http://www.kapook.com/images3/mobileportal.jpg	[5, 0, 0, 0, 0, 0, 3, 2,...	Image1.jpg	2007-03-06	1
2	http://www.kapook.com/images3/dookdikportal.jpg	[20, 0, 0, 0, 0, 0, 10,...	Image2.jpg	2007-03-06	1
3	http://www.kapook.com/images3/oscardodi.jpg	[9, 0, 0, 0, 0, 0, 6, 0,...	Image3.jpg	2007-03-06	1
4	http://www.kapook.com/images3/apartment1dodi.jpg	[7, 0, 0, 5, 0, 1, 0, 0,...	Image4.jpg	2007-03-06	2
5	http://www.kapook.com/images3/thequeen.jpg	[41, 0, 0, 0, 0, 0, 57,...	Image5.jpg	2007-03-06	2
6	http://www.kapook.com/images3/Living-Forest.jpg	[0, 0, 22, 0, 0, 1, 3,...	Image6.jpg	2007-03-06	2
7	http://www.kapook.com/images3/thebodyguard2.jpg	[0, 0, 0, 0, 0, 11, 3,...	Image7.jpg	2007-03-06	3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ **ตารางที่ 3.2 ตาราง table\_link\_image** อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง table\_link\_image เป็นตารางที่ใช้เก็บลิงค์รูปภาพและใช้แสดงชื่อลิงค์ของรูปภาพที่ได้จากเว็บเพจหลังจากการ โหลดรูป โดยจะแสดงชื่อลิงค์ที่ฟิลด์ของ Url\_Image และใช้เก็บค่าฮิสโตแกรมที่หาจากรูปที่ได้โหลดที่ฟิลด์ Histogram ส่วนฟิลด์ Date\_Image ใช้แสดงวันที่ ที่โรบอทได้เก็บตามลิงค์มาจากอินเทอร์เน็ต เป็นฟิลด์ที่บอกถึงลิงค์ย่อยที่ได้มา Id\_Web\_Chis จากตัว Id\_Web โดยข้อมูลจะแสดงได้ดังตารางที่ 3.2

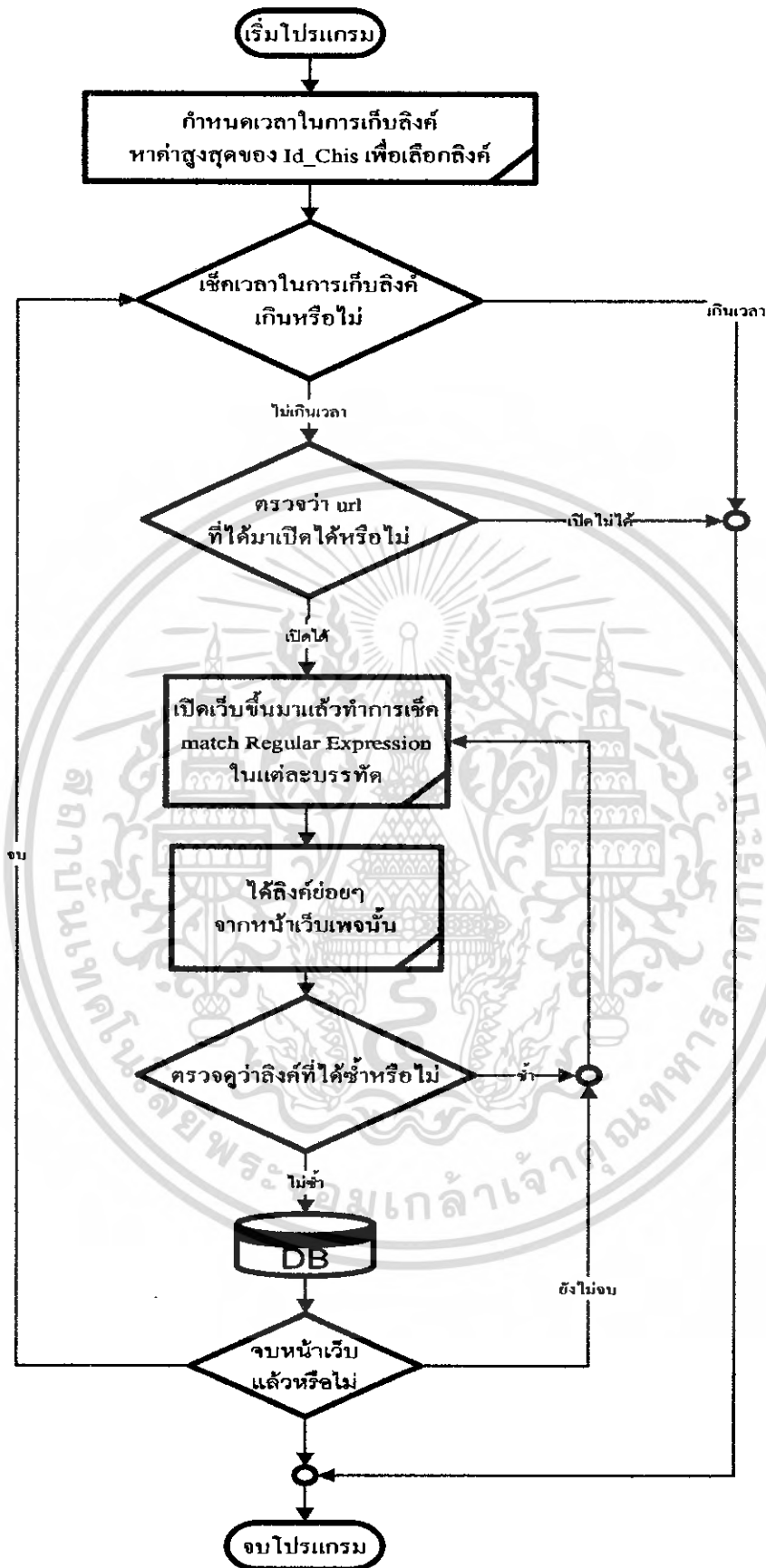
ตาราง table\_link\_image ฟิลด์แต่ละส่วนมีความสำคัญดังนี้

Id_Image	-	เป็นฟิลด์ตัวเลขที่กำหนดหมายเลขของลิงค์ในฐานข้อมูล
Url_Image	-	เป็นฟิลด์ที่ใช้แสดงชื่อลิงค์ของรูปภาพที่ได้จากเว็บเพจ
Histogram	-	เป็นฟิลด์ตัวเลขที่ใช้เก็บค่าฮิสโตแกรมของภาพเพื่อใช้เปรียบเทียบ
Date_Image	-	เป็นฟิลด์ที่แสดงวันที่ ที่โรบอทได้เก็บตามลิงค์มาจากอินเทอร์เน็ต
Id_Web_Chis	-	เป็นฟิลด์ที่บอกถึงลิงค์ย่อยที่ได้มา จากตัว Id_Web

ตาราง table\_link\_image ใช้ Id\_Image และ Url\_Image เป็น Primary Key ร่วมกัน

### 3.3 การทำงานของเสิร์จเอนจินเพื่อทำการเก็บลิงค์ของเว็บ

ระบบเสิร์จเอนจินจะทำงาน โดยใช้โปรแกรม Web crawler ทำหน้าที่ในการเก็บลิงค์ตามเว็บเพจต่างๆ โดยจะมีการเช็คเรื่องเวลาด้วยว่าถ้าเข้าไปภายในหน้าเว็บเพจแล้วไม่พบข้อมูลที่เป็นข้อมูลรูปภาพหรือว่าลิงค์ของรูปจะมีการจับเวลาแล้วถ้าจับเวลาตามที่ได้กำหนดไว้แล้วยังไม่สามารถเข้าไปยังลิงค์นั้นได้ โปรแกรมก็จะหยุดการทำงานจากการเก็บข้อมูลจากหน้าเว็บเพจนั้นทันที เพื่อแก้ไขปัญหการวนลูปหรือวิ่งเก็บลิงค์ที่ซ้ำๆ แล้วจะกลับไปเริ่มทำงานที่ลิงค์อื่นแทน และเมื่อพบข้อมูลที่เป็นได้คของลิงค์ก็จะทำการตัดเอาส่วนของ URL ของลิงค์นั้นไปเก็บไว้ในฐานข้อมูลซึ่งการทำงานในส่วนนี้สามารถสร้าง Web crawlers หลายๆ ตัวหรือการทำงานเป็นเทรคเพื่อช่วยกันทำงาน ซึ่งจะเก็บลิงค์รูปภาพเรื่อยๆ จนกระทั่งครบหน้าเว็บเพจนั้น จากนั้นก็จะทำการเข้าไปยังหน้าถัดไปโดยเลือกจากค่า URL ที่เก็บไว้ในฐานข้อมูล ซึ่งจะทำการเก็บข้อมูลไปเรื่อยๆ ตามข้อมูลในฐานข้อมูลขั้นตอนการทำงานสามารถแสดงในลักษณะ flowchart ได้ดังรูปที่ 3.2



รูปที่ 3.2 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.3.1 การทำงานของ Web crawler ที่ใช้เก็บลิงค์ของเว็บและลิงค์ในระดับถัดไป

การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ของเว็บสามารถแบ่งเป็นขั้นตอนได้ดังนี้

ขั้นตอนที่ 1 โปรแกรมทำการกำหนดค่าของเวลาในการเก็บลิงค์

ขั้นตอนที่ 2 โปรแกรมทำการเลือกลิงค์ในฟิลด์ Url\_Web โดยที่ฟิลด์ Load\_Web มีค่าเท่ากับ empty และฟิลด์ Id\_Web จะต้องมีค่าน้อยที่สุด แล้วจึงนำลิงค์ของฟิลด์ใน Url\_Web นั้นมาทำการเปิดเพื่อทำงานต่อไป

Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
1	http://www.kapook.com	0000-00-00	0	Load
2	http://movie.kapook.com/	2007-03-06	1	Load
3	http://game.kapook.com/	2007-03-06	1	Load
4	http://music.kapook.com/	2007-03-06	1	Load
5	http://mobile.kapook.com/	2007-03-06	1	Load
6	http://dookdik.kapook.com/	2007-03-06	2	Load
7	← http://kookkik.kapook.com/	2007-03-06	2	empty
8	http://planet.kapook.com	2007-03-06	2	empty

รูปที่ 3.3 แสดงค่า empty เพื่อหาลิงค์ที่จะนำมาโหลด

การที่จะนำลิงค์ใดมาโหลดนั้นฟิลด์ Load\_Web จะต้องมีค่าเท่ากับ empty และ Id\_Web จะต้องมีค่าน้อยที่สุด เช่นรูปที่ 3.3 ฟิลด์ Load\_Web ที่มีค่าเป็น empty จะตรงกับฟิลด์ Id\_Web ที่ 7 และ Id\_Web ที่ 7 ก็จะมี Url\_Web เป็น http://kookkik.kapook.com/ เพราะฉะนั้นจึงทำให้ทราบได้ว่า จะเริ่มทำงานจากลิงค์ใด

ขั้นตอนที่ 3 โปรแกรมทำการเช็คเวลาที่ใช้ในการเก็บลิงค์หมดหรือยัง ถ้ายังไม่หมดให้ทำขั้นตอนที่ 4 ต่อไป แต่ถ้าหมดแล้วให้จบโปรแกรมทันทีโดย

ขั้นตอนที่ 4 ทำการเช็ค URL ที่อ่านมาว่าสามารถเปิดได้หรือไม่เว็บเพจนั้นปิดไปแล้วหรือไม่ ถ้าไม่สามารถเปิดได้ก็จะจบฟังก์ชันการทำงานจากนั้นโปรแกรมก็จะกลับไปอ่านค่า URL ถัดไปของตาราง table\_link\_web อีกครั้ง

ขั้นตอนที่ 5 เปิดหน้าเว็บเพจนั้นขึ้นมาและทำการตรวจสอบการ match regular expression เพื่อที่จะได้ส่วน URL ที่ต้องการมาโดยรูปแบบการ match คำของ Regular Expression ที่ 1 คือ

```
re.compile("""href=(("[S+]*[A[a-zA-Z0-9-_*]+]*(.php|.com|.shtml|.html|.htm|.cgi|.th)*[^\.^pdf|^\.doc|^\.swf|^\.ico|^\.css])*(|)"")""", re.IGNORECASE)
```

โดยส่วนนี้เป็นส่วนที่สำคัญที่สุดในการ match คำเพื่อให้ได้ URL ที่ตรงกับความต้องการมากที่สุด โดยแต่ละเว็บในเครือข่ายอินเทอร์เน็ตจะใช้รูปแบบการ match ที่แตกต่างกันออกไป ส่วนนี้จึงเป็นปัญหาที่มีความสำคัญมาก

ส่วนรูปแบบการทำงานของการทำงานของการเก็บลิงค์ในระดับที่สองนั้นการ match คำในการเก็บลิงค์ในระดับนี้มีรูปแบบการ match คำของ Regular Expression เหมือนกับการเก็บลิงค์ในระดับที่แรก โดยรูปแบบที่ใช้ match ลิงค์ย่อยในหน้าเว็บเพจนั้นคือ href รูปแบบที่ใช้ดังตัวอย่าง

```
""""href=(("[S+]*[A[a-zA-Z0-9-_*]+]*(.php|.com|.shtml|.html|.htm|.cgi|.th)*[^\.^pdf|^\.doc|^\.swf|^\.ico])*(|)"")"""
```

ลักษณะการทำงานของการทำงานของการเก็บลิงค์ในระดับที่สองจะคล้ายกับการเก็บลิงค์ในระดับแรกโดยมีส่วนการทำงานที่แตกต่างกันคือ ส่วนของการ match regular expression รูปแบบที่ใช้จะมีลักษณะที่ต่างกันส่วนการทำงานขั้นตอนอื่นจะคล้ายกันทุกขั้นตอน

รูปแบบการค้นหาหน้าเว็บเพจนั้นฟังก์ชันที่ใช้ คือ search โดยรูปแบบของ search คือ re.search (<pattern>, line) โดยที่ <pattern> เป็นรูปแบบที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาในแต่ละบรรทัดในข้อส ไลค์ของหน้าเว็บเพจแต่ละเว็บเพจเพื่อค้นหาลิงค์ที่อยู่ในระดับถัดไป

ในการ match คำของ Regular Expression ที่ 1 เพื่อค้นหาลิงค์ของเว็บที่สามารถ  
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งยังมีเหตุที่เปลี่ยนแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ลำดับที่ 1 href =("|") อธิบายได้ว่า จะต้องขึ้นต้นด้วย href = และตามด้วยเครื่องหมาย (') หรือ (") เท่านั้น
- ลำดับที่ 2 [S+]\* อธิบายได้ว่าจะมีช่องว่างหรือไม่มีเลขก็ได้ แต่ถ้ามีจะต้องมีอย่างน้อย 1 ช่องหรือมากกว่า 1 ช่อง
- ลำดับที่ 3 [^[a-zA-Z0-9- ]+]\* อธิบายได้ว่า เครื่องหมาย (/) และอักขระจะมีหรือไม่มีเลขก็ได้ แต่ถ้ามีจะต้องมีอย่างน้อย 1 ตัวหรือมากกว่า 1 ตัว
- ลำดับที่ 4 (.php|.com|.shtml|.html|.htm|.cgi|.th)\* อธิบายได้ว่า มีอักขระ .php หรือ .com หรือ .shtml หรือ .html หรือ .htm หรือ .cgi หรือ .th จะมีหรือไม่มีเลขก็ได้ แต่ถ้ามีจะต้องเรียงลำดับกัน
- ลำดับที่ 5 [^.pdf^.doc |^.swf^.ico] อธิบายได้ว่า อักขระ .pdf หรือ.doc หรือ.swf หรือ .ico จะต้องไม่มีเลข
- ลำดับที่ 6 (/)\* อธิบายได้ว่า ช่องว่างหรือเครื่องหมาย (/) จะมีหรือไม่มีเลขก็ได้
- ลำดับที่ 7 (|")\* อธิบายได้ว่า เครื่องหมาย (') หรือ (") จะมีหรือไม่มีเลขก็ได้

จากการ match ค่าของ Regular Expression ทำให้เราสามารถค้นหาลิงค์ของเว็บที่อยู่ภายในเว็บเพจได้ ซึ่งตัวอย่างลิงค์ที่สามารถ match ได้นั้นมีตัวอย่างดังต่อไปนี้

href="/index.php"

href="http://photo.kapook.com/?action=k2.app.photo.user&user\_id=80302"

href="http://event.kapook.shtml"

href="http://planet.kapook.com/thesims"

href="http://www.thannews.th.com"

href="http://www.tj.co.th"

ขั้นตอนที่ 6 เปิดหน้าเว็บเพจนั้นขึ้นมาและทำการตรวจสอบการ match regular expression เพื่อที่จะได้ส่วน URL ที่ต้องการมาโดยรูปแบบการ match ค่าของ Regular Expression คือ

```
re.compile("""src=(|")*[S+]*[^[a-zA-Z0-9-
_]+](.jpg|.bmp)(/|")*""", re.IGNORECASE)
```

และรูปแบบของ Regular Expression ในการ match ลิงค์รูปภาพในหน้าเว็บเพจนั้น คือ src โดยมีรูปแบบดังตัวอย่าง

```
""src="(\\")*[\\S+]*[\\/[a-zA-Z0-9-_]+]*(.jpg|.bmp)(/)*(\\")*""
```

รูปแบบการค้นหาหน้าเว็บเพ็จนั้นฟังก์ชันที่ใช้ คือ search โดยรูปแบบของ search คือ re.search (<pattern>, line) โดยที่ <pattern> เป็นรูปแบบที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาในแต่ละบรรทัด ในข้อสไลด์ของหน้าเว็บเพ็จนั้นแต่ละเว็บเพ็จนั้นเพื่อค้นหาลิงค์ที่อยู่ในระดับถัดไป

ในการ match คำของ Regular Expression ในการค้นหาลิงค์ของรูปนั้นสามารถอธิบายออกได้เป็นตามลำดับได้ดังต่อไปนี้

- ลำดับที่ 1 src="(\\")\* อธิบายได้ว่า จะต้องขึ้นต้นด้วย src= และตามด้วยเครื่องหมาย (\\) หรือ (") ซึ่งเครื่องหมายนี้จะไม่มีหรือไม่มีเลขก็ได้
- ลำดับที่ 2 [\\S+]\* อธิบายได้ว่าจะมีช่องว่างหรือไม่มีเลขก็ได้ แต่ถ้ามีจะต้องมีอย่างน้อย 1 ช่องหรือมากกว่า 1 ช่อง
- ลำดับที่ 3 [\\/[a-zA-Z0-9-\_]+]\* อธิบายได้ว่า เครื่องหมาย (/) และอักขระจะมีหรือไม่มีเลขก็ได้ แต่ถ้ามีจะต้องมีอย่างน้อย 1 ตัวหรือมากกว่า 1 ตัว
- ลำดับที่ 4 (.jpg|.bmp) อธิบายได้ว่า จะต้องใช้อักขระ .jpg หรือ .bmp เท่านั้น โดยจะต้องเรียงลำดับกันด้วย
- ลำดับที่ 5 (/)\* อธิบายได้ว่า ช่องว่างหรือเครื่องหมาย (/) จะมีหรือไม่มีเลขก็ได้
- ลำดับที่ 6 (\\")\* อธิบายได้ว่า เครื่องหมาย (\\) หรือ (") จะมีหรือไม่มีเลขก็ได้

จากการ match คำของ Regular Expression ทำให้เราสามารถค้นหาลิงค์ของรูปที่อยู่ในเว็บเพ็จนั้นได้ ซึ่งตัวอย่างลิงค์ที่สามารถ match ได้นั้นมีตัวอย่างดังต่อไปนี้

```
src="images3/mobileportal.jpg"
```

```
src="images3/Living-Forest.jpg"
```

```
src="/movie_con_04.bmp"
```

```
src="http://planet.kapook.com/files/kapookhome/hollabakgirl.jpg"
```

```
src="http://movie.kapook.com/images/submovies_banner4_7.bmp"
```

ขั้นตอนที่ 7 ทำการเช็คในฐานข้อมูลว่า URL ลิงค์นั้นเคยมีในฐานข้อมูลแล้วหรือไม่ ถ้าเคยมีแล้วจะทำการเช็คเวลาจากฐานข้อมูลของลิงค์นั้น โดยจะเช็คจากฟิลด์ Date\_Image ในตาราง table\_link\_image ว่ามีความห่างของระยะเวลาในการเก็บลิงค์เกินกว่าที่กำหนดหรือไม่ เช่น มีการกำหนดระยะห่างของการเก็บลิงค์ซ้ำมากกว่า 60 วัน และหาระยะเวลาในการเก็บข้อมูลเกินจากระยะเวลาที่กำหนดก็จะเก็บลิงค์ลงในฐานข้อมูล แต่หากระยะเวลาในการเก็บลิงค์ไม่เกิน ก็จะกลับไปทำขั้นตอนที่ 5 อีกครั้งเพื่อหา URL ลิงค์ในบรรทัดต่อไป ถ้ายังไม่เคยมีจะทำการเก็บลงในฐานข้อมูลจนครบ

จากรูปที่ 3.4 จะเห็นได้ว่าฟิลด์ `Id_Chis` สามารถบอกได้ว่าลิงค์ของเว็บระดับถัดไป นั้นมาจากลิงค์ใดในเว็บเพจ ส่วนรูปที่ 3.5 แสดงลิงค์ของระดับถัดไปว่ามาจากลิงค์ของเว็บ เพจใดเช่นกันกับรูปที่ 3.4

Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
1	http://www.kapook.com	0000-00-00	0	Load
2	http://movie.kapook.com/	2007-03-06	1	Load
3	http://game.kapook.com/	2007-03-06	1	Load
4	http://music.kapook.com/	2007-03-06	1	Load
5	http://mobile.kapook.com/	2007-03-06	1	Load
6	http://dookdik.kapook.com/	2007-03-06	2	Load
7	http://kookkik.kapook.com/	2007-03-06	2	empty
8	http://planet.kapook.com	2007-03-06	2	empty

รูปที่ 3.4 แสดงลิงค์ของระดับถัดไปว่ามาจากลิงค์ฟิลด์ `Id_Web` ใดในตาราง `table_link_web`

Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
1	http://www.kapook.com	0000-00-00	0	Load

ลิงค์ของเว็บ



Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
2	http://movie.kapook.com/	2007-03-06	1	Load
3	http://game.kapook.com/	2007-03-06	1	Load
4	http://music.kapook.com/	2007-03-06	1	Load
5	http://mobile.kapook.com/	2007-03-06	1	Load

ลิงค์ระดับถัดไป

รูปที่ 3.5 แสดงลิงค์ของระดับถัดไปว่ามาจากลิงค์ของเว็บเพจใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.6 จะแสดงให้เห็นว่าฟิลด์ Id\_Web\_Chis ในตาราง table\_link\_image สามารถบอกได้ว่าลิงค์ของภาพนั้นมาจากลิงค์เว็บเพจใดในตาราง table\_link\_web

Id_Web	Url_Web	Date_Web	Id_Chis	Load_Web
1	http://www.kapook.com	0000-00-00	0	Load
2	http://movie.kapook.com/	2007-03-06	1	Load

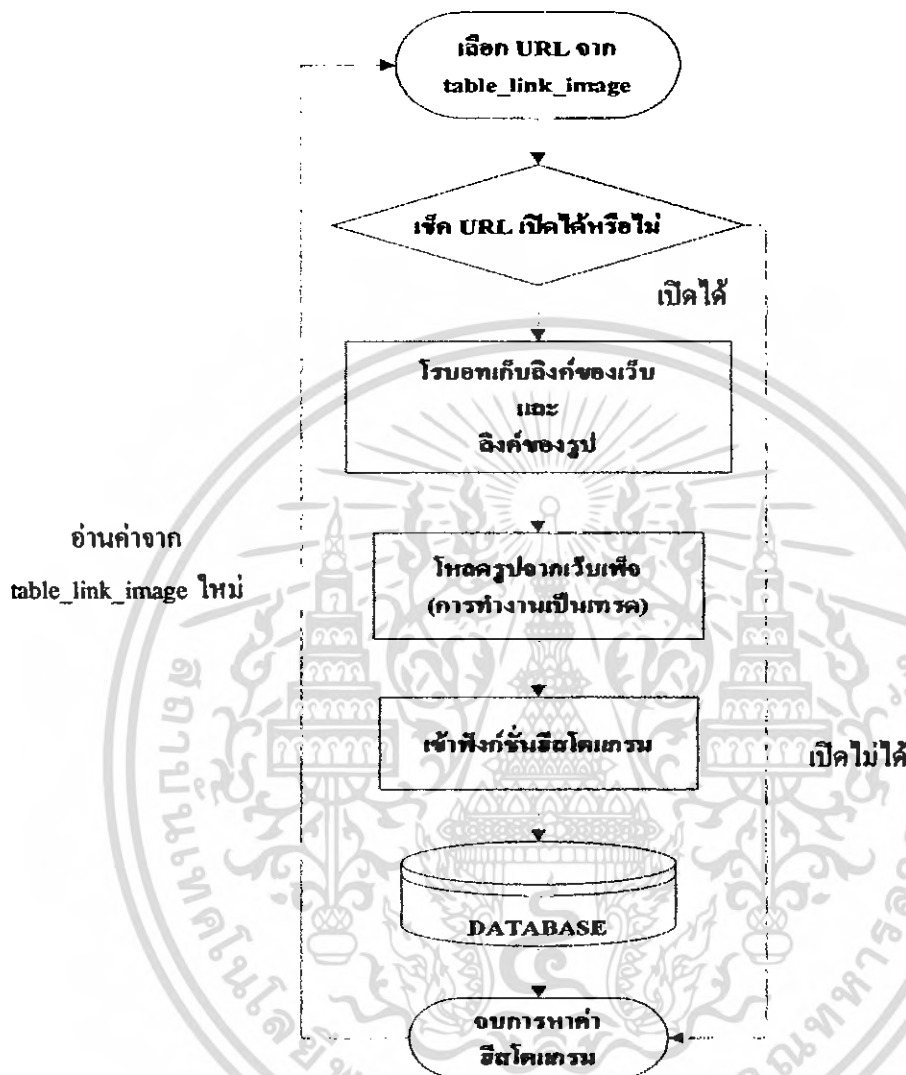
Id_Image	Url_Image	Histogram	File_Name	Date_Image	Id_Web_Chis
1	http://www.kapook.com/images3/mobileportal.jpg	[5, 0, 0, 0 , 8, 0, 3, 2,....	Image1.jpg	2007-03-06	1
2	http://www.kapook.com/images3/dookdikportal.jpg	[20, 0, 0, 0 , 0, 0, 10,....	Image2.jpg	2007-03-06	1
3	http://www.kapook.com/images3/oscardodi.jpg	[9, 0, 0, 0 , 0, 0, 6, 0,....	Image3.jpg	2007-03-06	1
4	http://www.kapook.com/images3/apartment1dodi.jpg	[7, 0, 0, 5 , 0, 1, 0, 0,....	Image4.jpg	2007-03-06	2
5	http://www.kapook.com/images3/thequeen.jpg	[41, 0, 0, 0 , 0, 0, 57,....	Image5.jpg	2007-03-06	2
6	http://www.kapook.com/images3/Living-Forest.jpg	[0, 0, 22, 0 , 0, 1, 3,....	Image6.jpg	2007-03-06	2
7	http://www.kapook.com/images3/thebodyguard2.jpg	[0, 0, 0, 0 , 0, 11, 3,....	Image7.jpg	2007-03-06	3

รูปที่ 3.6 แสดงลิงค์ของภาพว่ามาจากลิงค์ของเว็บเพจใดในตาราง table\_link\_web

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.4 การทำงานของเสิร์จเอนจินเพื่อทำการโหลดรูปภาพมาคำนวณหาค่าฮีสโตแกรม

#### 3.4.1 หลักการทำงาน



รูปที่ 3.7 ขั้นตอนการทำงานของเสิร์จเอนจินที่ใช้ในการคำนวณหาค่าฮีสโตแกรม

การทำงานของเสิร์จเอนจินที่ใช้ในการหาคำนวณค่าฮีสโตแกรมแบ่งเป็นขั้นตอนได้ดังนี้

ขั้นตอนที่ 1 โปรแกรมอ่านค่า URL จากตาราง `table_link_image`

ขั้นตอนที่ 2 เช็ค ว่า URL ที่ได้มาสามารถเปิดได้หรือไม่เว็บเพจนั้นปิดหรือไม่ ในกรณีที่เปิดลิงค์จากเว็บเพจนั้นไม่ได้โปรแกรมก็จะจบการทำงานทันที จากนั้นก็จะกลับไปอ่าน URL จากตาราง `table_link_image` อีกครั้งเพื่อค้นหาลิงค์ถัดไป

ขั้นตอนที่ 3 เมื่อพบลิงค์ที่สามารถเก็บได้การทำงานของโปรแกรมส่วนนี้จะแบ่งการทำงาน

ออกเป็น 2 ส่วนคือ ส่วนที่ทำหน้าที่เก็บลิงค์ของเว็บ และส่วนที่ทำหน้าที่โหลดรูปมาเก็บที่ฐานข้อมูล เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





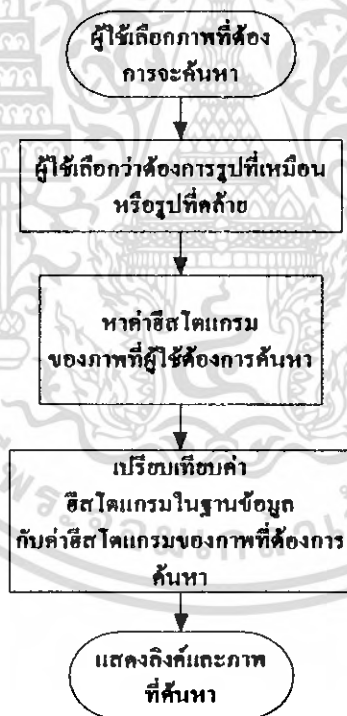




นั้นสามารถทำได้โดยนำอาเรย์ในทุกตำแหน่งของค่าฮิสโตแกรมทั้งสองภาพมาเปรียบเทียบความแตกต่าง โดยค่าความแตกต่างทุกๆอาเรย์ของค่าฮิสโตแกรมที่ได้ทำการเปรียบเทียบทั้งสองภาพจะต้องไม่เกินค่า Differential ที่ได้กำหนดเอาไว้ เมื่อค่า Differential เข้าใกล้ศูนย์มากเพียงใดจะทำให้ได้ภาพที่มีความเหมือนของภาพมากเท่านั้น จึงสามารถทำให้ทราบว่าควรจะใช้ค่าของ Differential มีค่าเท่าใดจึงจะเหมาะที่จะใช้ในการเปรียบเทียบรูปภาพซึ่งดูได้จากผลการทดลองการค้นหาข้อมูล โดยใช้ค่า Differential 0-5 จากบทที่ 4 หัวข้อที่ 4.4.1-4.4.6 ได้

จากตัวอย่างการทดลองข้างต้นจะทราบได้ว่าค่าฮิสโตแกรมของภาพในชุดทดลองเดียวกันจะมีค่าที่ใกล้เคียงกัน นั่นคือ ลักษณะสีของแต่ละภาพจะมีความใกล้เคียงกัน ในทุกๆลักษณะสีจะต่างกันเพียงไม่กี่ค่าเปอร์เซ็นต์ จากข้อมูลที่ได้นี้จะนำไปพัฒนาเป็นกระบวนการค้นหารูปภาพ หรือนำมาเปรียบเทียบเพื่อหาลักษณะเฉพาะของรูปแต่ละรูป จากนั้นจึงนำมาทำการเก็บลงในฐานข้อมูลเพื่อนำไปใช้งานในกระบวนการต่อไป

### 3.5 การค้นหารูปภาพของผู้ใช้



รูปที่ 3.13 ขั้นตอนการทำงานของผู้ใช้ในการค้นหารูปภาพ

ขั้นตอนการทำงานของผู้ใช้ในการค้นหารูปภาพมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 ผู้ใช้เลือกรูปที่ต้องการจะค้นหาเข้ามายังเซิร์ฟเวอร์เพื่อค้นหา

ขั้นตอนที่ 2 ผู้ใช้เลือกเมนูว่าต้องการได้รูปที่ใหม่หรือรูปที่คล้าย โดยจะมีให้เลือกที่หน้าเว็บ

ขั้นตอนที่ 3 หาค่าฮิสโตแกรมของภาพที่ค้นหาเพื่อหาลักษณะเฉพาะของรูปภาพนั้น

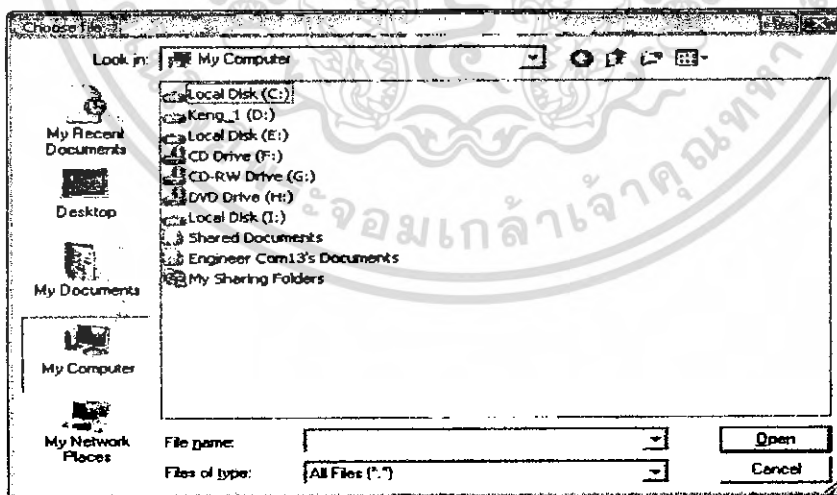
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 4 นำค่าฮิสโตแกรมที่ได้ไปเปรียบเทียบกับค่าฮิสโตแกรมในฐานข้อมูลโดยใช้ค่า Differential ในการเปรียบเทียบเพื่อหารูปที่คล้ายหรือรูปที่เหมือน จากนั้นผู้ใช้งานจะได้รูปที่เหมือนหรือมีความคล้ายตามที่ผู้ใช้งานต้องการ โดยค่า Differential ได้กล่าวไว้ในหัวข้อที่ 3.5.2

จากเครื่องผู้ใช้งานทำการเลือกรูปภาพจากเครื่องของผู้ใช้เองเป็นอินพุตผ่านทางหน้าเว็บเพื่อใช้ค้นหาว่ามีภาพที่เหมือนกันหรือคล้ายกัน โดยจะมีฟังก์ชันให้เลือกว่าต้องการรูปแบบใด โดยมีอยู่ 2 ฟังก์ชัน คือ เหมือน กับ ใกล้เคียง เมื่อผู้ใช้อัพโหลดรูปที่ต้องการจะค้นหาแล้ว โปรแกรมจะทำงานตามส่วนต่างๆ เมื่อการทำงานในแต่ละส่วนเสร็จสิ้นและค้นหารูปภาพผู้ใช้งานพบตามลิงค์ต่างๆ ก็จะแสดงรูปที่ผู้ใช้งานต้องการให้ผู้ใช้งานได้ทราบและจะแสดงลิงค์ของภาพดังกล่าว และภาพที่อยู่ตำแหน่งนั้นขึ้นมาให้ผู้ใช้งานดู

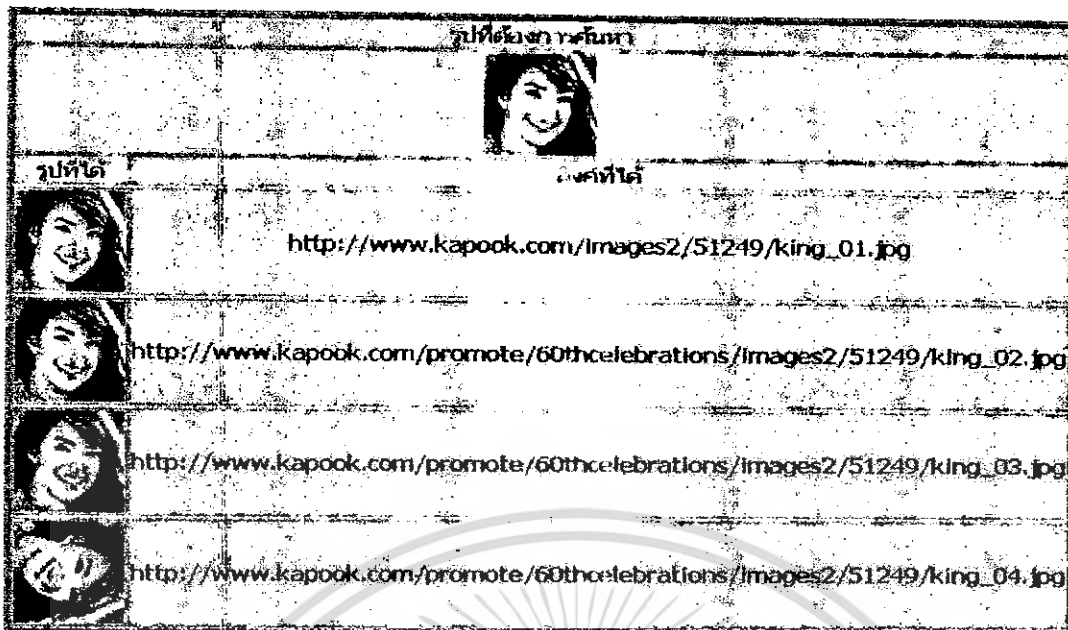


รูปที่ 3.14 หน้าจอของผู้ใช้งาน



รูปที่ 3.15 ผู้ใช้เลือกไฟล์รูปภาพที่ต้องการค้นหาจากเครื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



### รูปที่ 3.16 แสดงรูปภาพที่ผู้ใช้ต้องการค้นหา

เมื่อผู้ใช้ไม่ได้ไฟล์รูปที่ค้นหาหากกดปุ่ม Search โดยไม่ได้เลือกไฟล์จะแจ้งให้ทราบเมื่อทำการค้นหาว่า “กรุณาเลือกไฟล์ที่ต้องการค้นหา” และหากไฟล์รูปภาพที่เลือกค้นหา มีขนาดใหญ่เกิน 1Mb จะแจ้งว่าไฟล์มีขนาดใหญ่เกินไป โดยจะมีคำเตือนว่า “ไฟล์มีขนาดใหญ่เกิน 1Mb กรุณาเลือกไฟล์ที่ต้องการค้นหาอีกครั้ง”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### การทดลองและผลการทดลอง

#### 4.1 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้เก็บลิงค์ของเว็บ

ในการทดลองการเก็บลิงค์ของโปรแกรม Web Crawler เพื่อทดสอบประสิทธิภาพโดยการทดลองจะแบ่งการทำงานออกเป็นโปรเซสย่อย ซึ่งแต่ละโปรเซสจะแยกการทำงานออกเป็นดังนี้ โปรเซสแรกจะเริ่มเก็บลิงค์ที่อยู่ในเว็บเพจที่ค้นหาจากนั้นนำมาเก็บในฐานข้อมูลเมื่อได้ลิงค์ตามเว็บเพจแล้วจะนำลิงค์ที่ได้นั้นมาค้นหาลิงค์ของรูปภาพ ส่วนโปรเซสที่สองจะโหลดรูปจากลิงค์ที่ได้ค้นหาและโปรเซสสุดท้ายทำการหาค่าฮีโดแกรมจากนั้นจึงนำค่าที่ได้เก็บลงฐานข้อมูล จากการทดลองได้ผลการทดลองดังนี้จากตารางที่ 4.1 และตารางที่ 4.2 จะเห็นว่าลิงค์ที่ได้นั้นมีจำนวนมากว่ารูปที่ทำการโหลดมา เนื่องจาก บางลิงค์เป็นลิงค์ที่เว็บ Web Crawler ได้เก็บมานั้นเป็นลิงค์ที่ไม่สามารถโหลดรูปมาได้เนื่องจากเป็นลิงค์ที่ตาย (deadlink) จึงทำให้บางลิงค์ไม่สามารถโหลดรูปมาได้ การทดลองจะทดลองเก็บลิงค์ทุกๆ 10 นาทีเป็นระยะเวลา 2 ชั่วโมง โดยในฐานข้อมูลจะมีลิงค์ของรูปและลิงค์ของเว็บอยู่ด้วยเพื่อให้ได้ผลการทดลองจากการทำงานที่ดีที่สุดของแต่ละโปรแกรมซึ่งภายในฐานข้อมูลจะมีลิงค์ของเว็บอยู่แล้วจำนวน 3,000 ลิงค์การเก็บลิงค์จะแสดงดังตารางที่ 4.1

จำนวนลิงค์	เวลา (นาที)
5,230	10
6,655	20
7,773	30
8,580	40
10,054	50
11,875	60
12,460	70
14,570	80
17,287	90
19,650	100
20,207	110
21,178	120

ตารางที่ 4.1 ทดสอบการทำงานของโปรแกรม Web Crawler ที่ใช้เก็บลิงค์ของเว็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองการเก็บลิงค์ของ Web Crawler จะเห็นได้ว่าจำนวนลิงค์ของเว็บจะเพิ่มจำนวนขึ้นตามระยะเวลาซึ่งโปรเซสที่ใช้ค้นหาลิงค์ตามเว็บเพจจะค้นหาลิงค์ด้วยความเร็ว 176 ลิงค์/นาที

#### 4.2 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้เก็บลิงค์ของภาพ

จำนวนลิงค์ของรูปภาพ	เวลา (นาที)
900	10
1,650	20
2,350	30
2,676	40
3,087	50
3,692	60
4,183	70
4,570	80
4,757	90
4,960	100
5,265	110
5,567	120

ตารางที่ 4.2 ทดสอบการทำงานของโปรแกรม Web Crawler ที่ใช้เก็บลิงค์ของภาพ

จากการทดลองจะเห็นได้ว่าโปรเซสที่ใช้ค้นหาลิงค์ของรูปภาพมีความสามารถในการค้นหาลิงค์ของรูปภาพด้วยความเร็วประมาณ 46 ลิงค์/นาที ซึ่งปัญหาที่เกิดขึ้นในการทำงานของโปรเซสนี้คือ ลิงค์ที่ได้ค้นหานั้นบางลิงค์อาจจะค้นหาไม่ได้เนื่องจากลิงค์บางลิงค์เป็นลิงค์ที่เปิดไม่ได้จึงไม่สามารถค้นหาหรือเก็บลิงค์ดังกล่าวมาได้และความเร็วในการเก็บลิงค์นั้นอาจจะขึ้นอยู่กับความเร็วของอินเทอร์เน็ตด้วย ปัญหาอีกสิ่งหนึ่งคือ การค้นหาลิงค์ของเว็บ และลิงค์ของรูปภาพ จะมีความเร็วในการค้นหาไม่คงที่ขึ้นกับเว็บไซต์ที่ใช้เริ่มต้น ในการค้นหา เพราะว่าในแต่ละเว็บเพิ่จนั้นมีจำนวนของลิงค์ของเว็บและลิงค์ของรูปภาพเป็นจำนวนมากหรือน้อยแตกต่างกันจำนวนที่ได้ก็มีความแตกต่างกัน

### 4.3 การทดลองเพื่อทดสอบการทำงานของเสิร์จเอนจินที่ใช้คำนวณค่าฮิสโตแกรม

จำนวนชุดฮิสโตแกรมของภาพ	เวลา (นาทิต)
1,200	1
2,530	2
3,820	3
4,635	4
5,700	5
6,805	6
7,855	7
8,905	8
10,200	9
11,060	10

ตารางที่ 4.3 ทดสอบการทำงานของ โปรแกรมที่ใช้คำนวณหาค่าฮิสโตแกรม

โปรเซสในการทำฮิสโตแกรมสามารถทำงานด้วยความเร็วประมาณ 1,106 ภาพ/นาทิตโดยบางภาพเมื่อไม่สามารถเก็บลิงค์ของรูปภาพได้จึงไม่สามารถหาค่าฮิสโตแกรมได้ โดยจากตารางที่ 4.3 จะพบว่าปัญหาที่เกิดขึ้นจากการหาค่าฮิสโตแกรมได้ไม่เท่ากับจำนวนรูปคือ รูปจากลิงค์นั้นไม่สามารถดึงข้อมูลจากลิงค์นั้นมาได้เป็นลิงค์ที่ตาย (Deadlink)

จากการทดลองทั้ง 3 ตาราง จะเห็นความแตกต่างได้อย่างชัดเจน โดยโปรเซสที่ทำหน้าที่ค้นหาลิงค์ของเว็บและลิงค์ของรูปภาพจะมีความแตกต่างกันคือจำนวนลิงค์ของภาพจะมีจำนวนน้อยกว่าลิงค์ของเว็บ โดยในการเก็บลิงค์บางลิงค์ไม่สามารถเก็บมาได้หรือเป็นลิงค์ที่ตายจำนวนที่ได้จึงมีความแตกต่างกันโดยจะสามารถเช็คได้จากฐานข้อมูลว่าลิงค์ใดไม่สามารถดึงข้อมูลออกมาได้ ในส่วนของโปรเซสที่ทำงานโปรเซสที่ทำงานได้รวดเร็วที่สุดคือ โปรเซสที่ใช้คำนวณหาค่าฮิสโตแกรม รองลงมาคือ โปรเซสค้นหาลิงค์ของเว็บและเป็นโปรเซสค้นหาลิงค์ของรูปภาพตามลำดับ

#### 4.4 การทดลองเพื่อทดสอบค่าพารามิเตอร์ที่ใช้ค้นหาภาพที่เหมือน คล้ายและใกล้เคียง

การทดสอบหาค่า Differential นั้นก็เพื่อจะทดสอบว่ารูปภาพแต่ละรูปนั้นเมื่อมีขนาดที่ต่างกันหรือรูปที่เหมือนกันแต่มีลักษณะของรูปที่ต่างกันจะต้องทำการทดลองเพื่อหาค่าที่สามารถใช้ในการหาภาพที่เหมือนหรือคล้ายกันให้มีประสิทธิภาพมากที่สุด เนื่องจากภาพในเว็บส่วนใหญ่อาจมีขนาดเล็กทำให้ค่าฮิสโตแกรมเปลี่ยนไป ดังนั้นจึงทำการทดลองค่า Differential ตั้งแต่ 0-5 เพื่อให้ได้ค่า Differential ที่เหมาะสมที่สุดในการค้นหาและเมื่อได้เปรียบเทียบออกมาแล้วจะทำให้เราทราบว่าค่า Differential ที่ต่างกัมนั้นจะสามารถค้นหาภาพได้กี่ภาพและภาพที่ได้มีความคล้ายกันมากน้อยเพียงใด โดยภาพที่ใช้ทดลองนั้นจะเตรียมไว้ในฐานข้อมูลแล้วโดยมีภาพในฐานข้อมูลที่ปะปนอยู่ทั้งหมด 600 รูป ซึ่งการทดลองแต่ละชุดทดลองจะมีการลดขนาดของรูปหรือหมุนรูปในลักษณะต่างๆ เช่น ลดขนาดของรูปจาก 100 % เหลือ 60% , 50% ตามลำดับและหมุนรูปซ้ำๆ เพื่อให้รูปมีความแตกต่างกันออกไปจากนั้นจะทดสอบหาค่าความแตกต่างระหว่างรูปเพื่อนำไปใช้ค้นหารูปที่มีความเหมือนความคล้ายโดยการทดลองมีดังต่อไปนี้

จากที่ได้กล่าวไว้ในบทที่ 3 หัวข้อที่ 3.5.2 นั้นการหาค่า Differential นั้นจะเปรียบเทียบอารีย์ทุกตำแหน่งของค่าฮิสโตแกรมทั้งของภาพ เพื่อใช้วัดความเหมือนของภาพที่ต้องการเปรียบเทียบหรืออาจพูดได้ว่าค่า Differential นั้นเป็นค่าที่แตกต่าง(Differential)กันในแต่ละค่าลักษณะสี เพราะฉะนั้นจึงได้ทดลองหาค่า Differential ที่เหมาะสมที่สุดเพื่อนำไปใช้ในการค้นหารูปภาพว่าช่วงใดของค่า Differential คือตั้งแต่ 0-5 นั้นช่วงใดเหมาะสมที่สุดสำหรับภาพที่เหมือน ภาพที่คล้ายมาก และภาพที่คล้ายโดยผลการทดลองเป็นดังนี้

##### 4.4.1 การค้นหาข้อมูลโดยใช้ค่า Differential = 0



รูปที่ 4.1 ภาพชุดทดลองที่ 1



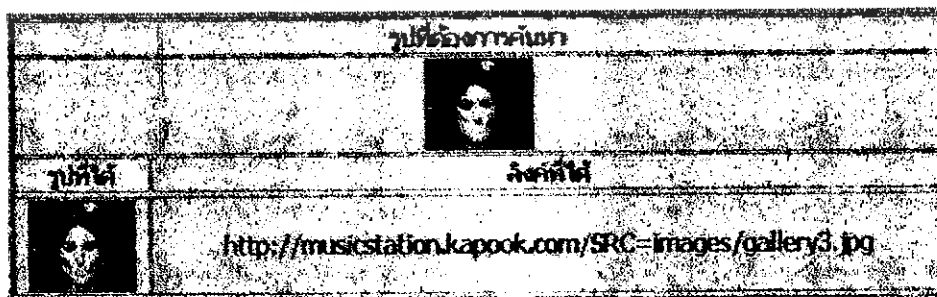
รูปที่ 4.2 ภาพชุดทดลองที่ 2



รูปที่ 4.3 ภาพชุดทดลองที่ 3



รูปที่ 4.4 ภาพชุดทดลองที่ 4



รูปที่ 4.5 ภาพชุดทดลองที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การใช้ค่า Differential = 0 จากการทดลองจะเห็นได้ว่ารูปที่ทำการค้นหานั้นสามารถค้นหาได้เฉพาะรูปที่เหมือนกับรูปต้นแบบเพียงอย่างเดียวเท่านั้น โดยรูปที่ไม่สามารถค้นหาได้นั้นเป็นรูปที่มีขนาดที่แตกต่างจากรูปต้นแบบหรือมีลักษณะที่ต่างกัน การทดลองโดยใช้ค่า Differential = 0 นี้จึงให้ผลเป็นรูปที่เหมือนและตรงตามความต้องการมากที่สุด

#### 4.4.2 การค้นหาข้อมูลโดยใช้ค่า Differential = 1



รูปที่ 4.6 ภาพชุดทดลองที่ 1

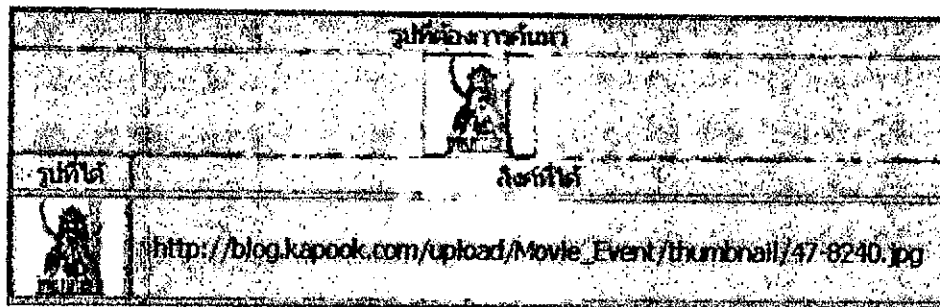


รูปที่ 4.7 ภาพชุดทดลองที่ 2



รูปที่ 4.8 ภาพชุดทดลองที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



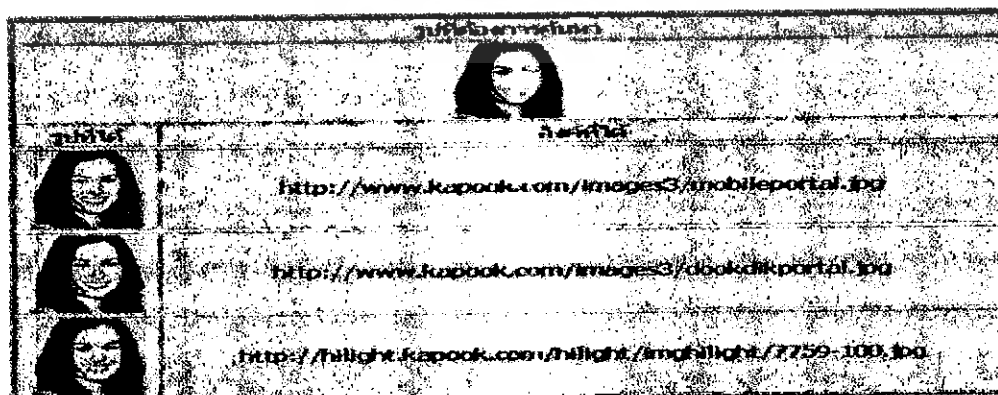
รูปที่ 4.9 ภาพชุดทดลองที่ 4



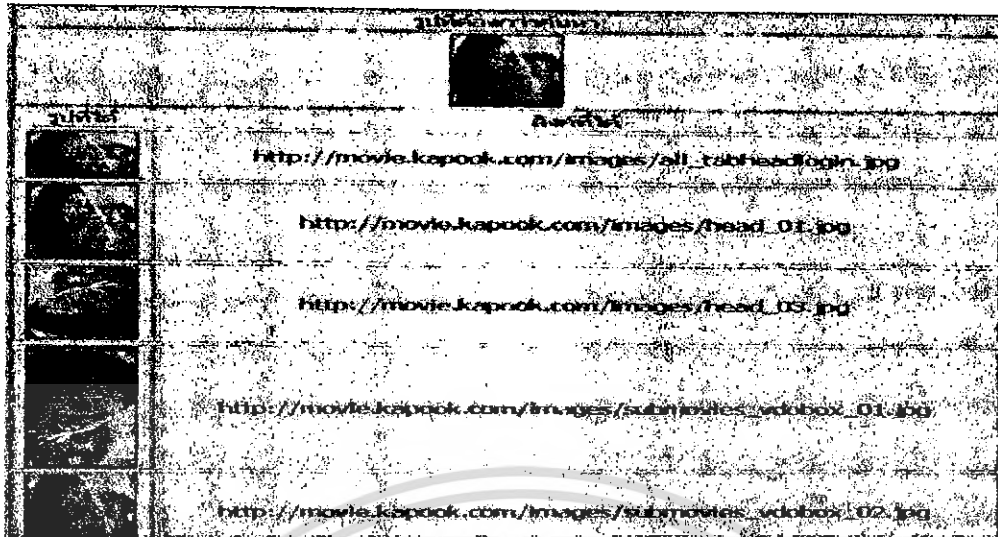
รูปที่ 4.10 ภาพชุดทดลองที่ 5

จากการทดลองใช้ค่า Differential = 1 จะ ได้รูปที่ตรงตามความต้องการเช่นเดียวกับการใช้ค่า Differential = 0 นั่นก็หมายความว่าค่าความแตกต่างของระดับสีของรูปที่ใช้ค้นหาความแตกต่างเพียงแต่เล็กน้อยจึงได้รูปที่มีความเหมือนกันกับรูปที่ทำการค้นหา

#### 4.4.3 การค้นหาข้อมูลโดยใช้ค่า Differential = 2



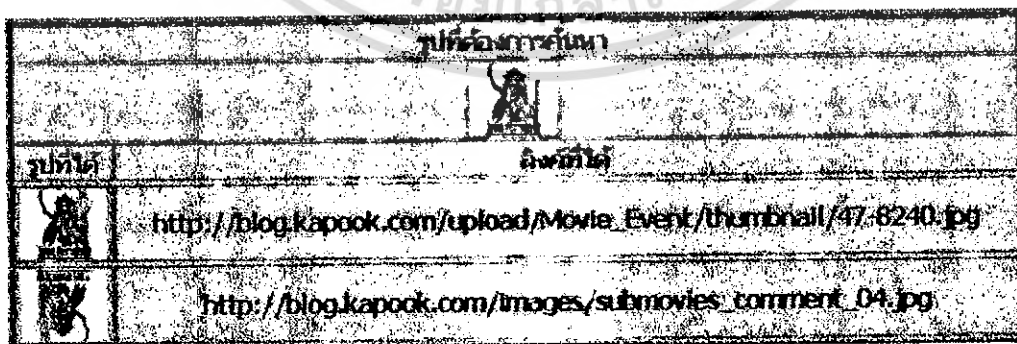
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 4.11 ภาพชุดทดลองที่ 1 ไม่นุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 ภาพชุดทดลองที่ 2

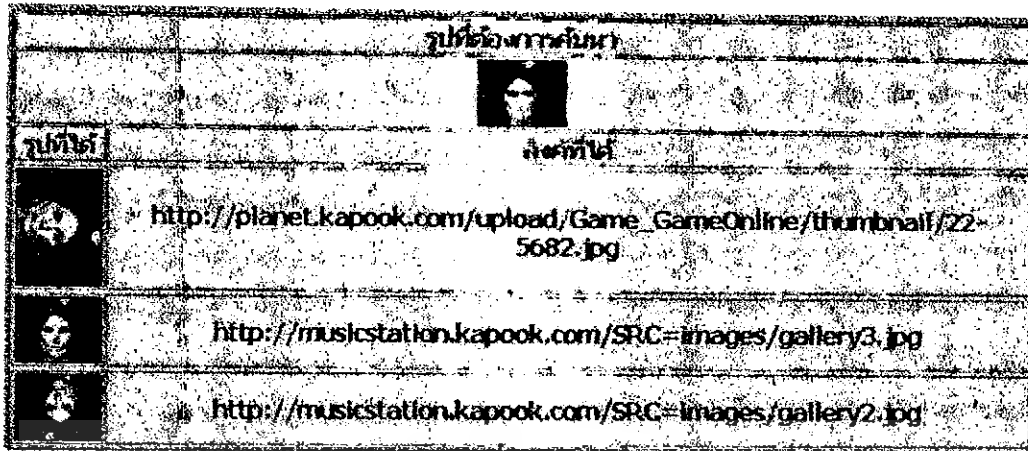


รูปที่ 4.13 ภาพชุดทดลองที่ 3



รูปที่ 4.14 ภาพชุดทดลองที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 ภาพชุดทดลองที่ 5

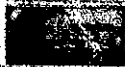




จากการทดลองใช้ค่า  $Differential = 2$  จะเห็นได้ว่าผลการทดลองมีความแตกต่างจากการใช้ค่า  $Differential = 0$  และ  $Differential = 1$  คือ รูปที่ได้นั้นจะมีจำนวนเพิ่มมากขึ้น เนื่องจากการใช้ค่า  $Differential$  ที่เพิ่มมากขึ้นจึงเป็นผลทำให้รูปที่มีลักษณะที่คล้ายกันแต่มีขนาดแตกต่างกันถูกแสดงออกมาด้วยการเลือกใช้ค่า  $Differential$  ที่เพิ่มขึ้นส่งผลให้การค้นหารูปมีความหลากหลายเพิ่มตามไปด้วย ส่วนรูปที่ค้นหาไม่ได้นั้นเกิดจากลักษณะของรูปที่ต่างกันทำให้ค่าสีสโตแกรมที่หามาได้อาจมีความแตกต่างกันมากเมื่อทดลองด้วยค่า  $Differential$  จึงได้แค่เพียงรูปที่มีขนาดที่ใกล้เคียงกัน

#### 4.4.4 การค้นหาข้อมูลโดยใช้ค่า $Differential = 3$

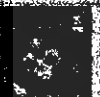

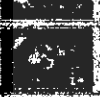
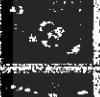



รูปที่ 4.16 ภาพชุดทดลองที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





รูปหน้าสารคดีในหา	
รูปหน้า	ลิงก์หน้า
	<a href="http://movie.kapook.com/images/all_tabheadlogin.jpg">http://movie.kapook.com/images/all_tabheadlogin.jpg</a>
	<a href="http://movie.kapook.com/images/head_01.jpg">http://movie.kapook.com/images/head_01.jpg</a>
	<a href="http://movie.kapook.com/images/head_03.jpg">http://movie.kapook.com/images/head_03.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_01.jpg">http://movie.kapook.com/images/submovies_vdobox_01.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_02.jpg">http://movie.kapook.com/images/submovies_vdobox_02.jpg</a>

รูปที่ 4.17 ภาพชุดทดลองที่ 2

รูปหน้าสารคดีในหา	
รูปหน้า	ลิงก์หน้า
	<a href="http://www.kapook.com/images/kapook_logo.jpg">http://www.kapook.com/images/kapook_logo.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_01.jpg">http://planet.kapook.com/images/newheadplanet220949_01.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_02.jpg">http://planet.kapook.com/images/newheadplanet220949_02.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_03.jpg">http://planet.kapook.com/images/newheadplanet220949_03.jpg</a>
	<a href="http://gamecenter.kapook.com/images/valentine_1.jpg">http://gamecenter.kapook.com/images/valentine_1.jpg</a>

รูปที่ 4.18 ภาพชุดทดลองที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปตัวอย่างค้นหา	
รูปที่ใด	ลิงค์ที่ใด
	<a href="http://blog.kapook.com/upload/Movie_Event/thumbnail/47-8240.jpg">http://blog.kapook.com/upload/Movie_Event/thumbnail/47-8240.jpg</a>
	<a href="http://blog.kapook.com/images/movies_comment.jpg">http://blog.kapook.com/images/movies_comment.jpg</a>
	<a href="http://blog.kapook.com/Images/submovies_comment_04.jpg">http://blog.kapook.com/Images/submovies_comment_04.jpg</a>
	<a href="http://blog.kapook.com/upload/Game_News/thumbnail/252-3252.jpg">http://blog.kapook.com/upload/Game_News/thumbnail/252-3252.jpg</a>

รูปที่ 4.19 ภาพชุดทดลองที่ 4







รูปตัวอย่างค้นหา	
รูปที่ใด	ลิงค์ที่ใด
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnail/22-5682.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnail/22-5682.jpg</a>
	<a href="http://musicstation.kapook.com/SRC=images/gallery3.jpg">http://musicstation.kapook.com/SRC=images/gallery3.jpg</a>
	<a href="http://musicstation.kapook.com/SRC=images/gallery2.jpg">http://musicstation.kapook.com/SRC=images/gallery2.jpg</a>

รูปที่ 4.20 ภาพชุดทดลองที่ 5








การทดลองด้วยค่า Differential = 3 ผลที่ได้จากการใช้ค่า Differential ในระดับนี้การทดลองที่ได้สามารถเปรียบเทียบรูปในฐานะข้อมูลได้ดียิ่งขึ้นเพราะรูปที่ได้เพิ่มมากขึ้นจากชุดทดลองที่ใช้ Differential = 0 , Differential =1 และ Differential =2 โดยรูปที่ค้นหาได้นั้นเป็นรูปที่เหมือนกับต้นฉบับแต่มีการกลับของรูปเท่านั้นเอง และมีรูปที่มีระดับสีที่มีความใกล้เคียงกันคิดมาด้วย ปัญหาในส่วนนี้คือ การหาลักษณะเฉพาะของรูปภาพโดยวิธีการหาค่าฮิสโตแกรมของภาพสีอย่างเดียวนั้นไม่สามารถที่จะค้นหารูปภาพได้อย่างมีประสิทธิภาพเท่าที่ควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4.5 การค้นหาข้อมูลโดยใช้ค่า Differential = 4




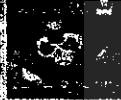
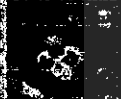
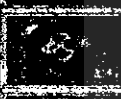
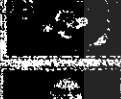

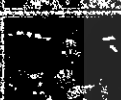

รูปตัวอย่างการค้นหา	
รูปที่ใด	ลิงค์ที่ใด
	<a href="http://www.kapook.com/images3/mobileportal.jpg">http://www.kapook.com/images3/mobileportal.jpg</a>
	<a href="http://www.kapook.com/images3/dookdikportal.jpg">http://www.kapook.com/images3/dookdikportal.jpg</a>
	<a href="http://www.postgang.com/prekapook/thumb/449.jpg">http://www.postgang.com/prekapook/thumb/449.jpg</a>
	<a href="http://www.kapook.com/msn/images/king_intro.jpg">http://www.kapook.com/msn/images/king_intro.jpg</a>
	<a href="http://highlight.kapook.com/ilight/ilight/7759-100.jpg">http://highlight.kapook.com/ilight/ilight/7759-100.jpg</a>
	<a href="http://movie.kapook.com/images/images/game_vdoleft_07.jpg">http://movie.kapook.com/images/images/game_vdoleft_07.jpg</a>

รูปที่ 4.21 ภาพชุดทดลองที่ 1

รูปตัวอย่างการค้นหา	
รูปที่ใด	ลิงค์ที่ใด
	<a href="http://movie.kapook.com/images/all_fabheadlogin.jpg">http://movie.kapook.com/images/all_fabheadlogin.jpg</a>
	<a href="http://movie.kapook.com/images/head_01.jpg">http://movie.kapook.com/images/head_01.jpg</a>
	<a href="http://movie.kapook.com/images/head_03.jpg">http://movie.kapook.com/images/head_03.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_01.jpg">http://movie.kapook.com/images/submovies_vdobox_01.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_02.jpg">http://movie.kapook.com/images/submovies_vdobox_02.jpg</a>
	<a href="http://www.kapook.com/promote/aismaewmong/images/mobile_javagame_01.jpg">http://www.kapook.com/promote/aismaewmong/images/mobile_javagame_01.jpg</a>
	<a href="http://gamecenter.kapook.com/images/valentine_9.jpg">http://gamecenter.kapook.com/images/valentine_9.jpg</a>

รูปที่ 4.22 ภาพชุดทดลองที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปตัวอย่างกราฟิก	
	
รูปตัวไอ	ลิงค์ที่ไอ
	<a href="http://planet.kapook.com/upload/Movie_News/thumbnail/240-3331.jpg">http://planet.kapook.com/upload/Movie_News/thumbnail/240-3331.jpg</a>
	<a href="http://www.egv.com/images/submovies_banner4_2.jpg">http://www.egv.com/images/submovies_banner4_2.jpg</a>
	<a href="http://www.kapook.com/images/kapook_logo.jpg">http://www.kapook.com/images/kapook_logo.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_01.jpg">http://planet.kapook.com/images/newheadplanet220949_01.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_02.jpg">http://planet.kapook.com/images/newheadplanet220949_02.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_03.jpg">http://planet.kapook.com/images/newheadplanet220949_03.jpg</a>
	<a href="http://highlight.kapook.com/hilight/Imghilight/7775-100.jpg">http://highlight.kapook.com/hilight/Imghilight/7775-100.jpg</a>
	<a href="http://highlight.kapook.com/hilight/Imghilight/7748-100.jpg">http://highlight.kapook.com/hilight/Imghilight/7748-100.jpg</a>
	<a href="http://gamecenter.kapook.com/images/valentine_1.jpg">http://gamecenter.kapook.com/images/valentine_1.jpg</a>





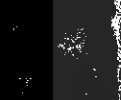
รูปที่ 4.23 ภาพชุดทดลองที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ของภาพค้นหา	
รูปถ่าย	ลิ้งค์ที่ไป
	<a href="http://movie.kapook.com/images/submovies_vdobox_05.jpg">http://movie.kapook.com/images/submovies_vdobox_05.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_06.jpg">http://movie.kapook.com/images/submovies_vdobox_06.jpg</a>
	<a href="http://blog.kapook.com/upload/Movie_Event/thumbnail/47-8240.jpg">http://blog.kapook.com/upload/Movie_Event/thumbnail/47-8240.jpg</a>
	<a href="http://blog.kapook.com/images/movies_comment.jpg">http://blog.kapook.com/images/movies_comment.jpg</a>
	<a href="http://blog.kapook.com/images/submovies_comment_02.jpg">http://blog.kapook.com/images/submovies_comment_02.jpg</a>
	<a href="http://blog.kapook.com/images/submovies_comment_03.jpg">http://blog.kapook.com/images/submovies_comment_03.jpg</a>
	<a href="http://blog.kapook.com/images/submovies_comment_04.jpg">http://blog.kapook.com/images/submovies_comment_04.jpg</a>
	<a href="http://movie.kapook.com/images/images/headmenugame.jpg">http://movie.kapook.com/images/images/headmenugame.jpg</a>
	<a href="http://movie.kapook.com/images/images/game_vdoleft_01.jpg">http://movie.kapook.com/images/images/game_vdoleft_01.jpg</a>
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnail/5-7545.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnail/5-7545.jpg</a>
	<a href="http://blog.kapook.com/upload/Game_News/thumbnail/253-5382.jpg">http://blog.kapook.com/upload/Game_News/thumbnail/253-5382.jpg</a>
	<a href="http://blog.kapook.com/upload/Game_News/thumbnail/252-3252.jpg">http://blog.kapook.com/upload/Game_News/thumbnail/252-3252.jpg</a>

รูปที่ 4.24 ภาพชุดทดลองที่ 4

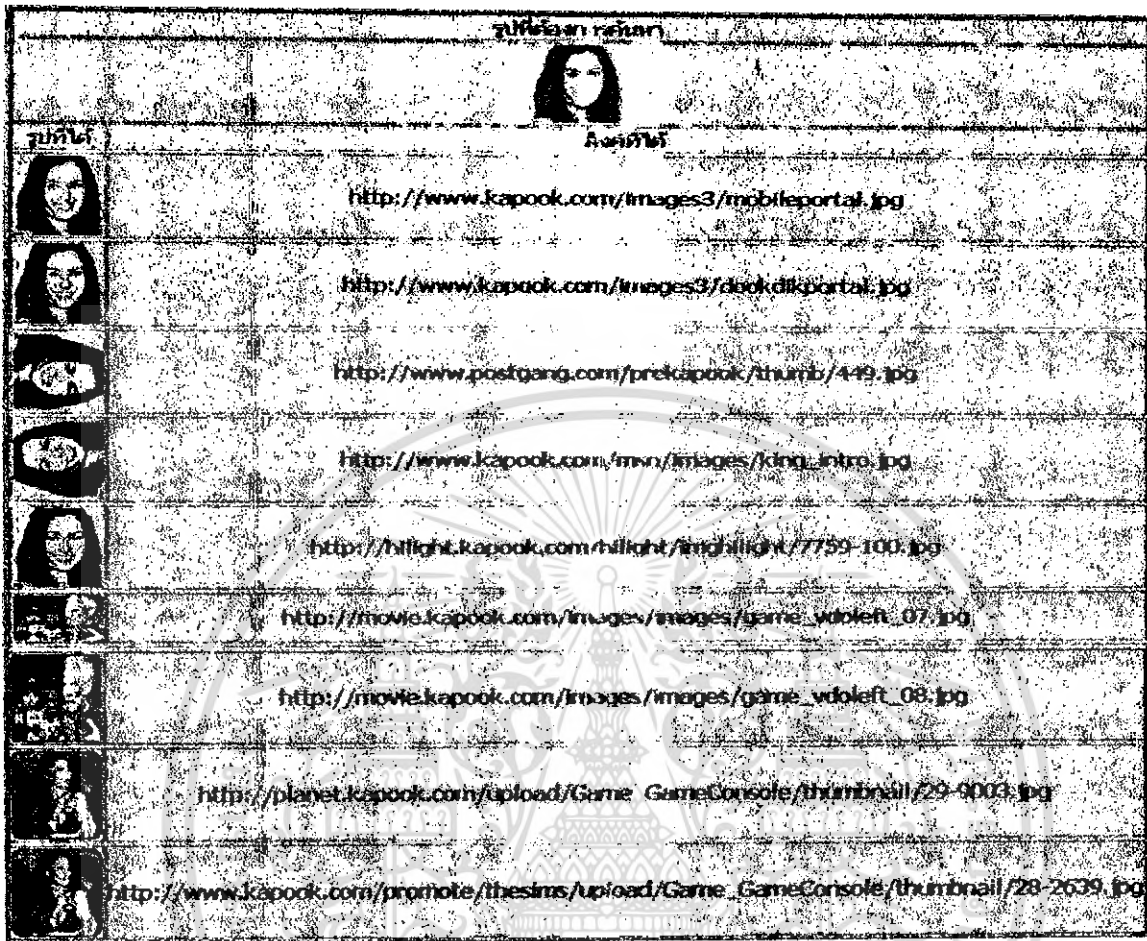
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปตัวอย่างการค้นหาก	
รูปที่ใด	ลิงค์ที่ใด
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/22-5682.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/22-5682.jpg</a>
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/21-2392.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/21-2392.jpg</a>
	<a href="http://musicstation.kapook.com/SRC=images/gallery3.jpg">http://musicstation.kapook.com/SRC=images/gallery3.jpg</a>
	<a href="http://musicstation.kapook.com/SRC=images/gallery2.jpg">http://musicstation.kapook.com/SRC=images/gallery2.jpg</a>
	<a href="http://musicstation.kapook.com/SRC=images/search_06.jpg">http://musicstation.kapook.com/SRC=images/search_06.jpg</a>

รูปที่ 4.25 ภาพชุดทดลองที่ 5

จากการทดลองใช้ค่า  $Differential = 4$  รูปที่ได้จากชุดทดลองสามารถหารรูปที่คล้ายกันได้เกือบทั้งหมด และมีรูปที่ไม่ต้องการคิดมาด้วย โดยรูปที่คิดมาด้วยนั้นจะเห็นได้ว่าลักษณะของรูปนั้นมีสีที่คล้ายกับรูปที่ค้นหาโดยค่าฮิสโตแกรมที่หาได้นั้นอาจมีความใกล้เคียงกันหรือมีค่าของระดับสีที่มีความใกล้เคียงกันจึงทำให้ได้รูปในลักษณะนี้คิดมาด้วย

#### 4.4.6 การค้นหาข้อมูลโดยใช้ค่า Differential = 5









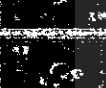



รูปที่ 4.26 ภาพชุดทดลองที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้







รูปสื่อภาพเคลื่อนไหว	
รูปถ่าย	รูปถ่ายทั่วไป
	<a href="http://movie.kapook.com/images/all_tabheadlogin.jpg">http://movie.kapook.com/images/all_tabheadlogin.jpg</a>
	<a href="http://movie.kapook.com/images/head_01.jpg">http://movie.kapook.com/images/head_01.jpg</a>
	<a href="http://movie.kapook.com/images/head_03.jpg">http://movie.kapook.com/images/head_03.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_01.jpg">http://movie.kapook.com/images/submovies_vdobox_01.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_02.jpg">http://movie.kapook.com/images/submovies_vdobox_02.jpg</a>
	<a href="http://www.kapook.com/promote/aismaewmong/images/mobile_jvgame_01.jpg">http://www.kapook.com/promote/aismaewmong/images/mobile_jvgame_01.jpg</a>
	<a href="http://gamecenter.kapook.com/images/valentine_9.jpg">http://gamecenter.kapook.com/images/valentine_9.jpg</a>
	<a href="http://gamecenter.kapook.com/random_9999/images/jvgame_01.jpg">http://gamecenter.kapook.com/random_9999/images/jvgame_01.jpg</a>

รูปที่ 4.27 ภาพชุดทดลองที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้












รูปตัวอย่างอื่นๆ	
	
รูปถ่าย	ลิงค์ที่ใด
	<a href="http://planet.kapook.com/upload/Movie_News/thumbnail/240-3331.jpg">http://planet.kapook.com/upload/Movie_News/thumbnail/240-3331.jpg</a>
	<a href="http://www.egv.com/images/submovies_banner4_2.jpg">http://www.egv.com/images/submovies_banner4_2.jpg</a>
	<a href="http://www.kapook.com/images/kapook_logo.jpg">http://www.kapook.com/images/kapook_logo.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_01.jpg">http://planet.kapook.com/images/newheadplanet220949_01.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_02.jpg">http://planet.kapook.com/images/newheadplanet220949_02.jpg</a>
	<a href="http://planet.kapook.com/images/newheadplanet220949_03.jpg">http://planet.kapook.com/images/newheadplanet220949_03.jpg</a>
	<a href="http://highlight.kapook.com/hilight/imghilight/7775-100.jpg">http://highlight.kapook.com/hilight/imghilight/7775-100.jpg</a>
	<a href="http://highlight.kapook.com/hilight/imghilight/7748-100.jpg">http://highlight.kapook.com/hilight/imghilight/7748-100.jpg</a>
	<a href="http://gamecenter.kapook.com/images/valentine_1.jpg">http://gamecenter.kapook.com/images/valentine_1.jpg</a>

รูปที่ 4.28 ภาพชุดทดลองที่ 3

รูปตัวอย่างอื่นๆ	
	
รูปถ่าย	ลิงค์ที่ใด
	<a href="http://movie.kapook.com/images/submovies_vdobox_05.jpg">http://movie.kapook.com/images/submovies_vdobox_05.jpg</a>
	<a href="http://movie.kapook.com/images/submovies_vdobox_06.jpg">http://movie.kapook.com/images/submovies_vdobox_06.jpg</a>
	<a href="http://blog.kapook.com/upload/Movie_Event/thumbnail/47-6240.jpg">http://blog.kapook.com/upload/Movie_Event/thumbnail/47-6240.jpg</a>
	<a href="http://blog.kapook.com/images/movies_comment.jpg">http://blog.kapook.com/images/movies_comment.jpg</a>
	<a href="http://blog.kapook.com/images/submovies_comment_02.jpg">http://blog.kapook.com/images/submovies_comment_02.jpg</a>





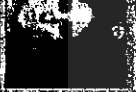
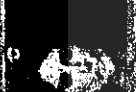



รูปที่ 4.29 ภาพชุดทดลองที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ **รูปที่ 4.29 ภาพชุดทดลองที่ 4** ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	<a href="http://blog.kapook.com/images/submovies_comment_03.jpg">http://blog.kapook.com/images/submovies_comment_03.jpg</a>
	<a href="http://blog.kapook.com/images/submovies_comment_04.jpg">http://blog.kapook.com/images/submovies_comment_04.jpg</a>
	<a href="http://blog.kapook.com/images/images/submovies_comment2_07.jpg">http://blog.kapook.com/images/images/submovies_comment2_07.jpg</a>
	<a href="http://blog.kapook.com/images/images/submovies_comment2_08.jpg">http://blog.kapook.com/images/images/submovies_comment2_08.jpg</a>
	<a href="http://movie.kapook.com/images/headmenugame.jpg">http://movie.kapook.com/images/headmenugame.jpg</a>
	<a href="http://blog.kapook.com/upload/Game_News/thumbnail/253-5382.jpg">http://blog.kapook.com/upload/Game_News/thumbnail/253-5382.jpg</a>
	<a href="http://event.kapook.com/images/7.jpg">http://event.kapook.com/images/7.jpg</a>
	<a href="http://event.kapook.com/images/8.jpg">http://event.kapook.com/images/8.jpg</a>
	<a href="http://event.kapook.com/images/indy35.jpg">http://event.kapook.com/images/indy35.jpg</a>
	<a href="http://www.kapook.com/images/k-fat-music.jpg">http://www.kapook.com/images/k-fat-music.jpg</a>
	<a href="http://www.kapook.com/promote/tang_jaice/images/head.jpg">http://www.kapook.com/promote/tang_jaice/images/head.jpg</a>
	<a href="http://www.kapook.com/imagepost/7741-100.jpg">http://www.kapook.com/imagepost/7741-100.jpg</a>

รูปที่ 4.29 (ต่อ) ภาพชุดทดลองที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.30 ภาพชุดทดลองที่ 5	
รูปที่	ลิงก์ไฟล์
	<a href="http://planet.kapook.com/upload/Movie_News/thumbnaill/235-8659.jpg">http://planet.kapook.com/upload/Movie_News/thumbnaill/235-8659.jpg</a>
	<a href="http://planet.kapook.com/upload/Movie_Event/thumbnaill/50-7384.jpg">http://planet.kapook.com/upload/Movie_Event/thumbnaill/50-7384.jpg</a>
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/22-5682.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/22-5682.jpg</a>
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/21-2392.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/21-2392.jpg</a>
	<a href="http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/20-1456.jpg">http://planet.kapook.com/upload/Game_GameOnline/thumbnaill/20-1456.jpg</a>
	<a href="http://musicstation.kapook.com/images/ssarch_08.jpg">http://musicstation.kapook.com/images/ssarch_08.jpg</a>
	<a href="http://musicstation.kapook.com/images/gallery2.jpg">http://musicstation.kapook.com/images/gallery2.jpg</a>
	<a href="http://musicstation.kapook.com/images/search_07.jpg">http://musicstation.kapook.com/images/search_07.jpg</a>
	<a href="http://musicstation.kapook.com/images/search_08.jpg">http://musicstation.kapook.com/images/search_08.jpg</a>

รูปที่ 4.30 ภาพชุดทดลองที่ 5

จากการทดลองใช้ค่า  $differential = 5$  จะเห็นได้ว่าชุดทดลองทั้งหมดสามารถหารรูปได้ครบทุกรูปครบทุกชุดในการเลือกค่า  $differential = 5$  นั้นจึงเป็นค่าที่เลือกใช้ในการทดสอบหารูปที่คล้ายและได้รูปที่มีลักษณะเหมือนกับรูปที่ต้องการค้นหาที่สุดแต่มีข้อเสียตรงที่ฮิสโตแกรมของสีไม่สามารถบอกได้ว่าระดับสีในแต่ละระดับอยู่ในตำแหน่งใดของภาพจึงทำให้ได้ภาพที่มีค่าสีที่ใกล้เคียงกันปะปนมากับรูปที่ทำการค้นหาด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4.7 การวิเคราะห์และสรุปผลการค้นหาด้วยค่า differential 0-5

จากการทดลองเปรียบเทียบค่า differential ในระดับต่างๆ ตั้งแต่ 0 – 5 เป็นผลทำให้เราสามารถทราบว่า การใช้ค่า differential ในระดับใดที่ตอบสนองความต้องการค้นหารูปได้มากที่สุด และค่าใดที่ทำให้ได้รูปที่ใกล้เคียง ค่า differential ที่ใช้แล้วมีผลการทดลองที่ให้ผลออกมาครบทุกชุด การเลือกใช้ค่า differential ที่ใช้ในการค้นหารูปเป็นดังนี้ การใช้ค่า differential ที่มีค่าเท่ากับ 5 โดยค่า differential = 5 นั้นเป็นค่าที่เราจะเลือกใช้ในการค้นหารูปที่ใกล้เคียงตามที่ต้องการเมื่อเลือกใช้ค่า differential = 5 ผลที่ได้คือ รูปภาพที่ได้อาจจะมิลักษณะที่เล็กกว่า ใหญ่กว่า มีจุดสีที่ใกล้เคียงกันหรือเอียงไปจากรูปต้นแบบติดมาด้วย การค้นหารูปที่คล้ายจะเลือกใช้ค่า differential = 3 ผลที่ได้คือ ได้ภาพที่มีลักษณะคล้ายกับรูปต้นแบบมาก ซึ่งลักษณะของภาพอาจจะมิลักษณะและจุดสีไม่ต่างมาก สุดท้ายการค้นหาที่ตรงตามความต้องการมากที่สุดจะเลือกใช้ค่า differential = 0 ค่า differential ในแต่ละระดับเมื่อเปรียบเทียบกันจะเห็นได้ว่าการเลือกใช้ค่า differential ที่มีค่าน้อย คือ differential = 0 และ differential = 1 จะเห็นได้จากการทดลองว่ารูปที่ค้นหานั้นจะตรงตามความต้องการมากที่สุดคือไม่มีรูปใดปะปนหรือรูปอื่นที่ไม่มีความเหมือนจากรูปที่ต้องการค้นหาติดมาด้วย และเมื่อเพิ่มค่า differential ขึ้นเป็น 2 , 3 , 4 , 5 ตามลำดับ ทำให้รูปที่ค้นหาจากชุดทดลองสามารถค้นหาจากชุดทดลองเพิ่มขึ้นดังตารางที่ 4.4 และมีรูปที่มีค่าของสีที่ใกล้เคียงกับรูปที่ค้นหาติดมาด้วย

ค่า differential ที่ใช้ มี 0-5 ระดับ	จำนวนภาพที่ได้จากการค้นของชุดภาพทดลอง				
	1	2	3	4	5
0	1	1	1	1	1
1	1	2	1	1	2
2	3	5	3	2	3
3	3	5	5	4	3
4	6	7	9	12	5
5	9	8	9	17	9

ตารางที่ 4.4 ตารางเปรียบเทียบการทำงานของ differential ในแต่ละระดับ

จากการทดลองปัญหาที่ยังคงมีในการเปรียบเทียบค่า differential นั้นคือทดลองแล้วได้รูปที่ไม่ต้องการติดเข้ามาด้วยปัญหานี้เกิดจากการหาลักษณะเฉพาะของรูปภาพ โดยวิธีการหาค่าฮิสโตแกรมของภาพสีอย่างเดียวนั้น ไม่สามารถที่จะค้นหาภาพได้อย่างมีประสิทธิภาพเท่าที่ควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.5 การทดลองหาภาพที่เหมือน คล้าย และใกล้เคียง

ในการทดลองนี้จะเป็นการค้นหารูปภาพที่เหมือน คล้าย และใกล้เคียง โดยใช้รูปที่สุ่มขึ้นมาเพื่อใช้ทดลองอย่างละ 3 ภาพ โดยในขณะที่ทำการทดลองมีภาพอยู่ในฐานข้อมูลประมาณ 17,000 ภาพ

##### 4.5.1 การทดลองหาภาพที่เหมือน

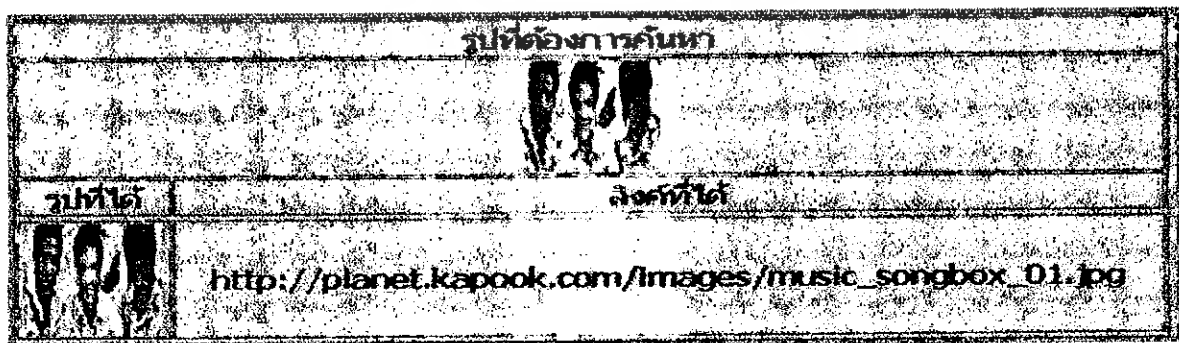
เป็นการทดสอบหารูปที่เหมือนกันทุกประการ โดยใช้ค่าความแตกต่างของรูปภาพ differential เท่ากับ 0 ซึ่งได้ผลลัพธ์ดังรูปที่ 4.31, 4.32 และ 4.33



รูปที่ 4.31 ภาพชุดทดลองหาภาพที่เหมือนกันชุดที่ 1



รูปที่ 4.32 ภาพชุดทดลองหาภาพที่เหมือนกันชุดที่ 2



รูปที่ 4.33 ภาพชุดทดลองหาภาพที่เหมือนกันชุดที่ 3


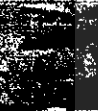
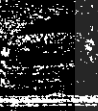


เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น เมื่อผู้ใดเห็นหน้าไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.5.2 การทดลองหาภาพที่คล้าย

เป็นการทดสอบหารูปที่คล้ายกัน โดยใช้ค่าความแตกต่างของรูปภาพ differential เท่ากับ 3 ซึ่งได้ผลลัพธ์ดังรูปที่ 4.34, 4.35 และ 4.36




รูปที่ต้องการค้นหา	
	
รูปที่ใด	ลิงก์ที่ใด
	<a href="http://movie.kapook.com/images/newpic08.jpg">http://movie.kapook.com/images/newpic08.jpg</a>
	<a href="http://movie.kapook.com/images/newpic09.jpg">http://movie.kapook.com/images/newpic09.jpg</a>

รูปที่ 4.34 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 1

รูปที่ต้องการค้นหา	
	
รูปที่ใด	ลิงก์ที่ใด
	<a href="http://www.kapook.com/promote/60thcelebrations/images2/51249/king_05.jpg">http://www.kapook.com/promote/60thcelebrations/images2/51249/king_05.jpg</a>
	<a href="http://www.kapook.com/promote/60thcelebrations/images2/51249/king_06.jpg">http://www.kapook.com/promote/60thcelebrations/images2/51249/king_06.jpg</a>
	<a href="http://www.kapook.com/promote/60thcelebrations/images2/51249/king_07.jpg">http://www.kapook.com/promote/60thcelebrations/images2/51249/king_07.jpg</a>
	<a href="http://www.kapook.com/promote/60thcelebrations/images2/51249/king_08.jpg">http://www.kapook.com/promote/60thcelebrations/images2/51249/king_08.jpg</a>

รูปที่ 4.35 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ต้องการค้นหา	
	
รูปที่ได้	ลิงค์ที่ได้
	<a href="http://planet.kapook.com/files/avatar/0493/Profile-Avatar-3907-0493.jpg">http://planet.kapook.com/files/avatar/0493/Profile-Avatar-3907-0493.jpg</a>
	<a href="http://planet.kapook.com/files/avatar/0097/Profile-Avatar-7968-0097.jpg">http://planet.kapook.com/files/avatar/0097/Profile-Avatar-7968-0097.jpg</a>

รูปที่ 4.36 ภาพชุดทดลองหาภาพที่คล้ายกันชุดที่ 3




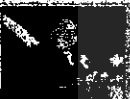

#### 4.5.3 การทดลองหาภาพที่ใกล้เคียง

เป็นการทดสอบหารูปที่ใกล้เคียงกัน โดยใช้ค่าความแตกต่างของรูปภาพ differential เท่ากับ 5 ซึ่งได้ผลลัพธ์ดังรูปที่ 4.37, 4.38 และ 4.39


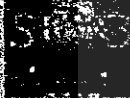



รูปที่ต้องการค้นหา	
	
รูปที่ได้	ลิงค์ที่ได้
	<a href="http://dookdik.kapook.com/images2/photo0001.jpg">http://dookdik.kapook.com/images2/photo0001.jpg</a>
	<a href="http://www.mobile-love.com/sr_01.jpg">http://www.mobile-love.com/sr_01.jpg</a>
	<a href="http://www.mobile-love.com/sr_02.jpg">http://www.mobile-love.com/sr_02.jpg</a>

รูปที่ 4.37 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่อ้างอิงกันหา	
รูปที่ใด	ลิงก์ที่ใด
	<a href="http://blog.kapook.com/upload/Game_Commentator/thumbnail/20-1582.jpg">http://blog.kapook.com/upload/Game_Commentator/thumbnail/20-1582.jpg</a>
	<a href="http://blog.kapook.com/images/pic3_ring02.jpg">http://blog.kapook.com/images/pic3_ring02.jpg</a>
	<a href="http://blog.kapook.com/images/pic3_ring03.jpg">http://blog.kapook.com/images/pic3_ring03.jpg</a>
	<a href="http://vdo.kapook.com/files/photo/3/274501_734822510_m.jpg">http://vdo.kapook.com/files/photo/3/274501_734822510_m.jpg</a>
	<a href="http://vdo.kapook.com/files/photo/0/270277_953142982_m.jpg">http://vdo.kapook.com/files/photo/0/270277_953142982_m.jpg</a>

รูปที่ 4.38 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 2

รูปที่อ้างอิงกันหา	
รูปที่ใด	ลิงก์ที่ใด
	<a href="http://highlight.kapook.com/hllight/lnghllight/7750-100.jpg">http://highlight.kapook.com/hllight/lnghllight/7750-100.jpg</a>
	<a href="http://www.majorcinplex.com/images/submovies_banner4_0.jpg">http://www.majorcinplex.com/images/submovies_banner4_0.jpg</a>
	<a href="http://www.majorcinplex.com/images/submovies_banner4_1.jpg">http://www.majorcinplex.com/images/submovies_banner4_1.jpg</a>
	<a href="http://movie.kapook.com/Images/Images/game_vdoleft_06.jpg">http://movie.kapook.com/Images/Images/game_vdoleft_06.jpg</a>
	<a href="http://blog.kapook.com/files/blogimages/g/gamers/Life-407x212.jpg">http://blog.kapook.com/files/blogimages/g/gamers/Life-407x212.jpg</a>

รูปที่ 4.39 ภาพชุดทดลองหาภาพที่ใกล้เคียงกันชุดที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยภาพที่ต่างกันไปจะได้ผลลัพธ์ที่ต่างกันไป ซึ่งจะขึ้นอยู่กับข้อมูลภาพที่มีอยู่ในฐานข้อมูลว่ามีมากเพียงใด และค่าสีของภาพนั้นมีภาพที่มีลักษณะสีคล้ายคลึงกันอยู่มากเพียงใด จึงทำให้การทดลองในการค้นหาภาพทั้งแบบเหมือน คล้าย และใกล้เคียง ซึ่งมีค่าของ Differential เท่ากับ 0, 3 และ 5 ตามลำดับ ของแต่ละชุดของการทดลองที่อาจจะไม่เท่ากันได้ ซึ่งการทดลองค้นหาภาพที่เหมือน ภาพที่ได้นั้นจะมีลักษณะที่เหมือนกันทุกประการแม้ว่าจะอยู่ในตำแหน่งของเว็บไซต์ที่ต่างกัน ส่วนการทดลองค้นหาภาพที่คล้ายกัน ภาพที่ได้จะมีลักษณะที่ใกล้เคียงกับภาพที่ต้องการค้นหาที่สุด ถึงแม้จะมีขนาดแล้วความเอียงของภาพที่ต่างกันไป และสุดท้ายการทดลองหาภาพที่ใกล้เคียงกัน ภาพที่ได้นั้นจะมีลักษณะที่คล้ายกันน้อยที่สุดในการทดลอง ซึ่งภาพที่ได้จากการค้นหาอาจมีภาพที่ไม่เหมือนกันภาพค้นฉบับเลขแต่อาจมีลักษณะของสีที่ใกล้เคียงกัน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### บทวิจารณ์และสรุป

#### 5.1 บทสรุป

จากการทดลองโปรแกรม Web Crawler ที่ได้พัฒนาขึ้นมา นั้น สามารถที่จะทำการค้นหา ลิงค์ของรูปภาพตามเว็บไซต์ต่างๆ โปรแกรม Web Crawler สามารถที่จะทำงานได้น่าพอใจโดย จากการทดลองการเก็บลิงค์ของ Web Crawler โปรแกรมที่ใช้ค้นหาลิงค์ตามเว็บไซต์จะค้นหาลิงค์ด้วยความเร็ว 176 ลิงค์/นาที โปรแกรมที่ใช้ค้นหาลิงค์ของรูปภาพมีความสามารถในการค้นหาลิงค์ของรูปภาพด้วยความเร็วประมาณ 46 ลิงค์/นาที ซึ่งปัญหาที่เกิดขึ้นในการทำงานของโปรแกรมนี้อาจเป็นลิงค์ที่ได้ค้นหานั้นบางลิงค์อาจจะค้นหาไม่ได้เนื่องจากลิงค์บางลิงค์เป็นลิงค์ที่เปิดไม่ได้จึงไม่สามารถค้นหาหรือเก็บลิงค์ดังกล่าวมาได้และความเร็วในการเก็บลิงค์นั้นอาจขึ้นอยู่กับความเร็วของ อินเทอร์เน็ตที่ใช้ด้วย ปัญหาอีกสิ่งหนึ่งคือ การค้นหาลิงค์ของเว็บ และลิงค์ของรูปภาพ จะมีความเร็วในการค้นหาไม่คงที่ขึ้นอยู่กับเว็บไซต์ที่ใช้เริ่มต้นในการค้นหา เพราะในแต่ละเว็บไซต์นั้นมีจำนวนลิงค์ของเว็บ และลิงค์ของรูปภาพเป็นจำนวนมากหรือน้อยแตกต่างกัน ปัญหาอีกสิ่งหนึ่งคือ Regular Expression (หรือ REs) ที่ใช้ในโปรแกรมยังไม่ดีเท่าที่ควรจึงทำให้บางครั้ง Web Crawler จะโหลดเอาลิงค์ที่ไม่ต้องการเข้ามาเก็บยังฐานข้อมูลด้วย

สรุปการทำงานในการค้นหารูปภาพ โดยใช้อิมเมจเสิร์จเอนจินผู้ใช้สามารถที่จะนำภาพที่ต้องการค้นหามาค้นหาภาพที่ต้องการได้ในฐานข้อมูล โดยมีวิธีการเปรียบเทียบ โดยการหาค่า ฮีชโตแกรมและเมื่อได้ภาพที่ต้องการจึงนำมาแสดงผลให้ผู้ใช้ การหารูปภาพในการทดลองผลลัพธ์ที่ได้จะต้องเป็นรูปที่มีความเหมือนกันทุกประการ หรือเป็นภาพเดียวกันแต่มีขนาดต่างกัน แต่ก็อาจจะเกิดกรณีที่ได้ภาพที่ไม่เหมือนกันเลยปะปนมาด้วย เช่น ภาพที่มีโทนสีโดยรวมใกล้เคียงกัน ดังนั้นในการออกแบบการทดลองต่อไปเราอาจจะค้นหารูปภาพที่มีความคล้ายคลึง โดยมีเป็นรูปที่มีความหมายใกล้เคียงกัน ซึ่งผลลัพธ์ที่ออกมาอาจมีความถูกต้องมากขึ้น แต่ผลลัพธ์ที่ดีที่สุดคือได้ผลลัพธ์ที่มีรูปภาพที่ไม่ตรงกับความต้องการเจอบนให้น้อยที่สุด แต่ในการทดลองยังพบอยู่

## 5.2 ปัญหาอุปสรรคและแนวทางในการแก้ไข

1. Regular Expression (หรือ REs) ที่ใช้ในโปรแกรมยังไม่ดีเท่าที่ควรต้องมีการพัฒนาเพิ่มเติมให้แมท สตริง ได้ตรงกับความต้องการมากที่สุด
2. ปรับปรุงอัลกอริทึมที่ใช้ในการเปรียบเทียบรูปที่ได้เพื่อเพิ่มประสิทธิภาพของโปรแกรมให้ดียิ่งขึ้น
3. สามารถหารูปภาพได้เพียงแค่จากลักษณะเฉพาะแบบฮิสโตแกรมเท่านั้นซึ่งอาจจะทำให้ได้ภาพที่มีความใกล้เคียงกันในลักษณะสีของภาพมาเท่านั้น

## 5.3 แนวทางการพัฒนาต่อ

1. พัฒนาอัลกอริทึมที่ใช้ในการค้นหารูปภาพให้มีประสิทธิภาพมากขึ้น เพื่อให้สามารถค้นหารูปภาพได้ตรงกับความต้องการของผู้ใช้มากที่สุดและพัฒนาในส่วนของ Regular Expression (หรือ REs) ให้มีประสิทธิภาพมากยิ่งขึ้น
2. พัฒนาส่วนของ โปรแกรมที่ใช้ในการเปรียบเทียบรูปเพื่อเพิ่มประสิทธิภาพให้แก่ตัวโปรแกรมให้ทำงานได้อย่างมีประสิทธิภาพ

## บรรณานุกรม

- [1] A.M. Kuchling , “Regular Expression HOWTO”, URL: <http://www.amk.ca/python/howto>
- [2] Guido van Rossum and Fred L. Drake, Jr., “Python Tutorial”
- [3] “MYSQL”, URL: <http://www.mysql.com>
- [4] Python How to Program
- [5] Python&XML
- [6] Python Programming
- [7] Python Imaging Library Overview
- [8] “Python programming language”, URL: <http://www.python.org>
- [9] Swaroop C H , 2004, “A Byte of Python”
- [10] การจัดการฐานข้อมูล
- [11] การเขียนเว็บด้วยภาษา PHP
- [12] สุทธิพล วรางกูร “การออกแบบและพัฒนาระบบเสิร์จเอนจินขนาดใหญ่” บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ เพื่อความสมบูรณ์แห่งปริญญาวิศวกรรมศาสตรมหาบัณฑิต (คอมพิวเตอร์) พ.ศ. 2544