

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน
IMAGE FINDER USING WEB SEARCH ENGINE

5/10



นายปนิพล อนอมเพชรสง่า
นายปณณวิช ธาราศักดิ์

เลขหมู่.....
เลขทะเบียน..... 63679
วัน,เดือน,ปี..... 30 ส.ค. 2549

b..... 11539611
i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน
IMAGE FINDER USING WEB SEARCH ENGINE



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญาานิพนธ์ปีการศึกษา 2548

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

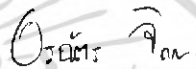
เรื่อง โปรแกรมค้นหารูปภาพ โดยใช้เว็บเสิร์จเอนจิน

Image Finder using Web Search Engine

ผู้จัดทำ

1. นายปนิพล ถนอมเพชรสง่า รหัสนักศึกษา 45010434

2. นายปณณวิช ธาราศักดิ์ รหัสนักศึกษา 45010456



Orattana Jitthaisakart

อาจารย์ที่ปรึกษา

(ผศ.ดร.อรณัตร์ จิตต์โสภาคย์)



อาจารย์ที่ปรึกษา

(อ.ธนัญชัย ตริภาค)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน

นายปนิพล ถนอมเพชรสง่า	45010434
นายปณณวิช ธาราศักดิ์	45010456
ผศ.ดร.อรฉัตร จิตต์โสภักตร์	อาจารย์ที่ปรึกษา
อ.ธนัญชัย ศรีภาค	อาจารย์ที่ปรึกษาร่วม
ปีการศึกษา 2548	

บทคัดย่อ

เนื่องจากในปัจจุบันในระบบเครือข่ายอินเทอร์เน็ตมีเว็บเพจอยู่เป็นจำนวนมาก และในแต่ละเว็บเพจนั้นก็มียูเอชทีเอ็มแอลเป็นจำนวนมากเช่นกัน เพราะฉะนั้นการค้นหารูปภาพในระบบเครือข่ายอินเทอร์เน็ตจึงมีความยากลำบากที่จะค้นหารูปภาพมาให้ได้ตามความต้องการ

ดังนั้น โปรแกรมค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจินจะช่วยทำการค้นหารูปภาพตามหน้าเว็บต่างๆแล้วนำมาจัดเก็บ โดยอัตโนมัติ เพื่อให้ผู้ใช้งานสามารถค้นหาภาพที่ต้องการ โดยจะได้ทั้งภาพที่เหมือนและคล้ายกัน รวมถึงที่อยู่ของรูปภาพนั้นด้วย โดยไม่ต้องเสียเวลาในการค้นหาเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Image Finder using Web Search Engine

Paniphon Tanompetsaga	45010434
Punnavit Tharasak	45010456
Asst.Prof.Dr. Orachat Chitsobhuk	Advisor
Mr. Thanunchai Threepak	Co-Advisor
Academic Year 2005	

ABSTRACT

With the advanced technology, there are a large number of websites in the internet. Each website also contains several images. If a user wants to search for an image from the internet, it would be very difficult to browse through such a large amount of information. Therefore, image search engine program developed in this project can help users to collect images from website in the internet while allow users to search the desired image from collected image database. Search engine gives not only similar images from the database but also the website of those images.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้อย่างดี ด้วยคำแนะนำ และคำปรึกษาจากอาจารย์ที่ปรึกษา
ผศ.ดร.อรฉัตร จิตต์โสภาคย์ และอาจารย์ ธานีชัย ตรีภาค ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์
ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านอาจารย์ทั้งสองท่านและขอขอบพระคุณเป็นอย่างสูง
ขอกราบพระคุณคณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบัน
เทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า
ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาวิศวกรรมคอมพิวเตอร์ สถาบันเทคโนโลยี
พระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมา
สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ
และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี
คุณค่าและประโยชน์อันพึงมาจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอบอบแต่ผู้มีพระคุณทุกท่าน

ปนิพล ถนอมเพชรสง่า

ปณณวิช ธาราศักดิ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มาของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 ขอบเขตของโครงการ.....	2
1.4 วิธีการดำเนินการ.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.6 ส่วนประกอบของปริญญานิพนธ์.....	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	4
2.1 กลไกการทำงานของ search engine.....	4
2.2 ประเภทของบริการค้นหาเว็บ.....	5
2.2.1 เสิร์จเอ็นจิน(search engine).....	5
2.2.2 ไคเรกทอรี(directory).....	6
2.2.3 ไฮบริดเสิร์จ เอ็นจิน(hybrid search engine).....	6
2.3 องค์ประกอบของเสิร์จเอ็นจิน.....	6
2.3.1 โรบอต (robots).....	6
2.3.2 อินเด็กเซอร์ (Indexer).....	6
2.3.3 ซอฟต์แวร์เสิร์จเอ็นจิน (Search Engine Software).....	6
2.4 Web robot (Web Crawler).....	7
2.4.1 Indexing Robot.....	8
2.4.2 การทำงานของโรบอต.....	8
2.5 อินเด็กเซอร์.....	9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.6 เสิร์จเอนจิน ซอฟต์แวร์.....	10
2.7 หลักการทำงานของภาษา Python.....	11
2.7.1 คอมไพเลอร์(Compiler).....	12
2.7.2 อินเตอร์พรีเตอร์(Interpreter).....	12
2.8 ข้อมูลภาพดิจิทัล (Digital Image Data).....	14
2.8.1 คุณสมบัติของกลุ่มสีที่แสดงผล.....	15
2.8.1.1 Uniformity.....	15
2.8.1.2 Completeness.....	15
2.8.1.3 Compactness.....	15
2.8.1.4 Naturalness.....	15
2.8.2 กลุ่มสีแสดงผลแบบ RGB.....	16
2.9 ฮิสโตแกรม (Histogram).....	18
2.9.1 การ Quantization.....	20
2.9.2 การ Normalization.....	20
2.9.3 ขั้นตอนการทำฮิสโตแกรม.....	21
บทที่ 3 กระบวนการค้นหารูปภาพโดยใช้เสิร์จเอนจิน.....	22
3.1 โครงสร้างฐานข้อมูล.....	22
3.1.1 table_image.....	22
3.1.2 table_link.....	23
3.1.3 table_id_process.....	24
3.2 การออกแบบและพัฒนาระบบค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน.....	25
3.2.1 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บ link.....	26
3.2.2 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บ link ในระดับถัดไป...28	
3.2.3 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บ link ของรูปภาพ.....	30
3.2.4 การคำนวณหาค่าฮิสโตแกรมของภาพ.....	32
3.3 การค้นหารูปภาพของผู้ใช้.....	33
3.4 ทดสอบหาค่าฮิสโตแกรมจากภาพต่างๆ.....	34
3.5 ส่วนหน้าของผู้ใช้.....	40

เอกสารนี้เป็นเอกสารที่สงวนเวลาหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

บทที่ 4 การทดลองและผลการทดลอง.....	42
4.1 ผลการทดลองเพื่อทดสอบประสิทธิภาพของ เซิร์จเอนจิน.....	42
4.2 ทดลองหาค่า Differential ที่เหมาะสม.....	44
4.3 การทดลองหาภาพตัวอย่าง	61
บทที่ 5 บทวิจารณ์และสรุป.....	65
5.1 บทสรุปและวิจารณ์ผลการทดลอง.....	65
5.2 ปัญหาอุปสรรคและแนวทางในการแก้ไข.....	66
5.3 แนวทางการพัฒนาต่อ.....	67
บรรณานุกรม.....	68
ภาคผนวก.....	69
ภาคผนวก ก. ตัวอย่างตาราง table image.....	70
ภาคผนวก ข. ตัวอย่างภาพในฐานข้อมูล.....	79

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตัวอย่างการกำหนดค่านัยสำคัญของแต่ละ TAG	11
2.2 การกำหนดระดับสี.....	17
2.3 ระดับของสีที่ใช้ในการทำ Quantization.....	20
3.1 ตัวอย่างตาราง table_image.....	22
3.2 ตัวอย่างตาราง table_link.....	23
3.3 ตัวอย่างตาราง table_id_process	24
3.4 ค่าฮิสโตแกรมของภาพชุดทดลองที่ 1	35
3.5 ค่าฮิสโตแกรมของภาพชุดทดลองที่ 2.....	36
3.6 ค่าฮิสโตแกรมของภาพชุดทดลองที่ 3.....	37
3.7 ค่าฮิสโตแกรมของภาพชุดทดลองที่ 4.....	38
3.8 ค่าฮิสโตแกรมของภาพชุดทดลองที่ 5.....	39
4.1 ผลการทดลองประสิทธิภาพการทำงานของ Web Crawler.....	43
4.2 ภาพที่หาได้จากชุดทดลองต่างๆ.....	60

สารบัญรูป

รูปที่	หน้า
2.1 การทำงานของส่วนประกอบแต่ละส่วนในระบบเสิร์จเอนจิน.....	7
2.2 การทำงานของโรบอต.....	9
2.3 รูปแบบการจัดเก็บข้อมูลของโปรแกรมสร้างคืบนี้.....	10
2.4 การทำงานของคอมพิวเตอร์.....	12
2.5 การทำงานของอินเทอร์เน็ต.....	12
2.6 ความสัมพันธ์ของสีในกลุ่มสีแสดงผลแบบ RGB	16
2.7 ภาพ RGB และค่าในพิกเซลของภาพ RGB	17
2.8 การผสมสีในรูปแบบ RGB	18
2.9 ภาพที่ทำฮีสโตแกรม.....	19
2.10 ภาพที่มีกราฟฮีสโตแกรมเหมือนกันแต่ภาพต่างกัน.....	19
2.11 ขั้นตอนการทำฮีสโตแกรม.....	21
3.1 ขั้นตอนการทำงานของอิมเมจเสิร์จเอนจิน.....	25
3.2 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์.....	27
3.3 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ในระดับถัดไป.....	29
3.4 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ของรูปภาพ.....	31
3.5 ขั้นตอนการทำงานของโปรแกรมที่ใช้ในการหาคำนวณค่าฮีสโตแกรมของภาพ.....	32
3.6 ขั้นตอนการทำงานผู้ใช้ในการค้นหารูปภาพ.....	33
3.7 หน้าแรกที่เป็นหน้าจอเพื่อให้ผู้ใช้งานทำการเลือกไฟล์รูปภาพที่จะทำการค้นหา.....	40
3.8 ผู้ใช้ทำการเลือกไฟล์รูปภาพ.....	40
3.9 ผลลัพธ์ของการค้นหา.....	41
3.10 เมื่อรูปภาพที่ใช้ค้นหาไม่พบ.....	41
3.11 ถ้าหากไฟล์รูปภาพที่เลือกมีขนาดเกิน 1Mb จะแจ้งว่าไฟล์มีขนาดใหญ่เกินไป.....	41
3.12 ถ้าหากกดปุ่ม Search โดยไม่ได้เลือกไฟล์จะแจ้งให้ทราบ.....	41
4.1 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 1.....	44
4.2 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 2.....	45
4.3 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 3.....	45
4.4 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 4.....	46
4.5 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1 ในภาพชุดทดลองที่ 5.....	46
4.6 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 1.....	47

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ขอสงวนสิทธิ์ในเนื้อหาและข้อมูลที่ได้เห็นไปใช้ประโยชน์ในการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.7 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 2.....	48
4.8 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 3.....	48
4.9 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 4.....	49
4.10 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 2 ในภาพชุดทดลองที่ 5.....	49
4.11 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 1.....	50
4.12 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 2.....	51
4.13 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 3.....	51
4.14 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 4.....	52
4.15 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 3 ในภาพชุดทดลองที่ 5.....	52
4.16 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 1.....	53
4.17 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 2.....	54
4.18 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 3.....	54
4.19 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 4.....	55
4.20 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 4 ในภาพชุดทดลองที่ 5.....	55
4.21 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 1.....	56
4.22 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 2.....	57
4.23 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 3.....	58
4.24 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 4.....	59
4.25 ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 5 ในภาพชุดทดลองที่ 5.....	60
4.26 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 1.....	61
4.27 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 2.....	61
4.28 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 3.....	62
4.29 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 4.....	62
4.30 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 5.....	63
4.31 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 6.....	63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของโครงการ

ในปัจจุบันข้อมูลต่างๆ ในเครือข่ายอินเทอร์เน็ตซึ่งโดยส่วนใหญ่แล้วจะเป็นข้อมูลที่อยู่ในเว็บไซต์ต่างๆ จากการพัฒนาของเทคโนโลยีในปัจจุบันทำให้การสร้างเว็บไซต์ต่างๆ สามารถทำได้ง่ายขึ้น ประกอบกับผู้คนมีความสนใจความรู้ทางด้านไอทีมากขึ้น จึงก่อให้เกิดเว็บไซต์ต่างๆ เกิดขึ้นใหม่อย่างมากมาย ข้อมูล ข่าวสารต่างๆ ในเครือข่ายอินเทอร์เน็ตจึงเพิ่มมากขึ้นเป็นเงาตามตัว จึงทำให้การค้นหาข้อมูลต่างๆ ในเครือข่ายอินเทอร์เน็ตมีความยากลำบากมากขึ้น

ข้อมูลรูปภาพเป็นข้อมูลที่นับได้ว่ามีความสำคัญในด้านต่างๆ มากมาย ทั้งที่ใช้รูปภาพเพื่อเก็บเหตุการณ์เรื่องราวต่างๆ ที่เกิดขึ้น หรือเป็นข้อมูลเกี่ยวกับความรู้ด้านต่างๆ นอกจากนั้นรูปภาพในปัจจุบันยังสามารถถูกสร้างขึ้นได้อย่างง่ายกว่าอดีต เช่น กล้องถ่ายภาพดิจิทัล (Digital Camera) การจับภาพจากกล้องถ่ายวิดีโอ (Video Camera) การสแกนรูปภาพด้วยสแกนเนอร์ (Scanner) โทรศัพท์เคลื่อนที่ (Mobile Phone) บางรุ่นที่มีกล้องถ่ายภาพติดอยู่ เป็นต้น จึงทำให้รูปภาพกลายเป็นข้อมูลที่นิยมใช้งานมากขึ้น ซึ่งในเครือข่ายอินเทอร์เน็ตนั้น มีข้อมูลรูปภาพอยู่เป็นจำนวนมาก ดังนั้นการที่จะได้มาซึ่งข้อมูลที่เราต้องการ จึงจำเป็นที่จะต้องหาเครื่องมือมาช่วยในการค้นหาข้อมูลให้ได้อย่างถูกต้องตรงกับความต้องการ และสามารถค้นหาว่ารูปภาพนั้นอยู่ที่ใดในเครือข่ายอินเทอร์เน็ต โดยใช้ภาพเป็นอินพุตในการค้นหา ซึ่งนั่นก็คือ อิมเมจเสิร์จเอนจินนั่นเอง

1.2 วัตถุประสงค์ของโครงการ

ปริญญาบัตรฉบับนี้มุ่งหวังเพื่อศึกษา และออกแบบอิมเมจเสิร์จเอนจิน โดยใช้ภาพจากเครื่องผู้ใช้งานเป็นตัวค้นหารูปภาพที่อยู่ในเครือข่ายอินเทอร์เน็ต โดยที่ส่วนของเครื่องเซิร์ฟเวอร์จะทำการรวบรวมรูปภาพจากเว็บไซต์ต่างๆ มาเก็บไว้ในฐานข้อมูลของตัวเครื่องเซิร์ฟเวอร์ และทำการหาลักษณะเฉพาะของรูปภาพแต่ละรูปด้วยกระบวนการหา ฮิสโตแกรม (Histogram) เมื่อได้ค่าดังกล่าวมาแล้วจะนำไปเก็บไว้ในฐานข้อมูลเพื่อประกอบกับรูปภาพที่รวบรวมมา เมื่อผู้ใช้งานต้องการค้นหา ก็จะนำภาพในเครื่องผู้ใช้งานเข้ากระบวนการหาลักษณะเฉพาะของภาพเช่นเดียวกัน เพื่อเป็นค่าที่ใช้ในการค้นหาภาพที่มีลักษณะเหมือนกันในฐานข้อมูลของเครื่องเซิร์ฟเวอร์ และแสดงผลออกทางหน้าเว็บเพจตามที่ผู้ใช้งานต้องการ

1.3 ขอบเขตของโครงการ

ในปฏิญญาพันธบัตรฉบับนี้ได้นำเสนอวิธีการศึกษา และออกแบบอิมเมจเสิร์จเอนจิน โดยแบ่งออกเป็น ส่วนของเครื่องมือที่ใช้ในการรวบรวมรูปภาพจากเว็บไซต์ต่างๆ (Web Crawler) แล้วทำการเก็บลิงค์ของรูปภาพต่างๆ ไว้ในฐานข้อมูล โดยที่เมื่อเข้าไปยังแต่ละเว็บไซต์จะทำการตรวจสอบว่าสามารถเข้าไปยังหน้าอื่นได้อีกหรือไม่ แล้วจึงทำการเก็บลิงค์ของรูปภาพต่างๆ ไว้ในฐานข้อมูล, ส่วนที่จัดการเกี่ยวกับฐานข้อมูล, ส่วนที่ใช้ในการประมวลผลภาพเพื่อหาค่าฮิสโตแกรมของภาพ เพื่อบอกลักษณะเฉพาะของภาพแต่ละภาพ และส่วนของเว็บเพื่อให้ผู้ใช้เข้าใช้งาน ค้นหารูปภาพจากเครื่องผู้ใช้

1.4 วิธีการดำเนินการ

1. ศึกษาการทำงานของเสิร์จเอนจิน ว่ามีการทำงานอย่างไรบ้าง
2. ศึกษาภาษาที่ใช้ในการสร้างส่วนต่างๆ ของเสิร์จเอนจิน
3. ศึกษาวิธีหาลักษณะเฉพาะของภาพรูปแบบต่างๆ
4. ศึกษาวิธีการจัดการฐานข้อมูล
5. ศึกษาภาษาที่ใช้ในการสร้างเว็บเพจ
6. จัดหาวัสดุอุปกรณ์ที่จำเป็นต้องใช้ในการพัฒนา
7. วิเคราะห์ และออกแบบระบบ
8. พัฒนาส่วนประกอบต่างๆ และนำมารวมเข้าด้วยกัน
9. ทดสอบระบบต่างๆ และแก้ไขจุดบกพร่อง

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้รับความรู้ความเข้าใจเกี่ยวกับเสิร์จเอนจิน
2. ได้รับความรู้ความเข้าใจเกี่ยวกับระบบฐานข้อมูล
3. ได้รับความรู้ความเข้าใจเกี่ยวกับการหาลักษณะเฉพาะของภาพ
4. ได้รับความรู้ความเข้าใจเกี่ยวกับการสร้างเว็บเพจ
5. ระบบอิมเมจเสิร์จเอนจินที่สามารถค้นหารูปภาพได้จริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 ส่วนประกอบของปฏิญญานิพนธ์

ปฏิญญานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 กล่าวถึงความสำคัญและที่มาของโครงการ วัตถุประสงค์ของโครงการ ขอบเขตของโครงการ วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบของปฏิญญานิพนธ์

บทที่ 2 กล่าวถึงทฤษฎีที่เกี่ยวข้องของระบบอิมเมจเสิร์จเอนจิน

บทที่ 3 กล่าวถึงกระบวนการค้นหาภาพโดยใช้เสิร์จเอนจิน

บทที่ 4 กล่าวถึงการทดลองและผลการทดลอง การดึงรูปภาพจากเว็บไซต์ต่างๆ และการค้นหารูปภาพของผู้ใช้

บทที่ 5 เป็นบทวิจารณ์และสรุป ซึ่งกล่าวถึงบทสรุปของโครงการ วิจารณ์สิ่งที่ได้รับจากโครงการ และข้อเสนอแนะสำหรับเป็นแนวทางในการพัฒนาต่อ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆที่เกี่ยวข้องในการวิจัย และพื้นฐานของระบบ อิมเมจเสิร์จเอนจิน ซึ่งเนื้อหาในบทนี้จะกล่าวถึงเรื่องของเสิร์จเอนจิน และเรื่องของข้อมูลภาพ คิวจัดอล ซึ่งเนื้อหาทั้งหมดนี้จำเป็นสำหรับการศึกษาระบบอิมเมจเสิร์จเอนจิน

2.1 กลไกการทำงานของเสิร์จเอนจิน

เสิร์จเอนจินทำงานโดยการเก็บข้อมูลเกี่ยวกับหน้าเว็บเพจ ซึ่งเสิร์จเอนจินสามารถดึงเอา ข้อมูลออกมาจากตัวเว็บ ไซต์ได้เอง โดย Web crawler จะทำหน้าที่ในการดึงข้อมูลของหน้านั้นมา เก็บไว้ ข้อมูลของเว็บแต่ละหน้าจะถูกวิเคราะห์เพื่อพิจารณาว่าจะเก็บเป็นอินเด็กซ์อย่างไร ข้อมูล เกี่ยวกับเว็บเพจจะถูกเก็บไว้ในฐานข้อมูลอินเด็กซ์เพื่อค้นหาในภายหลัง สำหรับเสิร์จเอนจินบาง ตัวอย่างเช่น Google จะเก็บทุกๆส่วนของ Source Page

เมื่อผู้ใช้งานทำการใช้งานเสิร์จเอนจิน และทำการถามหรือค้นหา ตัวอย่างเช่น การใส่ keyword เพื่อค้นหาซึ่งเสิร์จเอนจินจะทำการมองหาอินเด็กซ์ เพื่อจัดเตรียมรายชื่อของหน้าเว็บที่ ตรงกับความต้องการมากที่สุด โดยใช้กฎเกณฑ์ของเสิร์จเอนจิน แต่ก็มีเสิร์จเอนจินอีกแบบที่ไม่ได้ ใช้การเก็บอินเด็กซ์ อย่างเช่น Orase ข้อมูลที่เสิร์จเอนจินต้องการนั้นใช้เมื่อเกิดการถามขึ้นเท่านั้น ซึ่งจะมีข้อดีตรงที่ข้อมูลจะมีการอัปเดตมากกว่า ทำให้ไม่มี dead link แต่มีข้อเสียคือใช้เวลาในการ ค้นหาานกว่า

โดยสรุปแล้วอาจแบ่งกระบวนการทำงานของเสิร์จเอนจินออกเป็น 3 ส่วน คือ

- 1) โปรแกรมรวบรวมเอกสารเว็บ (Spider หรือ Crawler) จะทำหน้าที่ออกไปสำรวจเว็บ จากโดเมนต่าง ๆ โดย Spider จะเข้าไปสำรวจและอ่านเว็บเพจ และหากพบลิงค์ก็จะทำ การติดตามลิงค์ภายในเว็บไซต์จนครบ ซึ่งจากการทำงานในลักษณะโง่งงนี้ จึงเป็นที่มาของคำว่า Spider หรือ Crawler จากนั้น Spider จะนำข้อมูลเว็บดังกล่าวไปเก็บไว้ ในฐานข้อมูลของเสิร์จเอนจิน และ Spider จะกลับไปตรวจสอบข้อมูลในเว็บนั้นๆอย่าง สม่าเสมอ เช่น ทุก 1 หรือ 2 เดือนเพื่อสำรวจความเปลี่ยนแปลง
- 2) รายการครรชนีข้อมูลที่ Spider พบจะถูกส่งต่อมายังเก็บที่รายการครรชนี (อินเด็กซ์ หรือ catalog) ซึ่งเปรียบเสมือนเป็นสมุดเล่มใหญ่ที่จัดเก็บสำเนาของเว็บเพจที่ Spider พบ หากข้อมูลที่เว็บต้นฉบับมีการเปลี่ยนแปลงข้อมูลในสมุดครรชนีจะเปลี่ยนแปลง ด้วย เอกสารเว็บจะถูกทำครรชนีและจัดเก็บตามบัญชีครรชนีที่กำหนดไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) โปรแกรมการสืบค้น (เสิร์จเอนจิน ซอฟต์แวร์) จะทำหน้าที่ในการค้นหาข้อมูลจากฐานข้อมูลของเสิร์จเอนจิน จะเริ่มต้นการทำงานเมื่อผู้ใช้ป้อนคำค้นหาโปรแกรมจะทำหน้าที่นำคำค้นหาของผู้ใช้ไปจับคู่กับบรรณานุกรมในฐานข้อมูลแล้วทำการดึงข้อมูล (เอกสารเว็บ) ที่ตรงกับคำค้นหาออกมา และจัดลำดับผลการค้นหาตามระดับความเกี่ยวข้องที่โปรแกรมประเมินได้ซึ่งเสิร์จเอนจินแต่ละตัวจะใช้ตรรกะที่แตกต่างกันไป ตัวอย่างการสร้างเกณฑ์ในการกำหนดค่าน้ำหนักให้กับคำบรรณานุกรมของเสิร์จเอนจินอาจจัดลำดับ ดังนี้ (ลำดับที่ 1 จะให้ค่าน้ำหนักความเกี่ยวข้องกับคำค้นหามากที่สุด)

- ลำดับที่ 1 คำหรือวลี ที่ปรากฏใน Meta tag elements
- ลำดับที่ 2 คำหรือวลี ที่ปรากฏใน Title tag
- ลำดับที่ 3 คำหรือวลี ที่ปรากฏใน Main heading และ Sub heading
- ลำดับที่ 4 จำนวนครั้งที่คำ หรือวลี ปรากฏในส่วนเนื้อหาของเอกสาร
- ลำดับที่ 5 ความถี่ที่เว็บอื่น ๆ เชื่อมโยงเข้ามา

2.2 ประเภทของบริการค้นหาเว็บ

บริการค้นหาเว็บเพ็จที่มีอยู่ในปัจจุบันเช่น Google, AltaVista, Lycos, Yahoo, HotBot มักเรียกโดยรวมว่าเป็นเสิร์จเอนจินนั้นแท้ที่จริงแล้วสามารถแบ่งประเภทของเครื่องมือที่ช่วยในการค้นหาเว็บเพ็จได้เป็น 3 ประเภท ดังนี้

2.2.1 เสิร์จเอนจิน

เป็นระบบซอฟต์แวร์ที่มีโปรแกรม ไรบอท (Robot) ซึ่งบางครั้งเรียกว่า สไปเดอร์ (Spider) หรือ ครอว์เลอร์ (Crawler) ทำหน้าที่เดินทางไปยังเว็บไซต์ต่างๆ ในอินเทอร์เน็ตและอ่านเว็บเพ็จจากไซต์เหล่านั้นเพื่อนำมาสร้างดัชนีรายการของเว็บเพ็จโดยอัตโนมัติ การทำดัชนีรายการเว็บเพ็จด้วยเสิร์จเอนจินจึงสามารถสร้างดัชนีของเว็บเพ็จได้เป็นจำนวนที่มากกว่าของไครคทอรี เนื่องจากดัชนีของไครคทอรีจะถูกจัดการโดยมนุษย์ แทนที่จะใช้คอมพิวเตอร์ ถ้าหากเว็บเพ็จที่ถูกทำดัชนีแล้วเกิดการเปลี่ยนแปลงเสิร์จเอนจินจะทราบถึงการเปลี่ยนแปลงและจัดนำเว็บเพ็จที่มีการเปลี่ยนแปลงนั้นมาสร้างดัชนีใหม่ ซึ่งการสร้างดัชนีใหม่นี้อาจส่งผลต่อการจัดลำดับของเว็บเพ็จนั้น ตัวอย่างของเสิร์จเอนจินที่มีอยู่ในปัจจุบันเช่น HotBot, AltaVista เป็นต้น

2.2.2 ไคเรททอรี (Directory)

แตกต่างจากเสิร์จเอนจินตรงที่ไคเรททอรีถูกจัดการโดยมนุษย์ เว็บไซต์ใดที่ต้องการมีรายชื่อในไคเรททอรีต้องติดต่อผู้ดูแลไคเรททอรีเพื่อให้ผู้ดูแลไคเรททอรีจำแนกประเภทของเว็บเพจนั้นๆ ให้อยู่ในประเภทที่เหมาะสม เนื่องจากไคเรททอรีถูกจัดเป็นหมวดหมู่ การค้นหาเว็บเพจด้วยไคเรททอรีจึงให้ผลลัพธ์ที่ตรงกับความต้องการมากกว่าเสิร์จเอนจิน ถ้าหากมีการเปลี่ยนแปลงเกิดขึ้นกับเว็บเพจที่ได้รับการจัดลำดับแล้ว การเปลี่ยนแปลงนั้นจะไม่ส่งผลกระทบต่อลำดับเดิมที่ได้รับการจัดไว้ ตัวอย่างไคเรททอรีที่มีอยู่ในปัจจุบันเช่น Yahoo

2.2.3 ไฮบริดเสิร์จเอนจิน (Hybrid Search Engine)

เป็นเสิร์จเอนจินที่มีไคเรททอรีผสมอยู่ด้วย

2.3 องค์ประกอบของเสิร์ชเอนจิน

เสิร์ชเอนจินประกอบด้วย 3 ส่วนหลัก ได้แก่ โรบอท อินเด็กเซอร์และเสิร์ชเอนจินซอฟต์แวร์

2.3.1 โรบอท (robots) หรือ สไปเดอร์ หรือ ครอว์เลอร์

ทำหน้าที่เดินทางไปยังเว็บไซต์ต่างๆ เพื่อสะสมไฟล์ HTML เพื่อมาเป็นข้อมูลสำหรับสร้างดัชนีค้นหา โดยทั่วไปแล้วโรบอทจะรวบรวมเอกสารเฉพาะเว็บไซต์ที่ดูกำหนดอยู่ก่อนแล้ว หลังจากทีโรบอทได้อ่านเว็บเพจเรียบร้อยแล้วโรบอทจะกลับมามาอ่านเว็บไซต์ที่อ่านไปแล้วเพื่อตรวจสอบการเปลี่ยนแปลงตามระยะเวลาที่กำหนด ซึ่งโดยทั่วไปแล้วจะประมาณ 2 เดือน

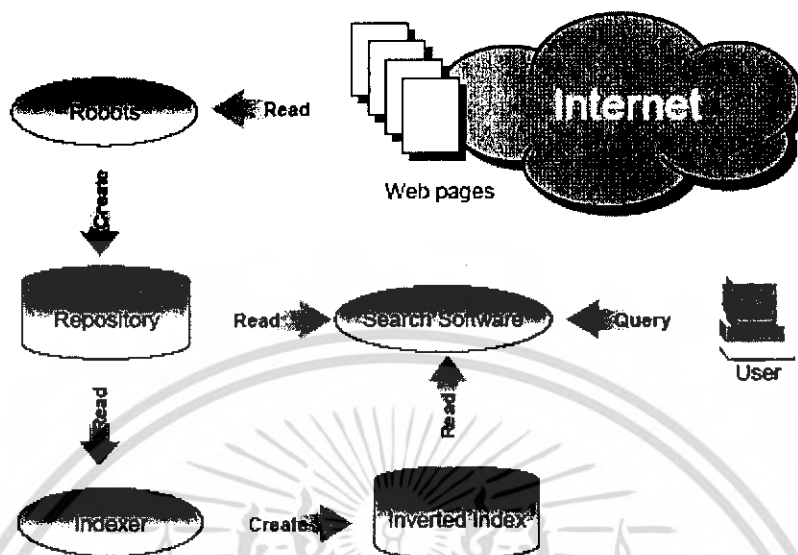
2.3.2 อินเด็กเซอร์ (Indexer, Catalog)

เป็นส่วนที่ทำหน้าที่สร้างดัชนีค้นหาจากไฟล์ HTML ที่โรบอทหาวิธีการสร้างดัชนีค้นหานั้นมีอยู่ด้วยกันหลายวิธี เช่น Inverted index, Vectorspace, N-Gram ซึ่งแต่ละวิธีก็มีข้อดีข้อเสียต่างกัน แต่วิธีที่เป็นที่นิยมมากสำหรับเสิร์ชเอนจินก็คือ Inverted index เว็บเพจใดๆ จะสามารถสืบค้นได้จากเสิร์ชเอนจินก็ต่อเมื่อเว็บเพจนั้นผ่านการทำดัชนีมาแล้วเท่านั้น

2.3.3 ซอฟต์แวร์เสิร์ชเอนจิน (Search engine software)

เป็นโปรแกรมทำหน้าที่รับคำศัพท์ที่ผู้ใช้ต้องการค้นหาผ่านทางซีจีไอ (CGI, Common Gateway Interface) มาค้นหาในดัชนีค้นหาที่ถูกสร้างโดยอินเด็กเซอร์เพื่อเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค้นหาเว็บเพจที่ตรงกับความต้องการของผู้ใช้เสร็จสิ้นจินบางระบบจะจัดอันดับความเกี่ยวข้องของเว็บเพจด้วย



รูปที่ 2.1 การทำงานของส่วนประกอบแต่ละส่วนในระบบเสิร์จเอนจิน

2.4 Web robot (Web Crawler)

คือโปรแกรมซึ่งจะทำการดึงทุกๆหน้าใน World Wide Web หรือระบบเครือข่ายอินเทอร์เน็ตโดยการกระจายสู่ Hyperlink ที่อ้างถึงจากหน้าแรกแล้วทำการดึงข้อมูลมาโดยที่การทำงานของเว็บโรบอท จะต้องขึ้นกับ robots.txt เป็นโปรโตคอล ซึ่งแต่ละเว็บไซต์จะมีเป็นของตัวเองเพื่อที่จะบอกว่าสามารถเข้าไปยังไฟล์ใดได้บ้างแต่โรบอทก็มีข้อเสีย อาทิเช่น

- โรบอททำให้เครือข่ายทำงานหนักมากขึ้น ซึ่งส่วนมากจะเกิดกับคนที่เริ่มเขียนโรบอท แต่ปัจจุบันมีข้อมูลเพียงพอที่จะป้องกันความผิดพลาดนี้แล้ว
- โรบอทถูกสร้างขึ้น โดยมนุษย์ ซึ่งถูกกระทำอย่างผิดพลาดในการตั้งองค์ประกอบ หรือไม่ได้พิจารณาการกระทำที่แฝงมาภายหลัง

เว็บโรบอทบางครั้งอาจจะถูกกล่าวถึงในชื่อ Web Wanderers, Web Crawlers หรือ Spiders

Web robot มีด้วยกันหลายชนิดดังนี้

- indexing
- HTML validation
- Link validation
- “ What 's new ” monitoring
- Mirroring

ซึ่งในที่นี้จะอธิบายเพียงอินเด็กซึ่ง ีโอบอทซึ่งจะใช้ในเสิร์จเอนจิน

2.4.1 Indexing Robot

โดยทั่วไป ีโอบอทจะเริ่มทำงานจากรายชื่อของ URL ที่เก็บไว้เดิม โดยเฉพาะไฟล์ที่มีการเชื่อมโยงไปที่อื่นมากๆ อย่างเช่นเซิร์ฟเวอร์ใหม่ๆ และเว็บไซต์ที่ได้รับความนิยม บริการของ Indexing ส่วนมาก จะอนุญาตให้เลือก URL เองซึ่งจะเข้าไปในคิวและจะถูกเข้าถึงโดย ีโอบอท

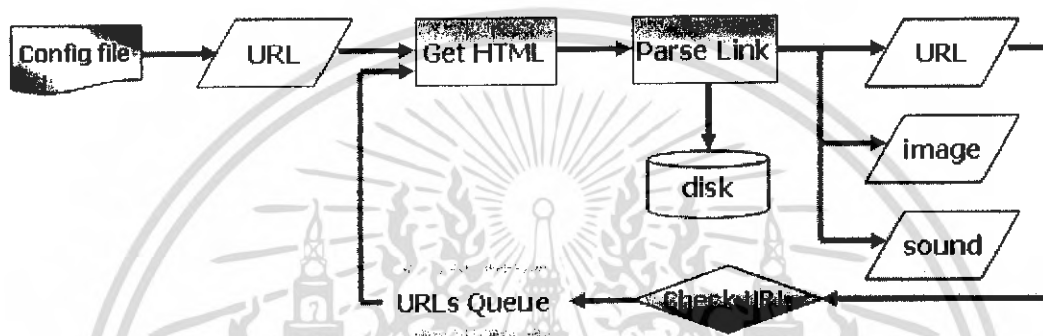
Indexing robot จะตัดสินใจว่าจะ ีโอบอทอินเด็กซ์โดยที่ถ้ารู้ข้อมูลเกี่ยวกับไฟล์นั้น Indexing robot อาจตัดสินใจที่จะตัดทิ้ง และทำการแทรกตัวเองเข้าไปในฐานข้อมูล ซึ่งการจะทำอะไรก็ขึ้นอยู่กับ ีโอบอทบาง ีโอบอท จะเก็บ title เป็นอินเด็กซ์ หรือ 2-3 ย่อหน้าแรก หรือว่าอาจจะเก็บทุกๆคำ ซึ่งนำหน้าของสิ่งที่จะเก็บก็ขึ้นอยู่กับโครงสร้างของ HTML บางส่วนจะเก็บ meta tag หรือ tag พิเศษอื่นๆ ที่ถูกซ่อนอยู่ ซึ่ง ีโอบอทจะสามารถเข้าไปโหลดได้ก็ต่อเมื่อเข้าไปตรวจสอบที่ไฟล์ robot.txt ซึ่งเป็นไฟล์ที่จะบอกว่าสามารถเข้าถึงส่วนใดได้บ้างของไซต์นั้น ซึ่งแต่ละเว็บไซต์ก็จะสามารถกำหนดโปรโตคอล robot.txt ได้เฉพาะตัว

2.4.2 การทำงานของ ีโอบอท

ีโอบอทเป็นโปรแกรมที่ท่องไปในอินเทอร์เน็ตเพื่อจุดประสงค์ใดจุดประสงค์หนึ่งเช่น เพื่อตรวจสอบว่ามีเว็บเพจใดบ้างที่ไม่สามารถอ่านได้ เป็นต้น แต่วัตถุประสงค์หลักของ ีโอบอทที่เกี่ยวกับเสิร์จเอนจินก็คือ ีโอบอทจะทำหน้าที่รวบรวมข้อมูลจากเว็บต่างๆ จากอินเทอร์เน็ต แล้วบันทึกไฟล์เหล่านั้นลงในฮาร์ดดิสก์เพื่อนำมาใช้ในส่วนของอินเด็กเซอร์ต่อไป

ในปัจจุบันได้มีมาตรฐาน Standard Robot Exclusion ซึ่งเป็นมาตรฐานที่ใช้กำหนดการทำงานของ ีโอบอท สำหรับเว็บไซต์แต่ละแห่ง โดยที่เว็บไซต์ใดๆสามารถกำหนดได้ว่า จะยอมให้ ีโอบอทตัวใดมาอ่านหรือไม่อนุญาตให้อ่านข้อมูลส่วนหนึ่งส่วนใดในเว็บไซต์ได้โดยการสร้างไฟล์ Robot.txt ขึ้นมา

การทำงานของโรบอทโดยทั่วไปจะมีลักษณะใกล้เคียงกันคือ เริ่มอ่านเว็บเพจ เริ่มต้นแล้วตรวจสอบดูว่าภายในเว็บเพจที่อ่านมาได้นั้นมีไฮเปอร์ลิงค์อะไรบ้าง เคยพบมาก่อนหรือไม่ ถ้ายังไม่เคยพบมาก่อน ก็จะมาวิเคราะห์ดูว่าสมควรที่จะอ่านเว็บเพจนั้นหรือไม่ โดยจะตรวจสอบโดยอาศัยกฎที่มีการตั้งไว้ เช่น จะอ่านเว็บไซต์ใดและไม่อ่านเว็บไซต์ใดบ้าง เมื่อเว็บเพจนั้นผ่านกฎที่มีการตั้งไว้แล้ว ตัวโรบอทก็จะตามไปอ่านเว็บเพจนั้นมา ทำซ้ำเช่นนี้จนกว่าลิงค์ที่ได้มาจะอ่านได้ครบ ดังรูปที่ 2.2



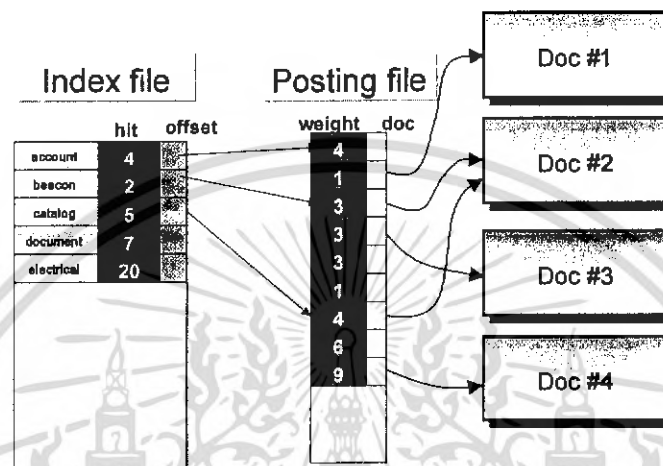
รูปที่ 2.2 การทำงานของโรบอท

2.5 อินเด็กเซอร์

การทำงานของโปรแกรมอินเด็กเซอร์นั้นมีวิธีการสร้างดัชนีค้นหาอยู่หลายวิธี เช่น แบบอินเวิร์ดอินเด็กซ์ (Inverted index), เวกเตอร์สเปซโมเดล (vector space model) เป็นต้น แต่ในที่นี้จะใช้เทคนิคในการสร้างดัชนีค้นหาแบบอินเวิร์ดอินเด็กซ์ ตัวอย่างเช่น สมมุติว่ามีหนังสือหนาๆอยู่เล่มหนึ่งแล้วต้องการค้นหาคำว่า “network” ในหนังสือเล่มนั้น คงไม่มีใครที่จะเปิดหนังสือทีละหน้า อ่านตัวหนังสือทีละบรรทัด เพื่อที่จะดูว่ามีคำว่า “network” อยู่ที่หน้าไหน สิ่งที่จะทำก็คือเปิดหนังสือไปที่ท้ายเล่มเพื่อดูดัชนีคำศัพท์ของหนังสือเล่มนั้น ว่ามีคำว่า “network” อยู่ที่หน้าไหน รูปแบบของดัชนีค้นหาของอินเวิร์ดอินเด็กซ์ก็คล้ายๆกันกับดัชนีท้ายเล่มของหนังสือ อินเวิร์ดอินเด็กซ์ก็คือตารางของคำศัพท์ที่ถูกสร้างขึ้นมาจากสำหรับค้นหา โดยที่ภายในของอินเวอ์อินเด็กซ์นั้นจะจัดเก็บคำศัพท์และตำแหน่งที่อยู่ของคำศัพท์เหล่านั้นว่าอยู่ที่ใด เพราะฉะนั้นเมื่อมีการค้นหาเกิดขึ้นก็ไม่จำเป็นต้องไปค้นหาคำศัพท์จากเอกสารทีละเอกสาร แต่สามารถดูที่อินเวอ์อินเด็กซ์ได้เลยว่าคำศัพท์คำที่ค้นหาอยู่ที่ไหนบ้าง

โปรแกรมอินเด็กเซอร์นั้นจะมีหน้าที่สร้างตารางอินเวิร์คอินเด็กซ์ ข้อมูลในอินเวิร์คอินเด็กซ์ นั้นนอกจากจะเก็บตำแหน่งที่อยู่ของคำศัพท์แล้ว อินเด็กเซอร์ยังมีหน้าที่เก็บข้อมูลลักษณะรูปแบบของคำศัพท์แต่ละคำว่ามีปรากฏอยู่ใน HTML Tag อะไรเพื่อเป็นข้อมูลให้กับเสิร์จเอนจิน ซอฟต์แวร์ นำเอาไปใช้ในการจัดเรียงลำดับคอนค้นหาต่อไป

Inverted Index



รูปที่ 2.3 รูปแบบการจัดเก็บข้อมูลของโปรแกรมสร้างดัชนี

2.6 ซอฟต์แวร์เสิร์จเอนจิน

ซอฟต์แวร์เสิร์จเอนจินมีหน้าที่รับคำศัพท์ที่ต้องการค้นหาจากผู้ใช้ มาค้นหาในดัชนีค้นหา แล้วจัดเรียงลำดับผลลัพธ์ให้ได้ตรงกับความต้องการของผู้ใช้ให้มากที่สุด โดยอาศัยข้อมูลที่ได้เตรียมไว้ก่อนแล้วจาก โปรแกรมอินเด็กเซอร์

การค้นหาแต่ละครั้งนั้น โปรแกรมค้นหาจะนำคำแต่ละคำในชุดวลีค้นหา (Query string) ที่ผู้ใช้ป้อนเข้ามาเปรียบเทียบกับตารางอินเวิร์คอินเด็กซ์ เพื่อหาเอกสารที่มีคำศัพท์นั้นๆปรากฏ หลังจากที่ได้ชุดของเอกสารที่มีคำแต่ละคำมาแล้ว โปรแกรมค้นหาก็จะนำชุดของเอกสารแต่ละชุดมารวมเข้าด้วยกัน แล้วจัดเรียงลำดับของเอกสารที่มีค่าน้ำหนักการค้นหาที่ดีที่สุด ไปน้อยที่สุด หลังจากนั้น โปรแกรมค้นหาจะส่งข้อมูลของเอกสารที่ได้มีการจัดเรียงลำดับเรียบร้อยแล้วไปยังผู้ใช้

เทคนิคที่ใช้ในการจัดอันดับก็คือ จะอาศัยจำนวนคำศัพท์และลักษณะรูปแบบของคำศัพท์นั้น ยกตัวอย่างเช่น หากมีเอกสารอยู่สองเอกสารคือ เอกสาร A และเอกสาร B เอกสาร A มีคำว่า network ปรากฏอยู่ทั้งหมด 4 แห่งในเอกสาร เอกสาร B มีคำว่า network ปรากฏอยู่ทั้งหมด 10 แห่ง เพราะฉะนั้นหากยูสเซอร์ค้นหาคำว่า network เสิร์จเอนจินก็จะแสดงผลลัพธ์ โดยที่ลำดับของเอกสารนี้เป็นเอกสารที่ส่งมอบไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

เอกสาร B จะอยู่ก่อนเอกสาร A

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการทำงานเดียวกันลักษณะรูปแบบของคำศัพท์ที่มีผลในการจัดอันดับคือ หากเอกสาร A มีคำว่า network อยู่ที่ Title ของเอกสารแต่เอกสาร B มีคำว่า network อยู่ที่ท้ายเอกสาร เพราะฉะนั้นหากผู้ใช้ค้นหาคำว่า network ลำดับของผลลัพธ์ที่ได้เอกสาร A ก็ควรอยู่ก่อนเอกสาร B

ชนิดของ Tag	นัยความสำคัญ
URL	10
TITLE	100
ANCHOR	20
META	50
B , 1 , U , H1 , H2 , H3	20
Font	1

ตารางที่ 2.1 ตัวอย่างการกำหนดค่านัยสำคัญของแต่ละ TAG

2.7 หลักการทำงานของภาษา Python

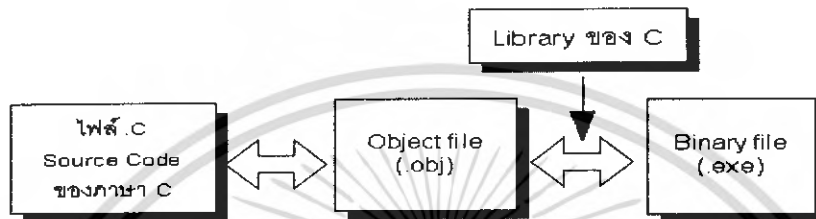
การใช้ภาษา Python ในการพัฒนา web application นี้เพราะ Python เป็นภาษาระดับสูง ภาษาหนึ่งที่เป็น object-oriented ที่มีความสามารถสูงถูกสร้างขึ้นในปี 1989 โดย Guido van Rossum ที่รวมเอาความสามารถแบบ dynamic ซึ่งทำให้เป็นที่น่าสนใจในการพัฒนา application ได้อย่างรวดเร็ว เช่นเดียวกับการใช้ในรูปแบบสคริปต์เพื่อเชื่อมต่อกับ component ที่มีอยู่แล้วเข้าด้วยกัน ทั้งนี้ยังเป็นภาษาที่ง่ายต่อการเรียนรู้ และยังสนับสนุน โมดูลและแพ็คเกจต่างๆซึ่งทำให้สามารถนำมาประกอบใช้ได้ใหม่ ซึ่งภาษานี้พัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ UNIX, Linux, Windows NT/2000/XP/2003 หรือแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาดังนี้เป็นภาษาลักษณะ Open Source ที่แจกจ่ายให้ใช้งานฟรี

เมื่อทำการเขียนโค้ดขึ้นมาตามโครงสร้างของโปรแกรมภาษาใดก็ตาม และการจะให้ได้คำสั่งเหล่านั้นทำงานได้ก็จะต้องมีตัวแปรภาษามาจัดการแปลงโค้ดคำสั่ง เพื่อให้ทำงานตามที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

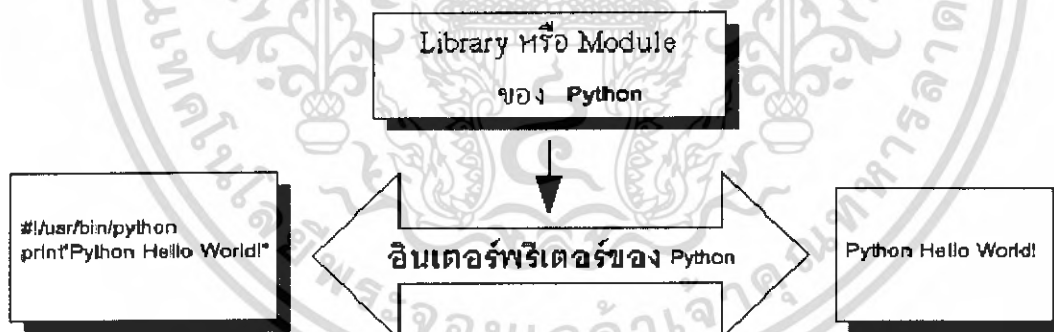
โดยลักษณะของตัวแปลภาษานั้นแบ่งได้ 2 ประเภทใหญ่ๆ คือ

2.7.1 คอมไพเลอร์ (Compiler) เป็นตัวแปลภาษาสำหรับภาษา C, C++, Pascal การทำงานก็คือจะตรวจสอบความผิดพลาดของโค้ดคำสั่งตั้งแต่ต้นจนจบก่อน หรือเรียกว่าการคอมไพล์ ถ้าไม่มีข้อผิดพลาดก็จะทำการแปลโค้ดคำสั่งให้เป็นไฟล์นามสกุล .obj (object file) จากนั้นก็ทำการแปลไฟล์ .obj ให้เป็นไบนารีไฟล์ .exe เพื่อทำงานต่อไป ดังตัวอย่างการทำงานของคอมไพเลอร์ภาษา C ดังรูป



รูปที่ 2.4 การทำงานของคอมไพเลอร์

2.7.2 อินเทอร์พรีเตอร์ (Interpreter) จะทำงานเป็นบรรทัดต่อบรรทัด คืออ่านโค้ดคำสั่งมาบรรทัดหนึ่งแล้วก็ทำงานให้ผลออกมาเลย ดังแสดงในรูป



รูปที่ 2.5 การทำงานของอินเทอร์พรีเตอร์

จากรูปตัวอย่างในกรณีที่มีการเรียกใช้ฟังก์ชันจากไลบรารี (Library) หรือ โมดูล (Module) ของภาษา Python อินเทอร์พรีเตอร์ของภาษา Python ก็จะไปทำการเรียกฟังก์ชันเหล่านั้นให้ทำงานแล้วจึงแสดงผลการทำงานออกมา

ในส่วนประสิทธิภาพการทำงานนั้นตัวแปลภาษาแบบคอมไพเลอร์จะทำงานได้เร็วกว่าตัวแปลภาษาแบบอินเทอร์พรีเตอร์เพราะโค้ดคำสั่งถูกคอมไพล์และลิงค์โดยตัวแปลภาษาแบบเอกคอมไพเลอร์ผ่านแล้วได้เป็นไฟล์ .exe ออกมาต่อจากนั้นก็เป็นส่วนของการทำงานเพียงอย่างเดียว การคำนวณที่หนักๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อีกสิ่งหนึ่งที่ใช้ในการพัฒนา คือ Regular Expression (หรือ REs) นั้นเป็นภาษาโปรแกรมมิ่งชนิดพิเศษที่มีขนาดเล็กโดยถูกฝังตัวไว้ภายใน Python โดยการใช้งาน Regular Expression นี้ต้องทำการระบุกฎเกณฑ์สำหรับกลุ่มของสตริงที่เป็นไปได้ที่ต้องการจะทำการตรวจจับ โดยที่กลุ่มสตริงเหล่านี้อาจจะเป็นประโยคภาษาอังกฤษ เช่น e-mail address เป็นต้น และยังสามารถที่จะหาความเหมือนของรูปแบบสตริงที่กำหนดว่าอยู่ในตำแหน่งใดหรือใช้ REs ในการแก้ไขสตริงหรือแยกเป็นส่วนๆ ได้หลายวิธี โดยนำ Regular Expression มาใช้ในส่วนของการ application ดังนี้

เมื่อต้องการที่จะดึงแท็ก Html ที่เกี่ยวข้องกับ link ของรูปภาพในเว็บไซต์ โดยในที่นี้เป็นการจะเป็นการดึงมาจาก <http://www.sanook.com> แท็กที่อยู่ในหน้านั้นเป็นดังนี้

```

```

ในตอนแรกต้อง import re ซึ่งเป็น Library ของภาษา Python ที่ไว้เรียกใช้งาน Regular Expression และเพื่อที่จะได้ส่วน URL ที่ต้องการโดยเมื่อจะทำงานต้องใช้ฟังก์ชัน match คำของ Regular Expression คือ re.compile("src=(|)"[S+][^a-zA-Z0-9-]+*(.jpg|.bmp)(|)"*"", re.IGNORECASE) โดยที่ ""src=(|)"[S+][^a-zA-Z0-9-]+*(.jpg|.bmp)(|)"*"" เป็น pattern ที่ต้องการจะ match link ของรูปภาพ คือ src ในภาษา html โดยต้องสร้าง pattern นี้ขึ้นมาเอง ต่อมาต้องค้นหา pattern ที่ต้องการในหน้าเว็บเพจนั้น โดยการใช้ฟังก์ชัน search ค้นหา re.search(<pattern>, line) โดยที่ <pattern> เป็น pattern ที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาถึงขั้นตอนนี้จะได้ผลลัพธ์ออกมาคือ "/story_picture/s/03542_001.jpg" เมื่อได้ผลลัพธ์ที่เป็น URL แบบนี้ออกมาแล้วจะยังไม่สามารถนำ URL นี้ไปเปิดในหน้าเว็บเพจได้ จึงต้องทำการรวม URL ในส่วนของ URL ที่เป็นหน้าหลักของเว็บนี้ และในส่วนผลลัพธ์ที่ได้จากการตัดคำมา โดยใช้ฟังก์ชันการทำงานของ Python คือ urljoin(site,match_jpg_name) โดยที่ site คือส่วนหน้าหลักของเว็บนั้น และ match_jpg_name คือส่วนผลลัพธ์ที่ได้มาจากการตัดคำได้ผลลัพธ์ที่เป็น URL ที่ใช้ในการลิงค์ไปหารูปภาพเป็น

http://www.sanook.com/story_picture/s/03542_001.jpg ซึ่งสามารถนำผลลัพธ์ที่ได้มาไปใช้ในส่วนอื่นๆต่อไป

ในส่วนของการดึงแท็ก Html ที่เกี่ยวข้องกับ link ย่อยๆ เช่น

```
<a href="http://www.kmitnb.ac.th" target="_blank"> ต้องใช้ฟังก์ชัน match คำของ Regular Expression คือ
```

```
re.compile("href=(|)"[S+][^a-zA-Z0-9- ]+*(.php|.com|.shtml|.html|.htm|.cgi|.th)*(|)"*"", re.IGNORECASE) โดยที่
```

```
""href=(|)"[S+][^a-zA-Z0-9- ]+*(.php|.com|.shtml|.html|.htm|.cgi|.th)*(|)"*""
```

เป็น pattern ที่ต้องการจะ match ลิงค์ย่อยในหน้าเว็บเพจ คือ href ในภาษา Html โดยที่ จะต้องสร้าง pattern นี้ขึ้นมา ต่อมาต้องค้นหา pattern ที่ต้องการในหน้าเว็บเพจนั้น โดยการใช้ ฟังก์ชัน search ค้นหา re.search(<pattern>,line) ได้ผลลัพธ์เป็น <http://www.kmitnb.ac.th>

2.8 ข้อมูลภาพดิจิทัล (Digital Image Data)

การทำการประมวลผลรูปภาพต้องศึกษาพื้นฐานของรูปภาพก่อนว่ามีการสร้างขึ้นได้อย่างไร ซึ่งในรูปภาพแต่ละรูปจะมีองค์ประกอบต่างๆมากมายแต่องค์ประกอบที่จะนำมาประมวลผลนั้นเรียกว่า “พิกเซล (Pixel)” พิกเซลนั้นจะมีลักษณะเป็นจุดสี่เหลี่ยม ซึ่งเป็นองค์ประกอบเล็กๆที่มีอยู่มากมายบนภาพการประมวลผลภาพต้องนำพิกเซลแต่ละรูปมาทำการประมวลผล โดยที่แต่ละพิกเซลจะประกอบด้วยรายละเอียดสีต่างๆมารวมกันเป็นพิกเซล ซึ่งเรียกกลุ่มของสีที่ใช้ว่า “Color Space” ซึ่งรูปแบบการบอกรายละเอียดของสีบนรูปภาพ โดยกลุ่มสีที่แสดงผลนั้นมีอยู่หลายแบบในที่นี้จะกล่าวถึงกลุ่มสีที่แสดงผลที่ใช้ในงานวิจัยนี้เท่านั้น

การเกิดภาพต่างๆบนจอคอมพิวเตอร์จะเกิดจากการเรียงตัวของจุดสี่เหลี่ยมสีต่างๆที่เรียกว่า “พิกเซล” ซึ่ง “พิกเซล” เป็นตัวกำหนดความคมชัดของภาพในจอคอมพิวเตอร์ ถ้ากำหนดให้ภาพแต่ละภาพมีจำนวน “พิกเซล” มากภาพก็จะมีรายละเอียดมากขึ้นตามไปด้วย แต่จะทำให้ไฟล์มีขนาดใหญ่ขึ้นด้วยตามรายละเอียดของภาพ

พิกเซล (Pixel มาจากคำว่า "Picture Element") เป็นหน่วยพื้นฐานของสีที่สามารถโปรแกรมได้บนจอแสดงผลหรือบนภาพ ขนาดทางกายภาพของพิกเซลจะขึ้นอยู่กับข้อกำหนด ความละเอียดของจอภาพ ดังนั้นถ้าต้องการแสดงผลให้มีความละเอียดสูงสุดขนาดทางกายภาพของพิกเซลก็จะเท่ากับขนาดทางกายภาพของจุด (Dot Pitch หรือ Dot Size) บนจอภาพแต่ถ้าตั้งความละเอียดไว้ต่ำกว่าค่าสูงสุด พิกเซลจะมีขนาดใหญ่กว่าขนาดทางกายภาพของจุดบนจอภาพ นั่นคือแต่ละพิกเซลจะใช้มากกว่าหนึ่งจุดในการแสดงผล

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

2.8.1 คุณสมบัติของกลุ่มสีที่แสดงผล

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

2.8.1.1 Uniformity

Uniformity เป็นคุณสมบัติที่ใช้ในการเปรียบเทียบความใกล้เคียงของเมตริกสี โดยกลุ่มสีที่แสดงผลจะมีคุณสมบัตินี้จะต้องคำนวณได้ง่าย อย่างเช่น การหาความใกล้เคียงของเมตริกสีโดยใช้ฟังก์ชันที่ไม่ขึ้นกับตำแหน่งในกลุ่มสีนั้น การทำการแปลงสีจะเป็นวิธีที่ทำให้กลุ่มสีมีคุณสมบัตินี้

2.8.1.2 Completeness

Completeness เป็นคุณสมบัติสำคัญที่กลุ่มสีที่แสดงผลจะต้องมีสีที่สามารถแยกแยะได้อย่างชัดเจน สามารถทำได้โดยการทำการลดระดับสี

2.8.1.3 Compactness

Compactness เป็นคุณสมบัติที่สีในกลุ่มสีที่แสดงผลจะต้องแตกต่างจากสีอื่นๆ ในกลุ่มสีที่แสดงผลเดียวกัน คุณสมบัตินี้จะเป็นตัวกำหนดมิติของภาพเพื่อให้แน่ใจว่า เมื่อผ่านการเปรียบเทียบจากกลุ่มสีที่แสดงผล RGB แล้วจะไม่มีสี 2 สีที่เหมือนกัน การลดระดับสีจะทำให้เกิดคุณสมบัตินี้

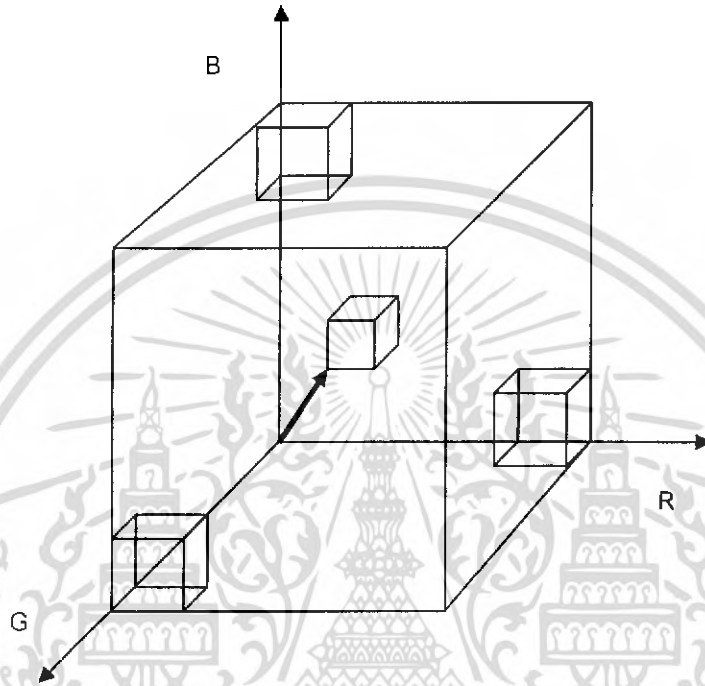
2.8.1.4 Naturalness

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

- ความสว่าง คือ คุณสมบัติที่มีผลกระทบต่อสีในด้านของแสง เช่น การเปลี่ยนแปลงระหว่างความสว่างถึงความมืด (Bright to Dim)
- ระดับสี คือ คุณสมบัติที่มีผลกระทบต่อสีที่จะทำให้สีเหมือนกับ สีแดง สีเหลือง สีเขียว สีน้ำเงิน สีฟ้าสีม่วง ฯลฯ
- ความเข้มของสี คือ คุณสมบัติที่มีผลกระทบต่อสีหนึ่งของระดับสี โดยพิจารณาถึงปริมาณของสี (Color Light) โดยไม่สนใจถึงความสว่าง

2.8.2 กลุ่มสีแสดงผลแบบ RGB

ภาพสีที่อยู่ในรูปแบบของ RGB จะเป็นรูปแบบพื้นฐานของภาพที่ใช้แสดงบนจอคอมพิวเตอร์ โดยแต่ละ พิกเซลของภาพจะประกอบด้วยองค์ประกอบของสี 3 ค่า คือสีแดง, สีเขียว, สีนํ้าเงิน โดยแสดงองค์ประกอบสีสีละ n บิต ดังนั้นจะสามารถแสดงเฉดสีได้องค์ประกอบสีละ 2^n เฉดสีที่แตกต่างกัน รูปที่ 2.4 แสดงความสัมพันธ์ของสีแดง สีเขียว และสีนํ้าเงินใน 1 พิกเซล



รูปที่ 2.6 ความสัมพันธ์ของสีในกลุ่มสีแสดงผลแบบ RGB

ในระบบแกน R จะแสดงปริมาณของสีแดงที่มีในพิกเซลนั้น แกน G จะแสดงปริมาณของสีเขียวที่มีในพิกเซลนั้น และแกน B จะแสดงปริมาณของสีนํ้าเงินที่มีในพิกเซลนั้น เนื่องจากเก็บสีแยกเป็นคณค่าของสีกันทำให้เมื่อมีการคำนวณต้องคำนวณแยกทีละค่าของสี

การผสมสีบนคอมพิวเตอร์นั้นอาศัยระดับสีหลัก 3 สีคือ สีแดง สีเขียว และ นํ้าเงิน หรือที่เรียกว่า “RGB Colors” (Red-Green-Blue) ความเหมือนจริงของสีคอมพิวเตอร์นั้น ขึ้นอยู่กับในหนึ่งจุดพิกเซลของการแสดงผลนั้นใช้ระดับของสี (Color depths) ที่มีค่าเท่าไร เช่นถ้าสี RGB มีระดับของสีเป็น 8 planes นั่นคือ ใช้ 8 บิตเก็บข้อมูลหนึ่งสี หมายความว่า เฉพาะระดับสีหลัก เช่น สีแดง สีเขียวก็จะมีค่าเป็นแดง อยู่ถึง $2^8 = 256$ ระดับ อธิบายง่ายๆก็ คือ ถ้าจะผสมสี หนึ่งสี จากแดง-เขียว-นํ้าเงิน (RGB) จะใช้สีแดง ก็ส่วนจาก 0 ถึง 255 ส่วน ใช้เขียว ก็ส่วนจาก 0 ถึง 255 และเช่นเดียวกัน สีนํ้าเงินก็ส่วน จาก 0 ถึง 255 สีที่เกิดขึ้นก็จะเกิดจากการผสมของสีทั้งสาม ในอัตราส่วนต่างๆ กัน เช่น สีเหลืองธรรมชาติ เกิดจากการผสมสี แดง 255 ส่วน สีเขียว 255 ส่วน และสีนํ้าเงิน 0 ส่วน

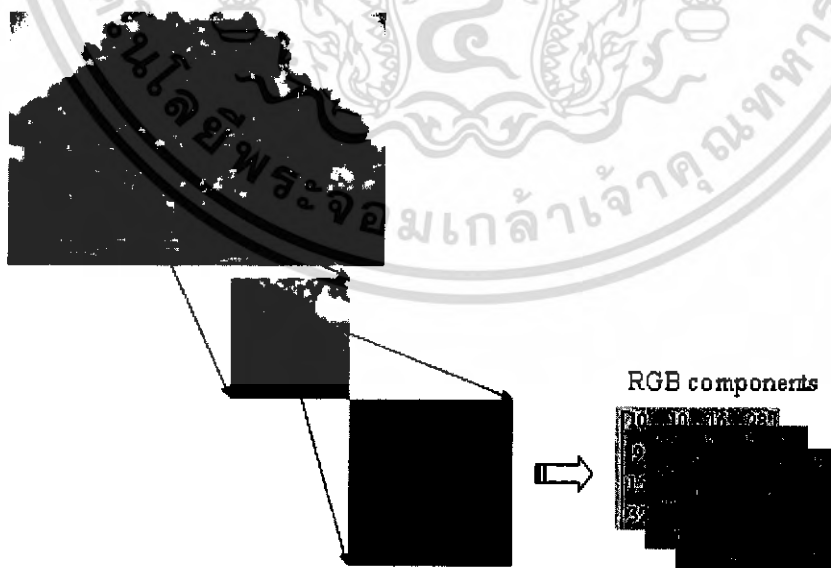
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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 ตารางที่ 2.2 แสดงถึงจำนวนสีที่เป็นไปได้ ซึ่งขึ้นกับจำนวนบิตที่ใช้กำหนดระดับของสี และรูปที่ 2.8 แสดงการผสมสีของแม่สีในระบบ RGB

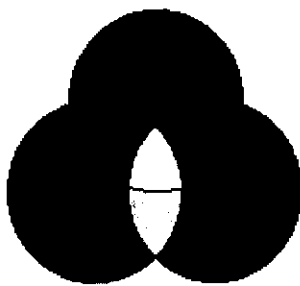
จำนวนบิตที่ใช้เก็บสีต่อหนึ่งจุด	Color Mode Name	จำนวนสีที่แสดงได้
1	Black and White	2
4	16-Color(EGA)	16
8	Pseudo Color	256
16	Hi-Color	65,536
24	True-Color	16,777,216

ตารางที่ 2.2 การกำหนดระดับสี

ค่าในแต่ละพิกเซลของสีของรูปภาพจะประกอบไปด้วยเวกเตอร์(Vector) ที่แสดงค่าของสีแดง สีเขียว และสีน้ำเงิน อย่างละ 8 บิต ดังนั้นรูปภาพ RGB 1 พิกเซล จะประกอบไปด้วยจำนวนบิตทั้งหมด 24 บิต ดังรูปที่ 2.7



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ในการเรียนการสอนเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
รูปที่ 2.7 ภาพ RGB และค่าในพิกเซลของภาพ RGB
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



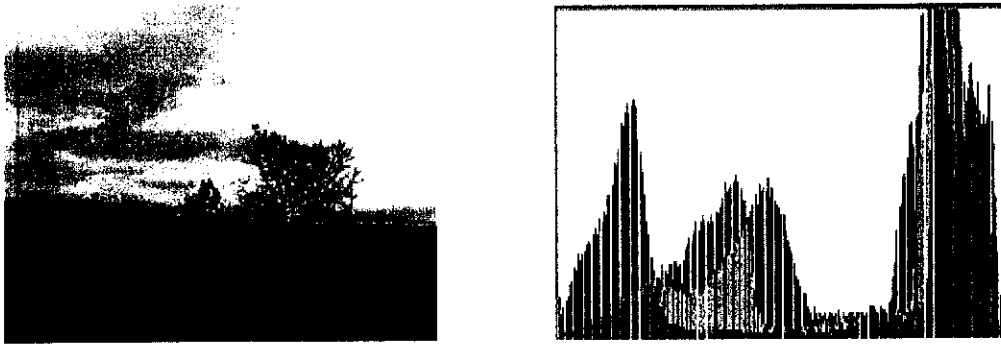
RGB

รูปที่ 2.8 การผสมสีในรูปแบบ RGB

2.9 ฮิสโตแกรม (Histogram)

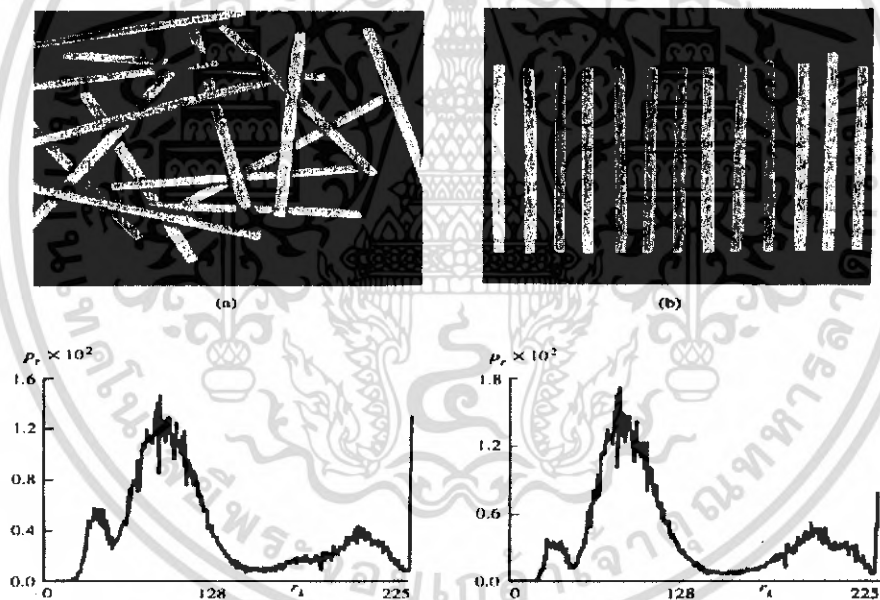
ฮิสโตแกรมของสีประกอบด้วยระดับต่างๆของสี ซึ่งระดับต่างๆนี้จะมีจำนวนของสีในแต่ละระดับที่ปรากฏในภาพ ฮิสโตแกรมของภาพสามารถอธิบายความมืด ความสว่าง ความคมชัดสูงหรือต่ำ ฮิสโตแกรมของสีจะมีความทนทานต่อการเปลี่ยนมุมมองของภาพเนื่องจากการเปลี่ยนมุมมองจะทำให้จำนวนพิกเซลรวมของสีไม่เกิดการเปลี่ยนแปลงมากนัก แต่ถ้ารูปร่างของภาพเมื่อมีการเปลี่ยนแปลงมุมมองแล้วจะเกิดการเปลี่ยนแปลงไ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

ฮิสโตแกรมจะถูกแสดงด้วยกราฟโดยค่าที่แสดงถึงความเข้มพิกเซลในแต่ละระดับ ซึ่งจะเริ่มตั้งแต่ระดับ 0 ถึง 255 ระดับของพิกเซลจะถูกวางอยู่บนแกน x ส่วนค่าที่แสดงถึงความถี่ที่เกิดขึ้นซ้ำๆกันของสีในระดับนั้นจะถูกแสดงบนแกน y รูปที่ 2.9 จะแสดงภาพที่ทำฮิสโตแกรม



รูปที่ 2.9 ภาพที่ทำฮิสโตแกรม

สำหรับฮิสโตแกรม (Global) ของภาพ 2 ภาพที่มีรูปร่างกราฟคล้ายๆกัน ไม่ได้หมายความว่าภาพทั้ง 2 จะคล้ายกัน ทั้งนี้เพราะฮิสโตแกรมของภาพจะไม่เกี่ยวข้องกับตำแหน่งของจุดภาพ กล่าวคือ จุดภาพที่มีค่าระดับสีเหมือนกัน จากกราฟฮิสโตแกรมทั้ง 2 ไม่ได้ปรากฏอยู่ที่ตำแหน่งเดียวกันของภาพ ดังนั้นภาพทั้ง 2 จึงไม่เหมือนกัน ดังแสดงในรูปที่ 2.10



รูปที่ 2.10 ภาพที่มีกราฟฮิสโตแกรมเหมือนกันแต่ภาพต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9.1 การ Quantization

ในงานวิจัยที่ได้จัดทำนี้ การสร้างฮีสโตแกรมของสีเพื่อเก็บสีที่ปรากฏอยู่ในภาพ โดยได้แบ่งค่า RGB ของรูปภาพถูกแบ่งออกเป็น 216 ระดับ (R = 6 ระดับ, G = 6 ระดับ, B = 6 ระดับ) เพื่อลดระดับความซับซ้อนของสีและเป็นการประหยัดเนื้อที่ในการจัดเก็บข้อมูลฮีสโตแกรมของแต่ละภาพ และทำให้ง่ายต่อการทำฮีสโตแกรม

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

ตารางที่ 2.3 ระดับของสีที่ใช้ในการทำ Quantization

เริ่มด้วยต้องการหาว่าในพิกเซลที่สนใจนั้นประกอบไปด้วยค่าของสีอะไรบ้าง (RGB) เท่าไหร่ แล้วนำค่าที่หาได้ไปเปรียบเทียบกับตารางว่าค่าของพิกเซลที่สนใจนั้นอยู่ในช่วงสีไหนแล้วใช้ค่า 0-5 แทนการแสดงด้วยค่าระดับสีที่แท้จริง ในสีแต่ละสีจะแบ่งค่าออกเป็นทั้งหมด 6 ช่วงสี ซึ่งเดิมจะถูกแบ่งไว้ทั้งหมด 255 สี เพราะฉะนั้นสามารถแสดงระดับสีจริงๆที่ผสมสีทั้ง 3 แล้วได้ทั้งหมด 216 ระดับ ตั้งแต่ค่า R=0, G=0, B=0 จนถึง R=6, G=6, B=6 เช่น ถ้า R = 192 , G = 64 , B = 255 ก็จะถูก quantized เป็น R = 4 , G = 1 , B = 5

2.9.2 การ Normalization

เนื่องจากขนาดของรูปภาพจะมีผลทำให้จำนวนพิกเซลในแต่ละรูปภาพไม่เท่ากัน ส่งผลให้ค่าฮีสโตแกรมของรูปภาพ 2 รูปภาพที่มีขนาดไม่เท่ากันนั้นไม่สามารถจะนำมาเปรียบเทียบกันได้ ดังนั้นจึงจำเป็นต้องทำการ Normalization เพื่อให้สามารถนำมาเปรียบเทียบกันได้ โดยเมื่อหาค่าฮีสโตแกรมของรูปภาพได้แล้ว จะต้องนำความถี่ของแต่ละระดับสีมาหารด้วยจำนวนพิกเซลทั้งหมดที่มีในภาพ ผลที่ออกมาจะได้ค่าของความน่าจะเป็น (Probability) ของแต่ละระดับสีที่มีในรูปภาพ ซึ่งเมื่อรวมค่าความน่าจะเป็นของแต่ละระดับสีเข้าด้วยกันแล้วจะมีค่าเป็น 1 โดยสมการคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

โดย

- $p(n)$ - ค่าความน่าจะเป็นของจำนวนพิกเซลที่มีระดับสีที่ n
 $f(n)$ - ค่าความถี่ของพิกเซลที่มีในระดับสีที่ n
 $P \times Q$ - จำนวนพิกเซลทั้งหมดที่มีในรูปภาพได้จากการนำจำนวน

พิกเซลตามความกว้างคูณกับจำนวนพิกเซลตามความยาว

2.9.3 ขั้นตอนการทำฮิสโตแกรม

หลังจากที่เลือกรูปภาพแล้วก็นำแต่ละภาพไปทำการ Quantization ซึ่งเป็นการลดระดับของสีลงเพื่อให้ง่ายต่อการคำนวณ โดยในงานวิจัยนี้จะแบ่งเป็น 6 ระดับของแต่ละสี คือ R 6 ระดับ G 6 ระดับ และ B 6 ระดับ รูปภาพที่ถูกทำ Quantization แล้วก็นำมาหาฮิสโตแกรมโดยนำแต่ละพิกเซลใน รูปภาพมาดูว่าอยู่ในระดับสีที่เท่าไร แล้วเก็บเป็นความถี่สะสมในแต่ละระดับสี เมื่อสแกนดูทุกพิกเซลในรูปภาพเสร็จแล้วก็จะนำความถี่สะสมของแต่ละระดับสีที่นำมาทำการ Normalized โดยการนำความถี่สะสมของแต่ละระดับสีมาหารด้วยจำนวนพิกเซลทั้งหมด แล้วจะได้ผลลัพธ์ออกมาเป็น ความน่าจะเป็นของแต่ละระดับสี และนำค่าความน่าจะเป็นของแต่ละระดับสีคูณกับ 100 เพื่อให้ได้เป็นค่าเปอร์เซ็นต์ (%) หลังจากนั้นก็ปิดให้เป็นเลขจำนวนเต็มไปเก็บไว้ในอาร์เรย์ เพื่อใช้ในการค้นหาต่อไป รูปที่ 2.11 แสดงขั้นตอนการทำฮิสโตแกรม



รูปที่ 2.11 ขั้นตอนการทำฮิสโตแกรม

บทที่ 3

กระบวนการค้นหารูปภาพโดยใช้เสิร์จเอนจิน

3.1 โครงสร้างฐานข้อมูล

แบ่งเป็น 2 ตาราง (table) คือ table_image และ table_link ดังนี้

3.1.1 table_image

id	urlimage	histogram	date	chis
1	http://www.postgang.com/prekapook/thumb/55.jpg	[1, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
2	http://www.kapook.com/images/wantjob2.jpg	[10, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
3	http://www.kapook.com/images2/chomsoan.jpg	[43, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
4	http://guide.kapook.com/images2/kb.jpg	[0, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
5	http://www.kapook.com/images2/banmanda.jpg	[2, 0, 0, 0, 0, 0, 2, 0, ...	2006-02-15	1
6	http://www.kapook.com/images2/chang.jpg	[11, 4, 0, 0, 0, 1, 0, ...	2006-02-15	1
7	http://www.kapook.com/images2/sonywalkman00001.jpg	[18, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
8	http://www.kapook.com/google/images/photo001.jpg	[0, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
9	http://www.kapook.com/images/mo.jpg	[0, 0, 0, 0, 0, 0, 0, 0, ...	2006-02-15	1
10	http://www.kapook.com/images/alive2bthumb.jpg	[9, 0, 0, 0, 0, 0, 4, 0, ...	2006-02-15	1
...				

ตารางที่ 3.1 ตัวอย่างตาราง table_image ที่เก็บ id, url, ค่าฮิสโตแกรม, วันที่ และ flag ของแต่ละภาพ โดย

- id - เป็นฟิลด์ตัวเลขที่เป็น auto increment เพื่อกำหนดหมายเลขของภาพ เช่น 1, 2, 3, ... จะมีขนาดเท่ากับจำนวนภาพที่มีอยู่ทั้งหมดในฐานข้อมูล
- urlimage - เป็นฟิลด์ชื่อของลิงค์ที่เป็นที่อยู่ของภาพในเครือข่ายอินเทอร์เน็ต
- histogram - เป็นฟิลด์ตัวเลขที่ใช้เก็บค่าฮิสโตแกรมของแต่ละภาพ
- date - เป็นฟิลด์ที่แสดงวันที่ทำการเก็บภาพมา
- chis - เป็นฟิลด์ที่ใช้เพื่อทำการเช็คว่าการหาค่าฮิสโตแกรมแล้วหรือไม่

* ใช้ id และ urlimage เป็น Primary Key ร่วมกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 table_link

id	urllink	date
1	http://my.kapook.com	2006-02-15
2	http://mobile.kapook.com	2006-02-15
3	http://bangkok.kapook.com	2006-02-15
4	http://guide.kapook.com	2006-02-15
5	http://gamecenter.kapook.com	2006-02-15
6	http://radio.prettygang.com	2006-02-15
7	http://picpost.kapook.com	2006-02-15
8	http://music.kapook.com	2006-02-15
9	http://download.kapook.com	2006-02-15
10	http://event.kapook.com	2006-02-15
...		

ตารางที่ 3.2 ตัวอย่างตาราง table_link ที่เก็บ id, urllink และวันที่เก็บของแต่ละลิงค์


โดย

- id - เป็นฟิลด์ตัวเลขที่เป็น auto increment เพื่อกำหนดหมายเลขของลิงค์ย่อยที่อยู่ภายในลิงค์หลัก
- urllink - เป็นฟิลด์ของชื่อลิงค์ย่อยที่อยู่ภายในลิงค์หลัก
- date - เป็นฟิลด์ที่แสดงวันที่เข้าสู่ลิงค์นั้นมา

* ใช้ id และ urllink เป็น Primary Key ร่วมกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 table_id_proces



id	start	stop
1	0	0
2	1	100
3	101	200
4	201	300
5	301	400
6	401	500
7	501	600
8	601	700
9	701	800
10	801	900
...		

ตารางที่ 3.3 ตัวอย่างตาราง table_id_process ที่เก็บค่าโปรเซสที่ทำงานกับ urlimage เพื่อประสานการทำงานระหว่างกัน

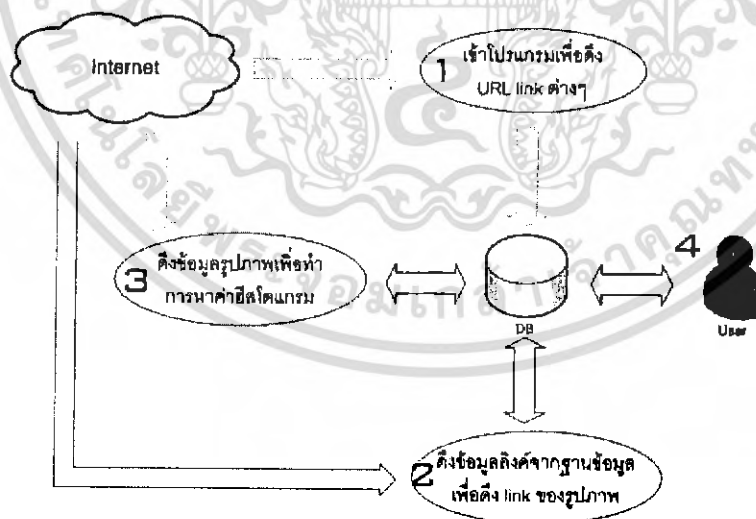
โดย

id	-	เป็นฟิลด์ตัวเลขที่เป็น auto increment เพื่อกำหนดหมายเลขของโปรเซสที่ทำงานจัดการกับภาพ
start	-	เป็นฟิลด์ตัวเลขที่เก็บ id ของภาพเริ่มต้น ที่แต่ละโปรเซสทำงาน
stop	-	เป็นฟิลด์ตัวเลขที่เก็บ id ของภาพสุดท้าย ที่แต่ละโปรเซสทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การออกแบบและพัฒนาระบบค้นหารูปภาพโดยใช้เว็บเสิร์จเอนจิน

จะทำงานในด้านของเครื่องเซิร์ฟเวอร์โดยใช้โปรแกรม Web crawler ที่พัฒนาขึ้นมาด้วย ภาษา Python ทำการกำหนดเว็บไซต์ที่ต้องการเข้าไปค้นหา และเก็บรูปภาพลงไปโดยโปรแกรมจะทำงานโดยอัตโนมัติซึ่งจะตรวจสอบซอสโค้ดของเว็บไซต์นั้น โดยเริ่มจากโปรแกรมแรกจะทำหน้าที่หา URL ซึ่งเป็นลิงค์ไปยังหน้าอื่นๆ นั่นคือเมื่อพบ tag ที่เป็นโค้ดของลิงค์ก็จะทำการตัดเอาส่วนของ URL ของลิงค์นั้นไปเก็บไว้ในฐานข้อมูล ซึ่งจะเก็บไปเรื่อยๆจนกระทั่งหมดหน้านั้น จากนั้นก็จะทำการเข้าไปยังหน้าถัดไปโดยเลือกจากค่า URL ที่เก็บไว้ในฐานข้อมูล ซึ่งจะทำการเก็บข้อมูลไปเรื่อยๆตามข้อมูลในฐานข้อมูล จากนั้น โปรแกรมที่สองจะเริ่มทำงานโดยทำการเลือก URL จากฐานข้อมูลเพื่อเข้าไปยังหน้านั้นๆแล้วทำการค้นหารูปภาพ นั่นคือเมื่อพบ tag ที่เป็นโค้ดของไฟล์รูปภาพก็จะทำการตัดเอาเฉพาะส่วนที่เป็น URL ของรูปภาพนั้น ด้วยการใช้ Regular Expression จากนั้นจะทำการเก็บข้อมูลรูปภาพไว้ในฐานข้อมูล จากนั้นโปรแกรมที่สามจะเริ่มทำงานโดยการเลือกรูปภาพจากฐานข้อมูล เพื่อนำมาเข้าฟังก์ชันการหาค่าฮีสโตแกรมของภาพ ทั้งนี้จะบันทึกวันที่เก็บข้อมูลต่างๆลงในฐานข้อมูลด้วยเพื่อใช้ในการอัปเดตข้อมูลในภายหลัง ส่วนโปรแกรมที่สี่เพื่อให้ผู้ใช้มาทำการค้นหารูปภาพผ่านทางเว็บ ซึ่งพัฒนาด้วย PHP ร่วมกับ Python โดยใช้รูปภาพเป็นอินพุทเพื่อไปหาค่าฮีสโตแกรมจากนั้นจะไปค้นหาภาพในฐานข้อมูลที่มีค่าฮีสโตแกรมคล้ายกัน แล้วแสดงผลออกมาเป็นลิงค์ของภาพและภาพจากฐานข้อมูล ซึ่งแสดงขั้นตอนต่างๆไว้ดังนี้



รูปที่ 3.1 ขั้นตอนการทำงานของโปรแกรมอิมเมจเสิร์จเอนจิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์

ขั้นตอนที่ 1 โปรแกรมทำการอ่านค่า URL ที่อยู่ใน textfile

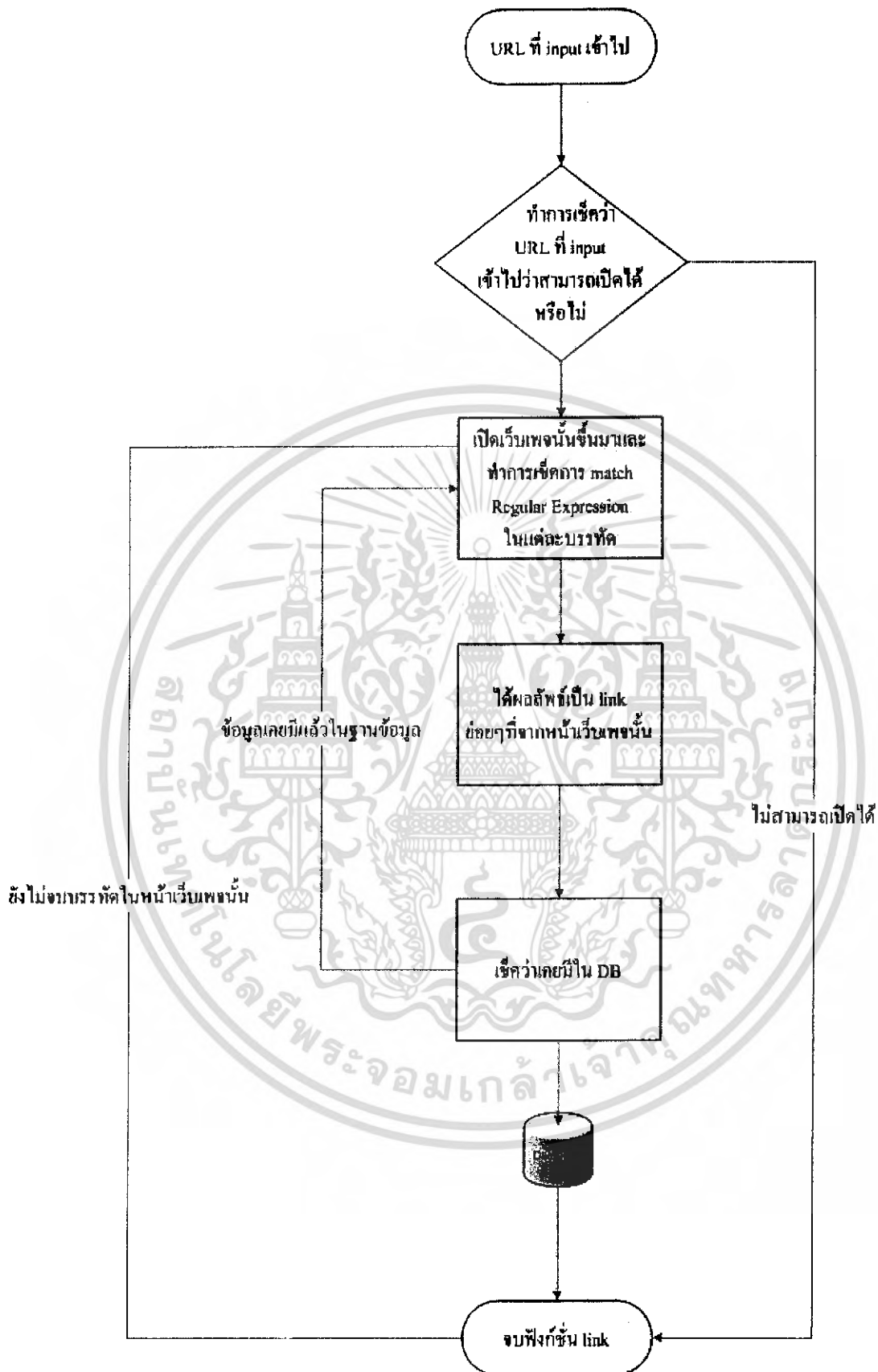
ขั้นตอนที่ 2 ทำการเช็ค URL ที่อ่านมาสามารถเปิดได้หรือไม่ เว็บเพจนั้นปิดไปแล้วหรือไม่ ถ้าไม่สามารถเปิดได้ก็จะจบฟังก์ชันการทำงาน จากนั้นโปรแกรมก็จะกลับไปอ่านค่า URL ใน textfile อีกครั้งหนึ่ง

ขั้นตอนที่ 3 เปิดหน้าเว็บเพจนั้นขึ้นมาและทำการเช็คการ match regular expression ในตอนแรก ต้อง import re ซึ่งเป็น library ของภาษา Python ที่ใช้เรียกใช้งาน Regular Expression และเพื่อที่จะได้ส่วน URL ที่เราต้องการ โดยการที่เราต้องใช้ฟังก์ชัน match ค่าของ Regular Expression คือ re.compile("""href=(|")\S+*\^[a-zA-Z0-9- _]*\.(php|.com|.shtml|.html|.htm|.cgi|.th)*(|)" """, re.IGNORECASE) โดยที่ """href=(|")\S+*\^[a-zA-Z0-9- _]*\.(php|.com|.shtml|.html|.htm|.cgi|.th)*(|)" """ เป็น pattern ที่เราต้องการจะ match ลิงค์ย่อยในหน้าเว็บเพจนั้น คือ href ในภาษา HTML โดยเราต้องสร้าง pattern นี้ขึ้นมาเอง ต่อมาจึงต้องค้นหา pattern ที่ต้องการในหน้าเว็บเพจนั้น โดยการใช้ฟังก์ชัน search ค้นหา re.search(<pattern>, line) โดยที่ <pattern> เป็น pattern ที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาในแต่ละบรรทัดในซอสโค้ดของหน้าเว็บเพจนั้น เพื่อหาลิงค์ย่อยๆที่อยู่ในหน้าเว็บเพจนั้นที่ต้องการ

ขั้นตอนที่ 4 ได้ผลลัพธ์เป็น URL ลิงค์ที่ต้องการแล้วนำไปเก็บในฐานข้อมูล

ขั้นตอนที่ 5 ทำการเช็คในฐานข้อมูลว่า URL ลิงค์นั้นเคยมีในฐานข้อมูลแล้วหรือไม่ ถ้าเคยมีแล้วก็จะกลับไปทำขั้นตอนที่ 3 อีกครั้งเพื่อหา URL ลิงค์ในบรรทัดต่อไปของซอสโค้ดในหน้าเว็บเพจนั้น ถ้ายังไม่เคยมีก็จะทำการเก็บลงในฐานข้อมูล

ขั้นตอนที่ 6 ทำการดูในซอสโค้ดในหน้าเว็บเพจนั้นว่าจบหน้าแล้วหรือไม่ ถ้าจบแล้วก็จะจบฟังก์ชันการทำงาน ถ้ายังไม่จบก็จะกลับไปขั้นตอนที่ 3 อีกครั้ง



รูปที่ 3.2 ขั้นตอนการทำงานของ โปรแกรม Web crawler ที่ใช้ในการเก็บ ลิงค์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ในระดับถัดไป

ขั้นตอนที่ 1 โปรแกรมจะทำการอ่านค่า URL จากฐานข้อมูลในตาราง table_link

ขั้นตอนที่ 2 ทำการเช็คค่า URL ที่อ่านมาสามารถเปิดได้หรือไม่เว็บเพจนั้นปิด

ไปแล้วหรือไม่ ถ้าไม่สามารถเปิดได้ก็จะจบฟังก์ชันการทำงาน แล้วโปรแกรมก็จะกลับไปอ่านค่า URL จาก table_link อีกครั้ง

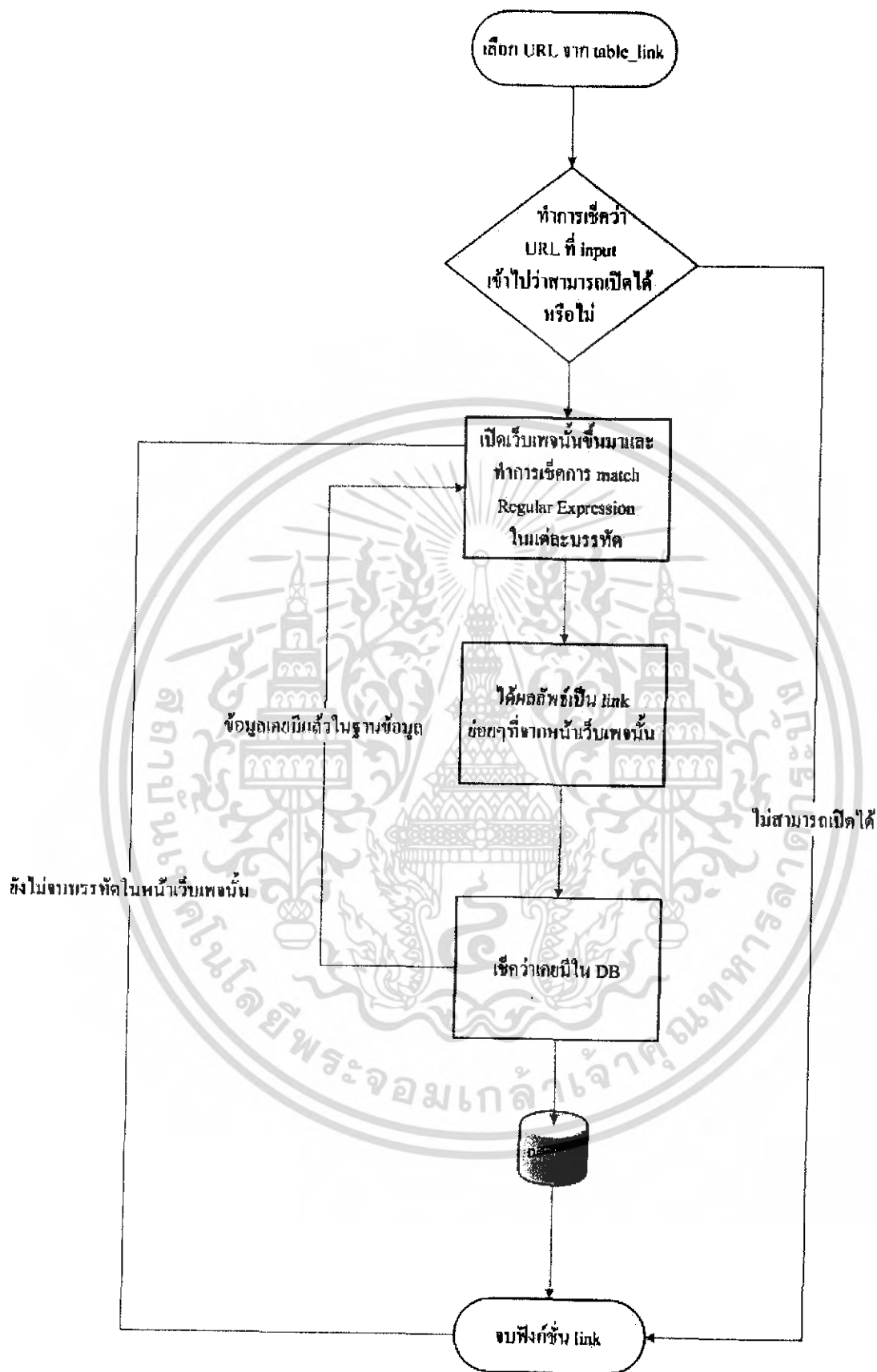
ขั้นตอนที่ 3 เปิดหน้าเว็บเพจนั้นขึ้นมาและทำการเช็คการ match regular

expression ในตอนแรก ต้อง import re ซึ่งเป็น library ของภาษา Python ที่ไว้เรียกใช้งาน Regular Expression และเพื่อที่จะได้ส่วน URL ที่ต้องการต้องใช้ฟังก์ชัน match ค่าของ Regular Expression คือ re.compile("""href="(\\S+)[^a-zA-Z0-9-]*\\.php\\.com|\\.shtml|\\.html\\.htm\\.cgi\\.th"(/)*(""" , re.IGNORECASE) โดยที่ ""href="(\\S+)[^a-zA-Z0-9-]*\\.php\\.com|\\.shtml|\\.html\\.htm\\.cgi\\.th"(/)*(""" เป็น pattern ที่ต้องการจะ match ลิงค์ย่อยในหน้าเว็บเพจนั้น คือ href ในภาษา HTML โดยต้องสร้าง pattern นี้ขึ้นมาเอง ต่อมาต้องค้นหา pattern ที่ต้องการในหน้าเว็บเพจนั้นโดยการใช้ฟังก์ชัน search ค้นหา re.search(<pattern>, line) โดยที่ <pattern> เป็น pattern ที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาในแต่ละบรรทัดในซอสโค้ดของหน้าเว็บเพจนั้น เพื่อหาลิงค์ย่อยๆที่อยู่ในหน้าเว็บเพจนั้นที่ต้องการ

ขั้นตอนที่ 4 ได้ผลลัพธ์เป็น URL ลิงค์ที่ต้องการจากนั้นจะนำไปเก็บในฐานข้อมูล

ขั้นตอนที่ 5 ทำการเช็คในฐานข้อมูล ว่า URL ลิงค์นั้นเคยมีในฐานข้อมูลแล้วหรือไม่ ถ้าเคยมีแล้วก็จะกลับไปทำขั้นตอนที่ 3 อีกครั้งเพื่อหา URL ลิงค์ในบรรทัดต่อไปของซอสโค้ดในหน้าเว็บเพจนั้น ถ้ายังไม่เคยมีก็จะทำการเก็บลงในฐานข้อมูล

ขั้นตอนที่ 6 ทำการดูในซอสโค้ดในหน้าเว็บเพจนั้นว่าจบหน้าแล้วหรือไม่ ถ้าจบแล้วก็จะจบฟังก์ชันการทำงาน ถ้ายังไม่จบก็จะกลับไปขั้นตอนที่ 3 อีกครั้ง



รูปที่ 3.3 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ในระดับถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 การทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บ ลิงค์ ของรูปภาพ

ขั้นตอนที่ 1 โปรแกรมทำการอ่านค่า URL ลิงค์จากฐานข้อมูลในตาราง table_link

ขั้นตอนที่ 2 ทำการเช็คค่า URL ที่อ่านมาสามารถเปิดได้หรือไม่เว็บเพจนั้นปิด

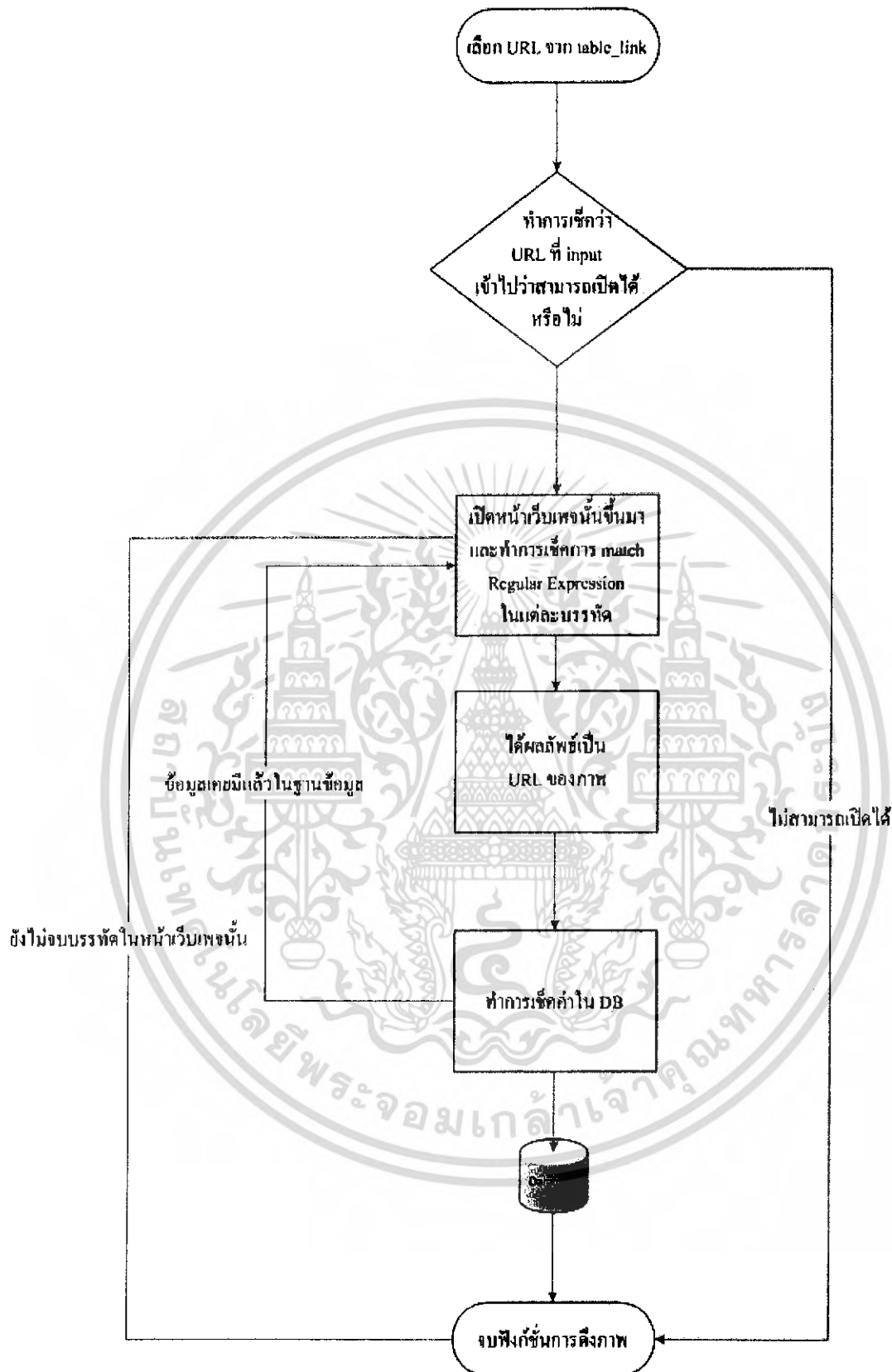
ไปแล้วหรือยัง ถ้าไม่สามารถเปิดได้ก็จะจบฟังก์ชันการทำงาน แล้วโปรแกรมก็จะกลับไปอ่านค่า URL จาก table_link อีกครั้ง

ขั้นตอนที่ 3 เปิดหน้าเว็บเพจนั้นขึ้นมาและทำการเช็คการ match Regular expression ในตอนแรก ต้อง import re ซึ่งเป็น library ของภาษา Python ที่ไว้เรียกใช้งาน Regular Expression และเพื่อที่จะได้ส่วน URL ที่ต้องการ โดยการใช้ฟังก์ชัน match ค่าของ Regular Expression คือ `re.compile("""src="(\\S+)*[\\[a-zA-Z0-9-_]+*(.jpg|.bmp)(/)*(\\S+)*""", re.IGNORECASE)` โดยที่ pattern นี้ `""src="(\\S+)*[\\[a-zA-Z0-9-_]+*(.jpg|.bmp)(/)*(\\S+)*""` เป็น pattern ที่ต้องการจะ match ลิงค์รูปภาพในหน้าเว็บเพจนั้น คือ src ในภาษา HTML โดยต้องสร้าง pattern นี้ขึ้นมาเอง ต่อมาทำการ search หา pattern ที่ต้องการในหน้าเว็บเพจนั้น โดยการใช้ฟังก์ชัน search ค้นหา `re.search(<pattern>, line)` โดยที่ `<pattern>` เป็น pattern ที่ต้องการค้นหา และ line เป็นส่วนที่จะทำการค้นหาในแต่ละบรรทัดในซอสโค้ดของหน้าเว็บเพจนั้นเพื่อหาลิงค์รูปภาพที่อยู่ในหน้าเว็บเพจที่ต้องการ

ขั้นตอนที่ 4 ได้ผลลัพธ์เป็น urlimage ที่ต้องการจากนั้นทำการเก็บในฐานข้อมูล

ขั้นตอนที่ 5 ทำการเช็คในฐานข้อมูล ว่า urlimage นั้นเคยมีในฐานข้อมูลแล้วหรือไม่ ถ้าเคยมีแล้วก็จะกลับไปทำขั้นตอนที่ 3 อีกครั้งเพื่อหา urlimage ในบรรทัดต่อไปของซอสโค้ดในหน้าเว็บเพจนั้น ถ้ายังไม่เคยมีก็จะทำการเก็บลงในฐานข้อมูล

ขั้นตอนที่ 6 ทำการเช็คในซอสโค้ดในหน้าเว็บเพจนั้นว่าจบหน้าแล้วหรือยัง ถ้าจบแล้วก็จะจบฟังก์ชันการทำงาน ถ้ายังไม่จบก็จะกลับไปขั้นตอนที่ 3 อีกครั้ง



รูปที่ 3.4 ขั้นตอนการทำงานของโปรแกรม Web crawler ที่ใช้ในการเก็บลิงค์ของรูปภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

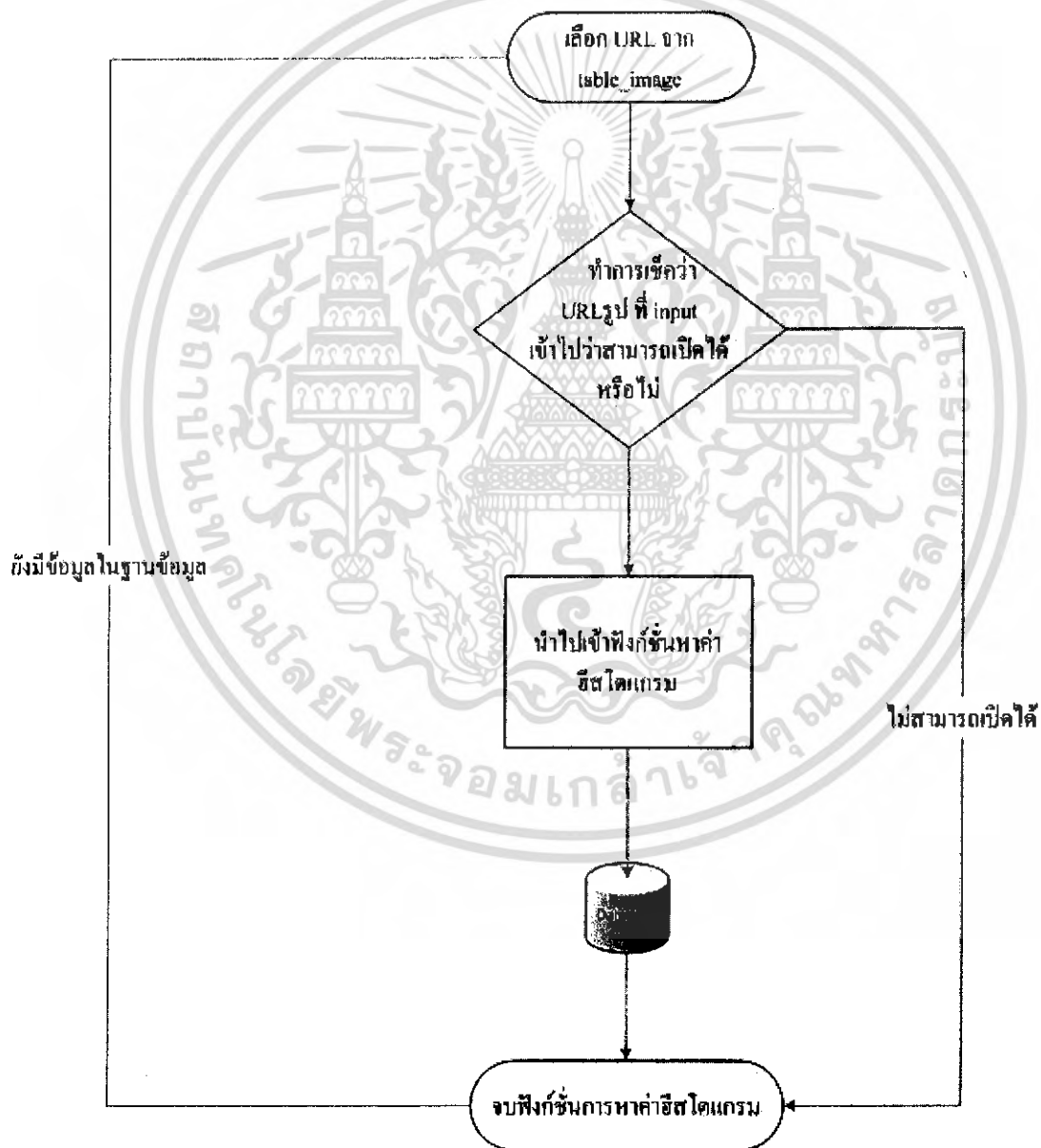
3.2.4 การคำนวณหาค่าฮีสโตแกรมของภาพ

ขั้นตอนที่ 1 โปรแกรมอ่านค่า url image จากฐานข้อมูลในตาราง table_image

ขั้นตอนที่ 2 ทำการเช็คค่า URL ที่อ่านมาสามารถเปิดได้หรือไม่เว็บเพจนั้นปิดไปแล้วหรือไม่ ถ้าไม่สามารถเปิดได้ก็จะจบฟังก์ชันการทำงาน จากนั้นโปรแกรมก็จะกลับไปอ่านค่า URL จาก table_image อีกครั้ง

ขั้นตอนที่ 3 อ่านค่าจาก urlimage และนำไปเข้าฟังก์ชันฮีสโตแกรม เมื่อได้ค่ามาก็จะจัดเก็บลงฐานข้อมูล

ขั้นตอนที่ 4 โปรแกรมทำการตรวจสอบฐานข้อมูล ใน table_image ว่ามีรูปภาพใดที่ยังไม่ได้เข้าฟังก์ชันฮีสโตแกรม

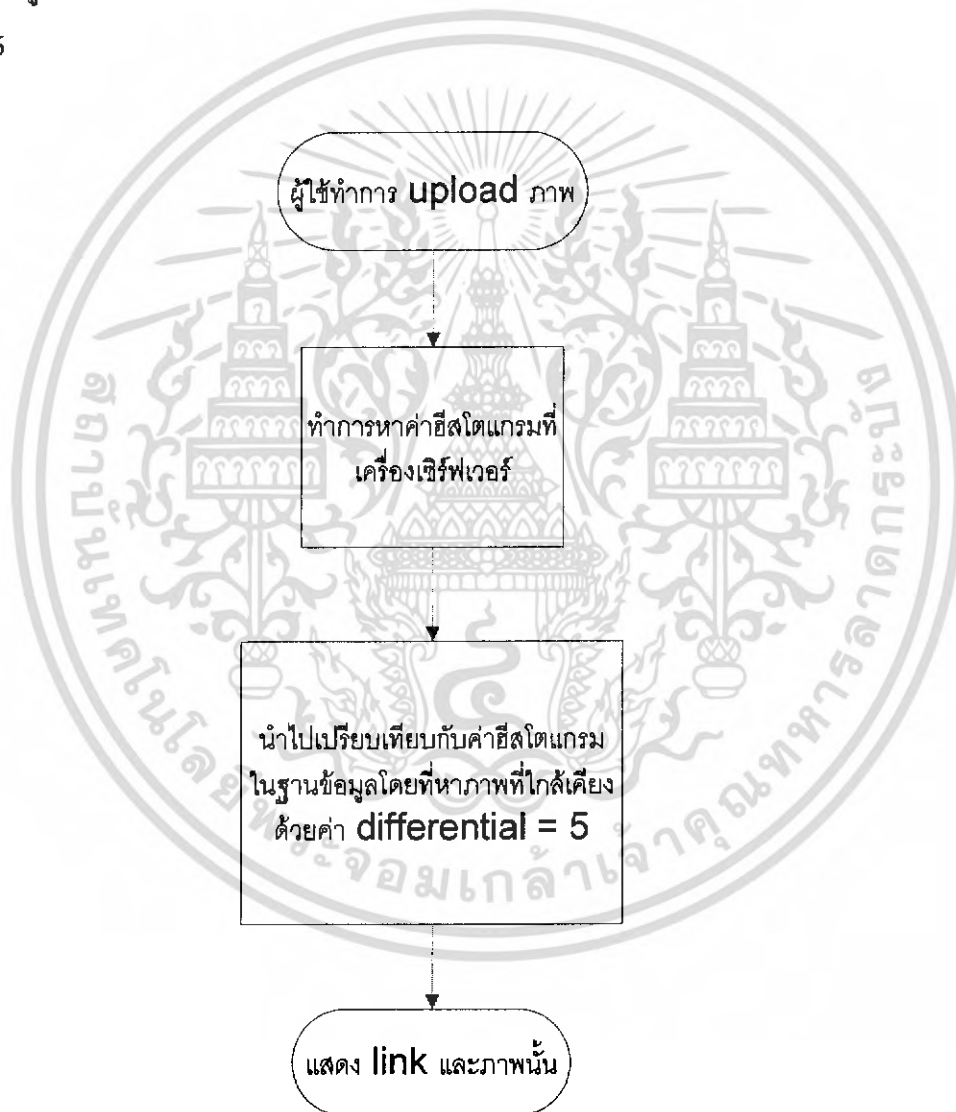


รูปที่ 3.5 ขั้นตอนการทำงานของโปรแกรมที่ใช้ในการหาคำนวณค่าฮีสโตแกรมของภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การค้นหารูปภาพของผู้ใช้

จากเครื่องผู้ใช้งานจะทำการเลือกรูปภาพจากเครื่องของคนเป็นอินเทอร์เน็ตผ่านทางหน้าเว็บเพื่อใช้ค้นหาว่ามีภาพที่เหมือนกันหรือคล้ายกันในเครือข่ายอินเทอร์เน็ตหรือไม่ จากนั้นจะทำการอัปโหลดไฟล์รูปภาพดังกล่าวขึ้นมาที่เครื่องเซิร์ฟเวอร์แล้วจึงนำไปเข้าฟังก์ชันการหาค่าฮิสโตแกรมของภาพและทำการค้นหาในฐานข้อมูล โดยที่กำหนดให้ค่า differential ในการหาภาพที่มีลักษณะสีในแต่ละแบบคล้ายกันได้ มีค่าต่างกันได้ ± 5 ค่า โดยจะเปรียบเทียบค่าฮิสโตแกรมทีละค่าถ้าหากว่ามีค่าต่างกันเกิน ± 5 ก็จะทำการข้ามไปเปรียบเทียบภาพอื่นต่อไป ถ้าพบจะแสดงลิงค์ของภาพดังกล่าวและภาพที่อยู่ตำแหน่งนั้นขึ้นมา แต่ถ้าหากไม่พบจะแจ้งให้ทราบ ซึ่งแสดงขั้นตอนต่างๆไว้ดังรูปที่ 3.6



รูปที่ 3.6 ขั้นตอนการทำงานผู้ใช้งานในการค้นหารูปภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 ทดสอบหาค่าฮีสโตแกรมจากภาพต่างๆ

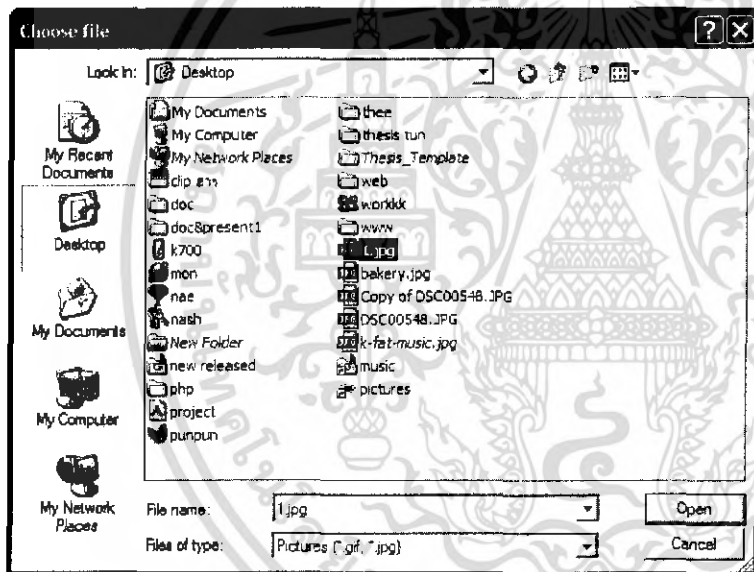
ในการทดสอบหาค่าฮีสโตแกรมจากภาพต่างๆนั้นได้ใช้โปรแกรมการหาค่าฮีสโตแกรมที่พัฒนาด้วยภาษา Python ร่วมกับ Python Image Library (PIL) ซึ่งเป็นไลบรารีที่เพิ่มความสามารถด้านอิมเมจโปรเซสซิ่ง จากนั้นเลือกรูปภาพตัวอย่างมา 5 ภาพ โดยในแต่ละภาพมีองค์ประกอบของภาพต่างกันพอสมควร โดยแต่ละภาพมีขนาดภาพ 600 x 400 และขนาดของไฟล์ภาพประมาณ 40-80 KB ซึ่งเป็นขนาดที่ใกล้เคียงกับภาพในเครือข่ายอินเทอร์เน็ต โดยจะนำภาพตัวอย่างมาย่อขนาดให้มีขนาดต่างๆ ดังนี้ 75%, 50%, 10% และภาพที่มีการหมุนไป 90° มาหาค่าฮีสโตแกรมโดยได้ค่า ซึ่งจากการทดสอบจะเห็นว่าค่าฮีสโตแกรมของภาพในชุดทดลองเดียวกันจะมีค่าที่ใกล้เคียงกัน นั่นคือ ลักษณะสีของแต่ละภาพจะมีความใกล้เคียงกันในทุกๆลักษณะสี ซึ่งจะต่างกันเพียงไม่กี่ค่าเปอร์เซ็นต์ ทำให้สามารถหาความคล้ายของภาพได้จากการหาค่าความต่าง (differential) ของค่าลักษณะสีที่เหมาะสม ซึ่งจากข้อมูลที่ได้นี้จะนำไปพัฒนาเป็นกระบวนการค้นหารูปภาพต่อไป โดยข้อมูลดังกล่าวจะแสดงดังตารางต่อไปนี้

3.5 ส่วนหน้าของผู้ใช้

IMAGE SEARCH ENGINE

Select your image to search :

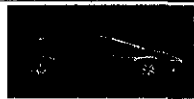
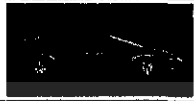
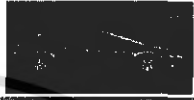
รูปที่ 3.7 หน้าแรกที่เป็นหน้าจอเพื่อให้ผู้ใช้งานทำการเลือกไฟล์รูปภาพที่จะทำการค้นหา



รูปที่ 3.8 ผู้ใช้ทำการเลือกไฟล์รูปภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img61.imageshack.us/img61/834/dvlon4zj.jpg	
http://img239.imageshack.us/img239/8683/dvlc5wz.jpg	
http://img141.imageshack.us/img141/8350/dvlcs3qz.jpg	

รูปที่ 3.9 ถ้าค้นพบรูปภาพที่มีค่าลักษณะเฉพาะที่เหมือนกันจะแสดง URL ของภาพและภาพออกมา

Search Results:

no match found

รูปที่ 3.10 เมื่อรูปภาพที่ใช้ค้นหาไม่พบรูปภาพที่มีลักษณะเฉพาะเหมือนกันในฐานข้อมูล

Your file is too big, please reduce size DSCF0843.JPG

รูปที่ 3.11 ถ้าหากไฟล์รูปภาพที่เลือกมีขนาดเกิน 1Mb จะแจ้งว่าไฟล์มีขนาดใหญ่เกินไป

Cannot search, please check your file

รูปที่ 3.12 ถ้าหากกดปุ่ม Search โดยไม่ได้เลือกไฟล์จะแจ้งให้ทราบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

4.1 ผลการทดลองเพื่อทดสอบประสิทธิภาพของเสิร์จเอนจิน

ในการทดลองการทำงานของโปรแกรม Web Crawler นั้น จะเริ่มทำงานโดยเริ่มต้นจากเว็บไซต์ <http://www.sanook.com> ซึ่งเป็นเว็บไคเร็กทอรีขนาดใหญ่มีลิงค์ที่เชื่อมต่อไปยังเว็บไซต์ต่างๆ และรูปภาพจำนวนมาก ทำให้สามารถเชื่อมต่อเพื่อไปเก็บข้อมูลลิงค์ต่างๆ และรูปภาพต่างๆ ได้มาก การทดลองนี้มีจุดประสงค์เพื่อทดสอบประสิทธิภาพการทำงานของโปรแกรมอิมเมจเสิร์จเอนจิน เนื่องจากความเร็วในการเก็บลิงค์, ลิงค์ของรูปภาพ และการทำฮิสโตแกรมมีความเร็วไม่เท่ากัน จึงต้องแยก 3 โปรแกรมออกจากกัน โดยที่เริ่มแรกจะทำการเก็บลิงค์ย่อยๆ ในหน้าเว็บเพจนั้นไว้ก่อน โดยทำงานจากโปรแกรมแรกที่ทำกรเก็บลิงค์ภายในหน้าเว็บเพจของลิงค์ที่กำหนดจาก textfile จากนั้นจะทำการบันทึกลงในฐานข้อมูล ต่อมาจะทำการนำลิงค์ย่อยๆ ที่เก็บไว้ก่อนหน้านั้นนำมาหาลิงค์ของรูปภาพโดยโปรแกรมที่สองจะทำการเก็บลิงค์ของรูปภาพในหน้าเว็บ เมื่อเก็บลิงค์ของรูปภาพได้แล้ว ขั้นตอนต่อมาจะนำลิงค์ของรูปภาพนั้นมาหาค่าฮิสโตแกรมของแต่ละภาพและเก็บเข้าไปในฐานข้อมูลตามลำดับ

ในการทดลองต้องมีข้อมูลของลิงค์และลิงค์ของรูปภาพอยู่ในฐานข้อมูลอยู่แล้วเพื่อให้ได้ผลการทดลองจากการทำงานที่ดีที่สุดของแต่ละโปรแกรม จากการทดลองได้ผลการทดลองดังนี้

จำนวนลิงค์	จำนวนลิงค์ของรูปภาพ	จำนวนชุดฮิสโตแกรมของภาพ	เวลา (นาที)
297	190	56	1
965	394	79	2
1778	471	122	3
2260	491	156	4
2394	499	195	5
2662	499	223	6
3544	556	276	7
3691	682	319	8
3967	786	373	9
4134	872	441	10

ตารางที่ 4.1 ผลการทดลองประสิทธิภาพการทำงานของ Web Crawler

จากการทดลองจะเห็นได้ว่าการทำงานในแต่ละโปรเซสมีความเร็วต่างกัน จึงต้องออกแบบให้แต่ละโปรเซสการทำงานแยกออกจากกัน โดยที่โปรเซสที่หนึ่งใช้เก็บลิงค์ โปรเซสที่สองใช้ในการเก็บลิงค์ของรูปภาพ และอีกหนึ่งโปรเซสใช้ในการทำฮิสโตแกรมโดยที่

- โปรเซสค้นหาลิงค์มีความสามารถในการค้นหาลิงค์ด้วยความเร็วประมาณ 413 ลิงค์/นาที
- โปรเซสค้นหาลิงค์ของรูปภาพมีความสามารถในการค้นหาลิงค์ของรูปภาพด้วยความเร็วประมาณ 87 ลิงค์/นาที
- โปรเซสในการทำฮิสโตแกรมสามารถทำงานด้วยความเร็วประมาณ 44 ภาพ/นาที

จะเห็นได้ว่าอัตราส่วนของการทำงานในโปรเซสค้นหาลิงค์ โปรเซสค้นหาลิงค์ของรูปภาพ และโปรเซสการทำฮิสโตแกรมจะมีค่าเป็น 10 : 2 : 1 โดยที่โปรเซสค้นหาลิงค์จะมีความเร็วในการทำงานมากที่สุด รองลงมาคือ โปรเซสค้นหาลิงค์ของรูปภาพ และท้ายที่สุดคือ โปรเซสการทำฮิสโตแกรม โดยที่อาจจะเพิ่มโปรเซสในการทำงานเป็นดังนี้คือ โปรเซสค้นหาลิงค์ของรูปภาพเป็น 5 โปรเซส โปรเซสการทำฮิสโตแกรมเป็น 10 โปรเซส แต่เนื่องจากการทดลองเพิ่มโปรเซสการทำงาน สามารถให้ CPU รองรับการทำงานได้ที่ 10 โปรเซสก็จะให้การทำงานของ CPU usage ขึ้นถึง 100 % ดังนั้นจึงสามารถแบ่งการทำงานของแต่ละโปรเซสให้มีการทำงานที่มีประสิทธิภาพมากที่สุดคือ โปรเซสค้นหาลิงค์ของรูปภาพแบ่งเป็น 5 โปรเซส โปรเซสการทำฮิสโตแกรมแบ่งเป็น 2 โปรเซส และโปรเซสการค้นหาลิงค์แบ่งเป็น 1 โปรเซส จึงจะไม่ทำให้การทำงานของ CPU usage ขึ้นถึง 100 % และเพื่อให้การทำงานของแต่ละโปรเซสไม่ช้าลงด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเพิ่มโปรเซสค้นหาลิงค์ของรูปภาพ มีการทำงาน โดยที่มีการจัดลำดับงานของหลายๆ โปรเซส ที่มาทำงานเดียวกันนี้ โดยการแบ่งเป็นช่วงๆ เช่น ในโปรเซส ที่ 1 ทำงานตั้งแต่ช่วง id ที่ 1 – 1000 ใน table_link และ โปรเซส ที่ 2 ทำงานตั้งแต่ในช่วง id ที่ 1001 – 2000 ใน table_link เป็นต้น




หมายเหตุ การค้นหาลิงค์ และลิงค์ของรูปภาพ จะมีความเร็วในการค้นหาไม่คงที่ขึ้นกับ เว็บไซต์ที่ใช้เริ่มต้นในการค้นหา เพราะว่าในแต่ละเว็บไซต์นั้นมีจำนวนของลิงค์ และลิงค์ของรูปภาพเป็นจำนวนมากหรือน้อยแตกต่างกัน เช่น ถ้าเริ่มต้นค้นหาจากเว็บ <http://www.ce.kmitl.ac.th> และ <http://www.sanook.com> จะได้ความเร็วในการค้นหาลิงค์ และลิงค์ของรูปภาพต่างกัน และ เว็บไซต์ที่ใช้เริ่มต้องเป็นเว็บไซต์ที่สามารถค้นหาข้อมูลได้ไม่ใช่เป็น deadlink

4.2 ทดลองหาค่า Differential ที่เหมาะสม

จากที่กล่าวไว้ในหัวข้อที่ 3.4 ในการหารูปที่มีลักษณะคล้ายกัน โดยต่างกันในเรื่องของขนาดภาพนั้น สามารถหาได้จากค่าลักษณะสีฮิสโตแกรมที่คล้ายกัน โดยที่ค่าสีที่แตกต่างกัน (Differential) กันในแต่ละค่าลักษณะสีนั้นจะต้องทำการทดลองเพื่อหาค่าที่สามารถใช้ในการหาภาพที่คล้ายกันให้มีประสิทธิภาพมากที่สุด เนื่องจากภาพในเว็บส่วนใหญ่อาจถูกย่อขนาดลง ทำให้ค่าฮิสโตแกรมเปลี่ยนไป ดังนั้นจึงทำการทดลองค่า Differential ตั้งแต่ 1-5 เพื่อให้ได้ค่า Differential ที่เหมาะสมที่สุดในการค้นหา โดยในการทดลองจัดทำกราฟชุดทดลองเช่นเดียวกับภาพในการทดลองที่ 3.4 ในฐานข้อมูล โดยจะปนกับภาพที่อยู่ในฐานข้อมูลเดิมที่ Web Crawler เก็บภาพมา ซึ่งมีภาพในฐานข้อมูลอยู่แล้วประมาณ 400 ภาพโดยให้ผลการทดลองดังนี้

ผลการทดลองเมื่อค้นหาภาพด้วยค่า Differential = 1




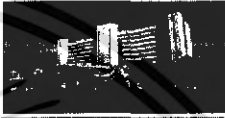

SEARCH RESULT:

Your Picture	
Source	Picture
http://img61.imageshack.us/img61/834/civcm4zj.jpg	
http://img239.imageshack.us/img239/8683/civc5wz.jpg	
http://img140.imageshack.us/img140/972/civc905kw.jpg	

รูปที่ 4.1 ภาพชุดทดลองที่ 1






เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img161.imageshack.us/img161/1818/e127lg.jpg	
http://img56.imageshack.us/img56/6455/e12m2xw.jpg	
http://img63.imageshack.us/img63/7783/e12s1rf.jpg	
http://img61.imageshack.us/img61/6358/e12904kj.jpg	

รูปที่ 4.2 ภาพชุดทดลองที่ 2


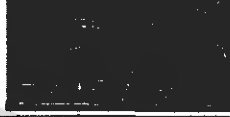
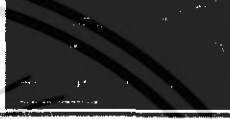
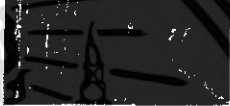
SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img156.imageshack.us/img156/5344/house4ox.jpg	
http://img62.imageshack.us/img62/8010/housem6iq.jpg	
http://img156.imageshack.us/img156/6458/houses6iy.jpg	
http://img100.imageshack.us/img100/7048/house907li.jpg	

รูปที่ 4.3 ภาพชุดทดลองที่ 3

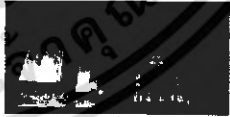
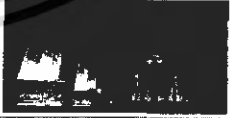


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img109.imageshack.us/img109/869/flag7xz.jpg	
http://img136.imageshack.us/img136/9581/flagm5zy.jpg	
http://img105.imageshack.us/img105/5151/flags2ax.jpg	
http://img66.imageshack.us/img66/6466/flag909vj.jpg	

รูปที่ 4.4 ภาพชุดทดลองที่ 4

SEARCH RESULT:

Your Picture	
Source	Picture
http://img86.imageshack.us/img86/9505/demo4mn.jpg	
http://img88.imageshack.us/img88/4094/demom8jh.jpg	
http://img95.imageshack.us/img95/4652/demos0ug.jpg	
http://img148.imageshack.us/img148/4193/demo909dm.jpg	





รูปที่ 4.5 ภาพชุดทดลองที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษานี้เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยค่า differential = 1 จะเห็นได้ว่าสามารถหาภาพในแต่ละชุดทดลอง ที่มีจำนวนทั้งหมด 5 ค่า โดยที่ ในชุดทดลองที่ 1 ค้นหาได้ 3 ภาพ โดยขาดภาพที่ย่อขนาดลงเหลือ 50% และ 10% , ชุดทดลองที่ 2 ค้นหาได้ 4 ภาพ ,ชุดทดลองที่ 3 ค้นหาได้ 4 ภาพ ,ชุดทดลองที่ 4 ค้นหาได้ 4 ภาพ ,ชุดทดลองที่ 5 ค้นหาได้ 4 ภาพ โดยขาดภาพที่เกิดจากการย่อขนาดลงเหลือ 10 % เท่านั้น

ผลการทดลองเมื่อค้นหาภาพด้วยค่า differential=2




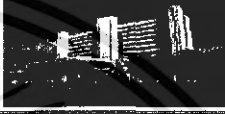

SEARCH RESULT:

Your Picture	
Source	Picture
http://img61.imageshack.us/img61/834/dv1cm4zj.jpg	
http://img239.imageshack.us/img239/8683/cv1c5wz.jpg	
http://img141.imageshack.us/img141/8350/cv1cs3qz.jpg	
http://img140.imageshack.us/img140/972/cv1c905kw.jpg	

รูปที่ 4.6 ภาพชุดทดลองที่ 1




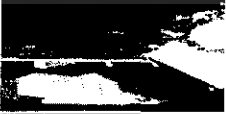

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img161.imageshack.us/img161/1818/e127lg.jpg	
http://img56.imageshack.us/img56/6455/e12m2xw.jpg	
http://img63.imageshack.us/img63/7783/e12s1rf.jpg	
http://img61.imageshack.us/img61/6358/e12904kj.jpg	

รูปที่ 4.7 ภาพชุดทดลองที่ 2


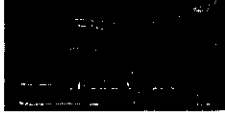


SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img156.imageshack.us/img156/5344/house4ox.jpg	
http://img62.imageshack.us/img62/8010/housem6fq.jpg	
http://img156.imageshack.us/img156/6458/houses6ly.jpg	
http://img100.imageshack.us/img100/7048/house907ll.jpg	

รูปที่ 4.8 ภาพชุดทดลองที่ 3



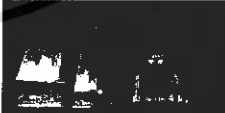

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img109.imageshack.us/img109/869/flag7xz.jpg	
http://img136.imageshack.us/img136/9581/flagm5zy.jpg	
http://img105.imageshack.us/img105/5151/flags2ax.jpg	
http://img66.imageshack.us/img66/6466/flag909vj.jpg	

รูปที่ 4.9 ภาพชุดทดลองที่ 4

SEARCH RESULT:

Your Picture	
Source	Picture
http://img86.imageshack.us/img86/9505/demo4mn.jpg	
http://img88.imageshack.us/img88/4094/demom8jh.jpg	
http://img95.imageshack.us/img95/4652/demos0ug.jpg	
http://img148.imageshack.us/img148/4193/demo909dm.jpg	




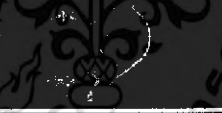
รูปที่ 4.10 ภาพชุดทดลองที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยค่า differential = 2 จะเห็นได้ว่าสามารถหาภาพในแต่ละชุดทดลอง ที่มีจำนวนทั้งหมด 5 ค่า โดยที่ ในชุดทดลองที่ 1 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 2 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 3 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 4 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 5 ค้นหาได้ 4 ภาพ โดยขนาดภาพที่เกิดจากการย่อขนาดลงเหลือ 10 % เท่านั้น

ผลการทดลองเมื่อค้นหาภาพด้วยค่า differential = 3




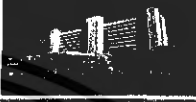

SEARCH RESULT:

Your Picture	
Source	Picture
http://img61.imageshack.us/img61/834/dvicm4zj.jpg	
http://img239.imageshack.us/img239/8683/dvic5wz.jpg	
http://img141.imageshack.us/img141/8350/dvics3qz.jpg	
http://img140.imageshack.us/img140/972/dvic905kw.jpg	

รูปที่ 4.11 ภาพชุดทดลองที่ 1






เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img161.imageshack.us/img161/1818/e127lg.jpg	
http://img56.imageshack.us/img56/6455/e12m2xw.jpg	
http://img63.imageshack.us/img63/7783/e12s1rf.jpg	
http://img164.imageshack.us/img164/921/e12107kn.jpg	
http://img61.imageshack.us/img61/6358/e12904kj.jpg	

รูปที่ 4.12 ภาพชุดทดลองที่ 2

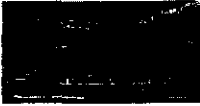
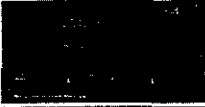

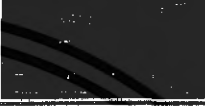
SEARCH RESULT:

Your Picture	
Source	Picture
http://img156.imageshack.us/img156/5344/house4ox.jpg	
http://img62.imageshack.us/img62/8010/housem6iq.jpg	
http://img156.imageshack.us/img156/6458/houses6ly.jpg	
http://img101.imageshack.us/img101/8025/house103vo.jpg	
http://img100.imageshack.us/img100/7048/house907li.jpg	

รูปที่ 4.13 ภาพชุดทดลองที่ 3

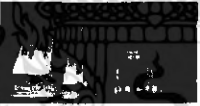
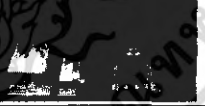
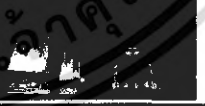
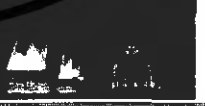

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img109.imageshack.us/img109/869/flag7xz.jpg	
http://img136.imageshack.us/img136/9581/flagm5zy.jpg	
http://img105.imageshack.us/img105/5151/flags2ax.jpg	
http://img66.imageshack.us/img66/6466/flag909vj.jpg	

รูปที่ 4.14 ภาพชุดทดลองที่ 4

SEARCH RESULT:

Your Picture	
Source	Picture
http://img86.imageshack.us/img86/9505/demo4mnh.jpg	
http://img88.imageshack.us/img88/4094/demom6jh.jpg	
http://img95.imageshack.us/img95/4652/demos0ug.jpg	
http://img146.imageshack.us/img146/3659/demo108yg.jpg	
http://img148.imageshack.us/img148/4193/demo909dm.jpg	






รูปที่ 4.15 ภาพชุดทดลองที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยค่า differential = 3 จะเห็นได้ว่าสามารถหาภาพในแต่ละชุดทดลอง ที่มีจำนวนทั้งหมด 5 ค่า โดยที่ ในชุดทดลองที่ 1 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 2 ค้นหาได้ 5 ภาพ, ชุดทดลองที่ 3 ค้นหาได้ 5 ภาพ, ชุดทดลองที่ 4 ค้นหาได้ 4 ภาพ, ชุดทดลองที่ 5 ค้นหาได้ 5 ภาพ โดยขาดภาพที่เกิดจากการย่อขนาดลงเหลือ 10 % เท่านั้น

ผลการทดลองเมื่อค้นหาภาพด้วยค่า differential = 4







SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img61.imageshack.us/img61/834/cv1cm4zj.jpg	
http://img239.imageshack.us/img239/8683/cv1c5wz.jpg	
http://img141.imageshack.us/img141/8350/cv1c3qz.jpg	
http://img140.imageshack.us/img140/972/cv1c905kw.jpg	

รูปที่ 4.16 ภาพชุดทดลองที่ 1







เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img161.imageshack.us/img161/1818/e127lg.jpg	
http://img56.imageshack.us/img56/6455/e12m2xw.jpg	
http://img63.imageshack.us/img63/7783/e12s1rf.jpg	
http://img164.imageshack.us/img164/921/e12107kn.jpg	
http://img61.imageshack.us/img61/6358/e12904kj.jpg	

รูปที่ 4.17 ภาพชุดทดลองที่ 2





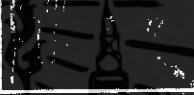
SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img156.imageshack.us/img156/5344/house4ox.jpg	
http://img62.imageshack.us/img62/8010/housem6iq.jpg	
http://img156.imageshack.us/img156/6458/houses6ly.jpg	
http://img101.imageshack.us/img101/8025/house103vo.jpg	
http://img100.imageshack.us/img100/7048/house907ll.jpg	

รูปที่ 4.18 ภาพชุดทดลองที่ 3

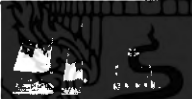
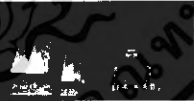

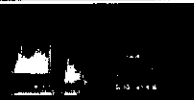

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
Source	Picture
http://img109.imageshack.us/img109/869/flag7xz.jpg	
http://img136.imageshack.us/img136/9581/flagm5zy.jpg	
http://img105.imageshack.us/img105/5151/flags2ax.jpg	
http://img138.imageshack.us/img138/4408/flag100ql.jpg	
http://img66.imageshack.us/img66/6466/flag909vj.jpg	

รูปที่ 4.19 ภาพชุดทดลองที่ 4

SEARCH RESULT:

Your Picture	
Source	Picture
http://img86.imageshack.us/img86/9505/demo4mn.jpg	
http://img88.imageshack.us/img88/4094/demom8jh.jpg	
http://img95.imageshack.us/img95/4652/demos0ug.jpg	
http://img146.imageshack.us/img146/3659/demo108yg.jpg	
http://img148.imageshack.us/img148/4193/demo909dm.jpg	






รูปที่ 4.20 ภาพชุดทดลองที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยค่า differential = 4 จะเห็นได้ว่าสามารถหาภาพในแต่ละชุดทดลอง ที่มีจำนวนทั้งหมด 5 ค่า โดยที่ ในชุดทดลองที่ 1 ค้นหาได้ 4 ภาพ ,ชุดทดลองที่ 2 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 3 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 4 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 5 ค้นหาได้ 5 ภาพ โดยขนาดภาพที่เกิดจากการย่อขนาดลงเหลือ 10 % เท่านั้น

ผลการทดลองเมื่อค้นหาภาพด้วยค่า differential = 5





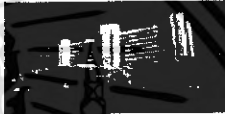

SEARCH RESULT:

Your Picture	
Source	Picture
http://img61.imageshack.us/img61/834/dvicm4zj.jpg	
http://img239.imageshack.us/img239/8683/dvic5wz.jpg	
http://img141.imageshack.us/img141/8350/dvics3qz.jpg	
http://img99.imageshack.us/img99/1648/dvic104vz.jpg	
http://img140.imageshack.us/img140/972/dvic905kw.jpg	

รูปที่ 4.21 ภาพชุดทดลองที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



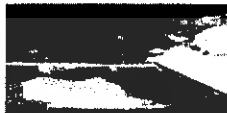



SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img161.imageshack.us/img161/1818/e127lg.jpg	
http://img56.imageshack.us/img56/6455/e12m2xw.jpg	
http://img63.imageshack.us/img63/7783/e12s1rf.jpg	
http://img164.imageshack.us/img164/921/e12107kn.jpg	
http://img61.imageshack.us/img61/6358/e12904kj.jpg	

รูปที่ 4.22 ภาพชุดทดลองที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




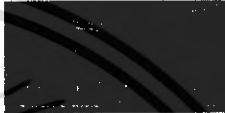
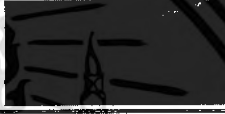

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img156.imageshack.us/img156/5344/house4ox.jpg	
http://img62.imageshack.us/img62/8010/housem6iq.jpg	
http://img156.imageshack.us/img156/6458/houses6iy.jpg	
http://img101.imageshack.us/img101/8025/house103vo.jpg	
http://img100.imageshack.us/img100/7048/house907ii.jpg	

รูปที่ 4.23 ภาพชุดทดลองที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้







SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img109.imageshack.us/img109/869/flag7xz.jpg	
http://img136.imageshack.us/img136/9581/flagm5zy.jpg	
http://img105.imageshack.us/img105/5151/flags2ax.jpg	
http://img138.imageshack.us/img138/4408/flag100qj.jpg	
http://img66.imageshack.us/img66/6466/flag909vj.jpg	

รูปที่ 4.24 ภาพชุดทดลองที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://img86.imageshack.us/img86/9505/demo4mn.jpg	
http://img88.imageshack.us/img88/4094/demom8jh.jpg	
http://img95.imageshack.us/img95/4652/demos0ug.jpg	
http://img146.imageshack.us/img146/3659/demo108yg.jpg	
http://img148.imageshack.us/img148/4193/demo909dm.jpg	

รูปที่ 4.25 ภาพชุดทดลองที่ 5

จากการทดลองด้วยค่า differential = 5 จะเห็นได้ว่าสามารถหาภาพในแต่ละชุดทดลอง ที่มีจำนวนทั้งหมด 5 ค่า โดยที่ ในชุดทดลองที่ 1 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 2 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 3 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 4 ค้นหาได้ 5 ภาพ ,ชุดทดลองที่ 5 ค้นหาได้ 5 ภาพ

ดังนั้นจึงสรุปได้ว่า ต้องใช้ ค่า differential = 5 เพื่อใช้ในการค้นหาภาพ เพราะจากการทดลองจะเห็นได้ว่า เมื่อใช้ค่านี้ในการค้นหา จะทำให้ค้นพบภาพจากชุดทดลองได้ครบที่สุด ดังตารางที่ 4.2

ค่า differential ที่ใช้	จำนวนภาพที่หาได้จากภาพชุดทดลองที่				
	1	2	3	4	5
1	3	4	4	4	4
2	4	4	4	4	4
3	4	5	5	4	5
4	4	5	5	5	5
5	5	5	5	5	5


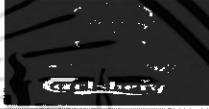


ตารางที่ 4.2 ภาพที่หาได้จากชุดทดลองต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การทดลองหาภาพตัวอย่าง





ในการทดลองนี้จะเป็นการค้นหารูปภาพโดยใช้รูปที่สุ่มขึ้นมาจำนวน 6 ภาพในการหาแล้วได้ผลลัพธ์ภาพซึ่งมีทั้งภาพที่เหมือนกับภาพที่ใช้ในการค้นหา และภาพที่มีความคล้ายกันในเรื่องของลักษณะสีของภาพซึ่งอาจจะเป็นภาพที่ไม่ต้องการติดมาด้วย โดยในขณะที่ทำการทดลองมีภาพอยู่ในฐานข้อมูลประมาณ 14,000 รูป

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://www.liverpoolfc.tv/team/squad/gertard/images/profile.jpg	
http://www.liverpoolfc.tv/team/squad/carragher/images/profile.jpg	
http://www.liverpoolfc.tv/team/squad/warnock/images/profile.jpg	

รูปที่ 4.26 ผลการค้นหาภาพที่สุ่มมาใช้ค้นหาที่ 1





SEARCH RESULT:

Your Picture	
	
Source	Picture
http://football.sanook.com/story_picture/s/09767_001.jpg	
http://football.sanook.com/story_picture/s/09728_001.jpg	
http://sport.sanook.com/story_picture/s/09767_001.jpg	

รูปที่ 4.27 ผลการค้นหาแสดงภาพที่สุ่มมาใช้ค้นหาที่ 2


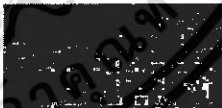


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://football.sanook.com/story_picture/s/09753_001.jpg	
http://sport.sanook.com/story_picture/s/09753_001.jpg	
http://sports.mweb.co.th/story_picture/s/09753_001.jpg	

รูปที่ 4.28 ผลการค้นหภาพที่สุ่มมาใช้ค้นหาที่ 3




SEARCH RESULT:

Your Picture	
	
Source	Picture
http://travel.sanook.com/story_picture/s/08610_023.jpg	
http://travel.mweb.co.th/story_picture/s/08610_023.jpg	
http://www.thalmet.tmd.go.th/./image/e_1.jpg	

รูปที่ 4.29 ผลการค้นหภาพที่สุ่มมาใช้ค้นหาที่ 4

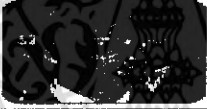




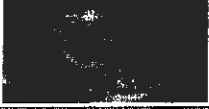
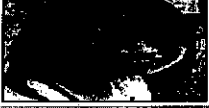

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://www.awd.ws/photos/thumbs/1089-01.jpg	
http://www.kmitl.ac.th/images/photo/11.jpg	

รูปที่ 4.30 ผลการค้นหภาพที่สุ่มมาใช้ค้นหาที่ 5

SEARCH RESULT:

Your Picture	
	
Source	Picture
http://www.sanook.com/story_picture/s/03809_001.jpg	
http://women.sanook.com/story_picture/s/16971_001.jpg	
http://women.sanook.com/story_picture/s/16703_001.jpg	
http://women.sanook.com/story_picture/s/16705_001.jpg	
http://www.mthel.com/scoop/bof/upload/d2PJWtue100746.jpg	
http://qq.sanook.com/qqidol/pare/images/pare_side6.jpg	
http://kids.mweb.co.th/story_picture/s/03809_001.jpg	

รูปที่ 4.31 ผลการค้นหภาพที่สุ่มมาใช้ค้นหาที่ 6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับบริการเชิงวิชาการเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลองด้วยภาพที่ต่างกันไปจะได้ผลลัพธ์ที่ต่างกันไป ซึ่งจะขึ้นอยู่กับข้อมูลภาพที่มีอยู่ในฐานข้อมูลว่ามีมากเพียงใด และค่าสีของภาพนั้นมีภาพที่มีลักษณะสีคล้ายคลึงกันอยู่มากเพียงใด ดังเช่นภาพที่สุ่มมาใช้ค้นหาที่ 1 จะได้ผลลัพธ์เป็นภาพที่เหมือนกัน 100% ออกมาและพบภาพของนักกีฬาซึ่งมีลักษณะสีผิวและสีหมวกคล้ายกัน รวมถึงชุดที่ใส่เป็นสีเดียวกัน และพื้นหลังของภาพเหมือนกัน ออกมาด้วย ภาพที่สุ่มมาใช้ค้นหาที่ 2 จะได้ผลลัพธ์เป็นภาพที่เหมือนกัน 100% ทุกภาพ แต่เป็นภาพที่อยู่ในตำแหน่งที่ต่างกันของเว็บไซต์ ภาพที่สุ่มมาใช้ค้นหาที่ 3 จะได้ผลลัพธ์เป็นภาพที่เหมือนกัน 100% ทุกภาพ แต่เป็นภาพที่อยู่ในตำแหน่งที่ต่างกันของเว็บไซต์เช่นกัน ภาพที่สุ่มมาใช้ค้นหาที่ 4 จะได้ผลลัพธ์เป็นภาพที่เหมือนกัน 100% ทุกภาพ แต่เป็นภาพที่อยู่ในตำแหน่งที่ต่างกันของเว็บไซต์ และได้ภาพที่มีลักษณะสีใกล้เคียงกันมาด้วย ภาพที่สุ่มมาใช้ค้นหาที่ 5 จะได้ผลลัพธ์เป็นภาพที่เหมือนกัน 100% ทุกภาพ แต่เป็นภาพที่อยู่ในตำแหน่งที่ต่างกันของเว็บไซต์ และได้ภาพที่มีลักษณะสีใกล้เคียงกันมาด้วยเช่นกัน ส่วนภาพที่สุ่มมาใช้ค้นหาที่ 6 จะเป็นภาพใบหน้าคนซึ่งภาพที่มีลักษณะสีลักษณะนี้ที่ใกล้เคียงกันมีจำนวนหลายภาพ จึงได้ผลลัพธ์เป็นภาพใบหน้าคนที่เหมือนกัน 100% และภาพใบหน้าคนต่างๆออกมาด้วย

บทที่ 5

บทวิจารณ์และสรุป

5.1 บทสรุปและวิจารณ์ผลการทดลอง

ในการเก็บข้อมูลของรูปภาพต่างๆมาจากอินเทอร์เน็ตนั้น โปรแกรม Web Crawler ที่ใช้สามารถทำการค้นหาข้อมูลลิงค์ และข้อมูลรูปภาพจากหน้าเว็บต่างๆได้ในปริมาณที่น่าพอใจ แต่ไม่สามารถเก็บข้อมูลมาได้ครบทุกๆลิงค์และรูปภาพ เนื่องจากบางลิงค์และรูปภavnั้นเสียอยู่ จึงไม่อาจเข้าถึงข้อมูลได้

ในการทดลองเก็บข้อมูลลิงค์และรูปภาพเริ่มต้นที่เว็บ <http://www.sanook.com> โดยเฉลี่ยในแต่ละโปรเซสได้ผลการทดลองดังนี้คือ โปรเซสค้นหาลิงค์มีความสามารถในการค้นหาลิงค์ด้วยความเร็วประมาณ 413 ลิงค์/นาที โปรเซสค้นหาลิงค์ของรูปภาพมีความสามารถในการค้นหาลิงค์ของรูปภาพด้วยความเร็วประมาณ 87 ลิงค์/นาที โปรเซสในการทำฮิสโตแกรมมีความสามารถในการทำด้วยความเร็วประมาณ 44 ภาพ/นาที ซึ่งเป็นจำนวนที่มากพอสมควรแต่ยังคงไม่เพียงพอเนื่องจากรูปในอินเทอร์เน็ต มีปริมาณมากกว่านี้มาก

โปรแกรมค้นหารูปภาพโดยเสิร์จเอนจินนี้ เป็นสิ่งที่สามารถตอบสนองความต้องการค้นหาข้อมูลของผู้คนได้ในระดับหนึ่ง โดยในการค้นหารูปภาพนั้นผู้ใช้สามารถนำรูปภาพที่ต้องการมาค้นหาในฐานะข้อมูลของเราได้จากการเทียบค่าฮิสโตแกรมและแสดงผลออกมาทางหน้าจอ โดยที่รูปภาพในฐานะข้อมูลที่ดึงมาจากเว็บต่างๆบางภาพที่เป็นรูปภาพเดียวกันแต่อยู่คนละเว็บไซต์ก็จะมีค่าฮิสโตแกรมที่เหมือนกันและค่าฮิสโตแกรมในแต่ละรูปภาพที่ไม่ใช่รูปเดียวกัน โอกาสที่จะซ้ำกันหรือคล้ายกันน้อย ทำให้เราสามารถค้นหารูปภาพที่มีความเหมือนกันทุกประการหรือความคล้ายได้

การหารูปภาพในการทดลองผลลัพธ์ที่ได้จะต้องเป็นรูปที่มีความเหมือนกันทุกประการหรือเป็นภาพเดียวกันแต่มีขนาดต่างกัน แต่ก็อาจจะเกิดกรณีที่ได้ภาพที่ไม่เหมือนกันเลย แต่เนื่องจากลักษณะเฉพาะเหมือนกัน เช่น ภาพที่มีโทนสีโดยรวมใกล้เคียงกัน ดังนั้นในการออกแบบการทดลองต่อไปเราอาจจะค้นหารูปภาพที่มีความคล้ายคลึงโดยมีเป็นรูปที่มีความหมายใกล้เคียงกัน ซึ่งผลลัพธ์ที่ออกมาอาจจะมีความถูกต้องมากขึ้น แต่ผลลัพธ์ที่ดีที่สุดคือได้ผลลัพธ์ที่มีรูปภาพที่ไม่ตรงกับความต้องการเจือปนให้น้อยที่สุด

5.2 ปัญหาอุปสรรคและแนวทางในการแก้ไข

1. สามารถหารูปภาพได้เพียงแค่ว่าจากลักษณะเฉพาะแบบฮิสโตแกรมเท่านั้นซึ่งอาจจะทำให้ได้ภาพที่มีความใกล้เคียงกันในลักษณะสีของภาพ แต่ได้ภาพที่ไม่ตรงกับความต้องการ ซึ่งแนวทางในการแก้ปัญหาที่นี้ คือ การปรับปรุงอัลกอริทึมที่ใช้ให้ดีขึ้นต่อไป
2. เนื่องจากเครื่องเซิร์ฟเวอร์ที่ใช้มีประสิทธิภาพไม่เพียงพอ และรูปภาพในเครือข่ายอินเทอร์เน็ตมีจำนวนมากมหาศาล จึงทำให้การทำงานของโปรแกรม Web crawler ไม่สามารถรวบรวมรูปภาพมาได้เพียงพอต่อการค้นหา ซึ่งแนวทางในการแก้ปัญหาที่นี้ คือ การเพิ่มเครื่องเซิร์ฟเวอร์ให้มีหลายๆเครื่องเพื่อช่วยกันทำงาน
3. ปัญหาที่เกิดขึ้นในกรณีที่เกิดไฟดับหรือระบบถูกปิดลงจะทำให้การทำงานของโปรเซสต่างๆที่ทำอยู่หยุดการทำงานไป ทำให้เมื่อเริ่มทำงานใหม่จะข้ามโปรเซสที่ยังทำไม่เสร็จในส่วนที่อยู่ในการทำงานตอนที่เกิดไฟดับไป ทำให้อาจมีบางข้อมูลได้ข้อมูลไม่ครบ ซึ่งสามารถแก้ไขด้วยการปรับปรุงซอฟต์แวร์ให้ดีขึ้น โดยอาจจะมีการ logfile เพื่อบันทึกการทำงาน ณ จุดที่เกิดปัญหานั้นแล้วเมื่อเริ่มโปรแกรมขึ้นมาใหม่ก็ให้ไปอ่านในค่า logfile ว่าทำถึง ณ จุดไหนแล้วจึงทำต่อจากจุดนั้น
4. ในการทำงานของโปรแกรม ยังมีปัญหาในส่วนของ CPU usage ซึ่งยังใช้งานถึง 100 % ทำให้โปรเซสแต่ละตัวทำงานได้ช้าลง ทำให้ดึงข้อมูลรูปภาพ และลิงค์ได้น้อยลง สามารถแก้ไขด้วยการปรับปรุงซอฟต์แวร์ให้ดีขึ้น โดยปรับปรุงในส่วนของการเก็บ URL link ให้มีการแบ่งโปรเซสการทำงานโดยให้แต่ละโปรเซสแบ่งหน้าที่รับผิดชอบ URL link ของตนเอง เพราะการทำงานในส่วนของ URL link โปรเซสเดียวยังไม่ดีพอทำให้ CPU usage มีการใช้งานถึง 100 %
5. ในการทำงานของโปรแกรมเพื่อค้นหาข้อมูลรูปภาพและลิงค์ ในตอนเริ่มต้นยังคงต้องเริ่มจากการใส่เว็บใน textfile ยังไม่สามารถเริ่มต้นได้จากการ generate เว็บขึ้นมาเอง ทำให้เมื่อถึงจุดๆหนึ่งที่เว็บใน textfile ไม่สามารถค้นหาลิงค์และรูปภาพได้แล้วจะทำให้ข้อมูลไม่ขยายตัวต่อไปและยังคงได้แค่ข้อมูลที่ซ้ำกับที่มีอยู่ในฐานข้อมูล สามารถแก้ไขด้วยการปรับปรุงซอฟต์แวร์ให้ดีขึ้น โดยพัฒนาโปรแกรมในส่วนที่ทำให้ตัวโปรแกรมเองสามารถ generate เว็บขึ้นมาเองได้แทนการใส่ค่า URL จาก textfile เพื่อให้ได้ URL ที่ไม่จำกัดอยู่แค่ใน text file สามารถขยายข้อมูลออกไปได้อีก

5.3 แนวทางการพัฒนาต่อ

1. พัฒนาในส่วนที่ติดต่อกับผู้ใช้งานให้มีความสวยงามมากขึ้นและให้ใช้งานได้ง่ายมากขึ้น
2. พัฒนาอัลกอริทึมที่ใช้ในการค้นหารูปภาพให้มีประสิทธิภาพมากขึ้น เพื่อให้สามารถค้นหารูปภาพได้ตรงกับความต้องการของผู้ใช้มากที่สุด
3. พัฒนาส่วนของ crawler ให้สามารถ generate เว็บขึ้นมาเพื่อค้นหารูปภาพและ link ได้
4. พัฒนาในส่วนของกรณีไฟดับแล้ว โพรเซสยังทำงานค้างอยู่ พัฒนาให้โปรแกรมสามารถทำต่อ ณ จุดที่เกิดไฟดับนั้น
6. พัฒนาในส่วนของการทำงานของโปรแกรมให้มีการจำกัดการใช้งาน CPU usage ไม่ให้มีการใช้งานถึง 100% ซึ่งจะให้ทำโปรเซสในการดึงรูปภาพและลิงค์ซ้ำลง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

- [1] A.M. Kuchling , “Regular Expression HOWTO”, URL: <http://www.amk.ca/python/howto>
- [2] “Dive Into Python”, URL: <http://diveintopython.org/>
- [3] “Python programming language”, URL: <http://www.python.org>
- [4] “MYSQL”, URL: <http://www.mysql.com>
- [5] สุทธิพล วรารุทธิ์ “การออกแบบและพัฒนาระบบเสิร์ชเอนจินขนาดใหญ่” บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ เพื่อความสมบูรณ์แห่งปัญญา วิศวกรรมศาสตรมหาบัณฑิต (คอมพิวเตอร์) พ.ศ. 2544
- [6] ชีร์ ไรจงวัฒน์ และ ชีระวุฒิ กิ่งไพบูลย์ “การศึกษาและเปรียบเทียบประสิทธิภาพในการค้นหาข้อมูลรูปภาพ” วิทยานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชา วิศวกรรมคอมพิวเตอร์ ปีการศึกษา 2547 บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
- [7] Swaroop C H , 2004, “A Byte of Python”
- [8] Guido van Rossum and Fred L. Drake, Jr., “Python Tutorial”



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก ก.

ตัวอย่างตาราง table_image

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

ตัวอย่างตาราง table_image

id	urlimage	histogram	date
1	http://www.sanook.com/story_picture/s/03798_001.jpg	[8, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
2	http://www.sanook.com/story_picture/s/03807_001.jpg	[27, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
3	http://www.sanook.com/story_picture/s/03809_001.jpg	[21, 0, 0, 0, 0, 0, 0, 0..	2006-03-16
4	http://www.sanook.com/story_picture/s/01821_001.jpg	[7, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
5	http://www.sanook.com/story_picture/s/01766_001.jpg	[1, 0, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
6	http://www.sanook.com/story_picture/s/01794_001.jpg	[1, 0, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
7	http://www.sanook.com/story_picture/s/01820_001.jpg	[19, 0, 0, 0, 0, 0, 0, 0, 1,..	2006-03-16
8	http://www.sanook.com/story_picture/s/01822_001.jpg	[5, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
9	http://www.sanook.com/story_picture/s/00735_001.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
10	http://www.sanook.com/story_picture/s/03101_001.jpg	[19, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
11	http://www.sanook.com/editimages/pict01.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
12	http://www.sanook.com/editimages/pict02.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
13	http://www.sanook.com/story_picture/s/03671_001.jpg	[17, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
14	http://www.sanook.com/story_picture/s/03670_002.jpg	[7, 0, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
15	http://www.sanook.com/story_picture/s/03668_001.jpg	[37, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
16	http://www.sanook.com/story_picture/s/03667_001.jpg	[26, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
17	http://www.sanook.com/story_picture/s/03669_001.jpg	[10, 0, 0, 0, 0, 0, 0, 0,..	2006-03-16
18	http://uploaded.auction.sanook.com/uploaded/_rad270360eacee6e7-0176-4950-b539-d84f5f491ab4.jpg	[1, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
19	http://uploaded.auction.sanook.com/uploaded/_rad1682244a2ffa24-6016-48ed-9585-a07d7448b620.jpg	[12, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
20	http://uploaded.auction.sanook.com/uploaded/_rad16016011b7e09f-4342-4d60-9c3c-2e82f847cfb6.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
21	http://uploaded.auction.sanook.com/uploaded/_rad2423222214fc02-792a-4537-a262-42cccfaad796.jpg	[2, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
22	http://uploaded.auction.sanook.com/uploaded/_rad3392548f878222-c311-4a87-845f-ca6d5da36445.jpg	[4, 3, 0, 0, 0, 0, 0, 0,..	2006-03-16
23	http://uploaded.auction.sanook.com/uploaded/_rad2491454ccda102-acf1-4eac-9145-0f3c26e4833d.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
24	http://uploaded.auction.sanook.com/uploaded/_rad3094508a0e7f46-bd0c-4459-abc3-6af285188b61.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
25	http://uploaded.auction.sanook.com/uploaded/_rad24232212371ea5-4798-4887-95ff-ffc0618fd9bb.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
26	http://tv.sanook.com/http://radio.sanook.com/cover/14023.jpg'	NULL	0000-00-00
27	http://blog.sanook.com/PortalPics/_Blogspace/images/default/icon002.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
28	http://blog.sanook.com/PortalPics/_Blogspace/images/default/icon001.jpg	[0, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
29	http://women.sanook.com/story_picture/m/16431_001.jpg	[8, 0, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
30	http://women.sanook.com/story_picture/m/16318_001.jpg	[17, 0, 0, 0, 0, 0, 0, 0,..	2006-03-
31	http://women.sanook.com/story_picture/m/16151_001.jpg	[24, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
32	http://women.sanook.com/story_picture/m/15993_001.jpg	[22, 0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
33	http://webindex.sanook.com/images/web_highlight.jpg	[12, 0, 0, 0, 0, 0, 0, 1, ..	2006-03-16

34	http://qq.sanook.com/story_picture/s/01067_001.jpg	[7, 0, 0, 0, 0, 0, 0, ...	2006-03-16
35	http://qq.sanook.com/story_picture/s/01069_001.jpg	[9, 8, 0, 0, 0, 0, 0, ...	2006-03-16
36	http://qq.sanook.com/story_picture/s/01064_001.jpg	[7, 0, 0, 0, 0, 0, 0, ...	2006-03-16
37	http://qq.sanook.com/story_picture/s/01063_001.jpg	[17, 0, 0, 0, 0, 0, 0, ...	2006-03-16
38	http://qq.sanook.com/story_picture/s/01051_001.jpg	[7, 0, 0, 0, 0, 0, 0, ...	2006-03-16
39	http://qq.sanook.com/php/my_pic/3429112/my_1.jpg	[6, 0, 0, 0, 0, 0, 0, ...	2006-03-16
40	http://qq.sanook.com/php/my_pic/7008570/my_1.jpg	[9, 6, 0, 0, 0, 0, 0, ...	2006-03-16
41	http://qq.sanook.com/php/my_pic/7028442/my_1.jpg	[8, 0, 0, 0, 0, 0, 0, ...	2006-03-16
42	http://qq.sanook.com/php/my_pic/7013839/my_1.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
43	http://qq.sanook.com/php/my_pic/2791154/my_1.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
44	http://qq.sanook.com/php/my_pic/7092720/my_1.jpg	[6, 1, 0, 0, 0, 0, 0, ...	2006-03-16
45	http://qq.sanook.com/story_picture/s/00977_002.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
46	http://qq.sanook.com/story_picture/s/01066_001.jpg	[7, 0, 0, 0, 0, 0, 0, ...	2006-03-16
47	http://qq.sanook.com/story_picture/s/01068_001.jpg	[9, 8, 0, 0, 0, 0, 0, ...	2006-03-16
48	http://qq.sanook.com/story_picture/s/01062_001.jpg	[17, 0, 0, 0, 0, 0, 0, ...	2006-03-16
49	http://friend.sanook.com/060217_blinddate.jpg	[40, 0, 0, 0, 0, 0, 0, ...	2006-03-16
50	http://world.sanook.com/web/homehighlight16.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
51	http://world.sanook.com/web/intro20051215.jpg	[2, 0, 0, 0, 0, 0, 0, ...	2006-03-16
52	http://radio.sanook.com//cover/11362.jpg'	NULL	0000-00-00
53	http://football.sanook.com/story_picture/s/09768_001.jpg	[12, 0, 0, 0, 0, 0, 0, ...	2006-03-16
54	http://football.sanook.com/story_picture/s/09767_001.jpg	[5, 0, 0, 0, 0, 0, 0, ...	2006-03-16
55	http://football.sanook.com/story_picture/s/09766_001.jpg	[3, 0, 0, 0, 0, 0, 0, ...	2006-03-16
56	http://football.sanook.com/story_picture/s/09743_001.jpg	[22, 4, 1, 0, 0, 0, 0, ...	2006-03-16
57	http://football.sanook.com/story_picture/s/09753_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
58	http://football.sanook.com/story_picture/s/09752_001.jpg	[7, 0, 0, 0, 0, 0, 0, ...	2006-03-16
59	http://voice.sanook.com/content/p01.jpg	[9, 1, 0, 0, 0, 0, 0, ...	2006-03-16
60	http://voice.sanook.com/content/taro.jpg	[0, 0, 0, 0, 0, 0, 0, ...	
61	http://voice.sanook.com/thm_foreign/img_main_bottom.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
62	http://game.sanook.com/images/th_chess.jpg	NULL	0000-00-00
63	http://game.sanook.com/images/th_defen_chess.jpg	NULL	0000-00-00
64	http://game.sanook.com/images/th_checkers.jpg	NULL	0000-00-00
65	http://game.sanook.com/images/chess.jpg	NULL	0000-00-00
66	http://game.sanook.com/images/chinese_chess.jpg	NULL	0000-00-00
67	http://game.sanook.com/images/shogi.jpg	NULL	0000-00-00
68	http://game.sanook.com/images/go.jpg	NULL	0000-00-00
69	http://game.sanook.com/images/othello.jpg	NULL	0000-00-00
70	http://game.sanook.com/images/crossword.jpg	NULL	0000-00-00
71	http://game.sanook.com/images/amath.jpg	NULL	0000-00-00
72	http://game.sanook.com/story_picture/s/02085_001.jpg	[15, 3, 0, 0, 0, 0, 0, ...	2006-03-16
73	http://game.sanook.com/story_picture/s/02074_001.jpg	[16, 0, 0, 0, 0, 0, 0, ...	2006-03-16
74	http://game.sanook.com/story_picture/s/02073_001.jpg	[2, 0, 0, 0, 0, 0, 0, ...	2006-03-16
75	http://game.sanook.com/story_picture/s/02091_001.jpg	[2, 0, 0, 0, 0, 0, 0, ...	2006-03-16
76	http://game.sanook.com/story_picture/s/02086_001.jpg	[9, 0, 0, 0, 0, 0, 0, ...	2006-03-16
77	http://game.sanook.com/story_picture/s/02078_001.jpg	[11, 1, 0, 0, 0, 0, 0, ...	2006-03-16
78	http://game.sanook.com/story_picture/s/02077_001.jpg	[11, 2, 0, 0, 0, 0, 0, ...	2006-03-16
79	http://game.sanook.com/story_picture/s/02072_001.jpg	[9, 1, 0, 0, 0, 0, 0, ...	2006-03-16

80	http://game.sanook.com/story_picture/s/02075_001.jpg	[18, 14, 0, 0, 0, ..	2006-03-16
81	http://game.sanook.com/story_picture/s/02064_001.jpg	[6, 0, 0, 0, 0, 0, ..	2006-03-16
82	http://game.sanook.com/story_picture/s/02062_001.jpg	[3, 0, 0, 0, 0, 0, ..	2006-03-16
83	http://game.sanook.com/story_picture/s/02102_001.jpg	[6, 0, 0, 0, 0, 0, ..	2006-03-16
84	http://game.sanook.com/story_picture/s/02090_001.jpg	[8, 0, 0, 0, 0, 0, ..	2006-03-16
85	http://game.sanook.com/story_picture/s/02084_001.jpg	[1, 0, 0, 0, 0, 0, ..	2006-03-16
86	http://game.sanook.com/story_picture/s/02076_001.jpg	[25, 0, 0, 0, 0, 0, ..	2006-03-16
87	http://game.sanook.com/story_picture/m/02115_001.jpg	[51, 1, 0, 0, 0, 0, ..	2006-03-16
88	http://game.sanook.com/story_picture/m/02110_001.jpg	[4, 1, 6, 0, 0, 0, 3, ..	2006-03-16
89	http://game.sanook.com/story_picture/m/02109_001.jpg	[1, 0, 0, 0, 0, 0, ..	2006-03-16
90	http://world.sanook.com/web/club_200509161711_idolfocushighlight.jpg	[11, 0, 0, 0, 0, 0, ..	2006-03-16
91	http://world.sanook.com/web/060310_1_385701.jpg	[2, 1, 0, 0, 0, 0, ..	2006-03-16
92	http://world.sanook.com/web/060310_2_385166.jpg	[0, 0, 0, 0, 0, 0, ..	2006-03-16
93	http://world.sanook.com/web/060310_3_381651.jpg	[6, 6, 0, 0, 0, 0, ..	2006-03-16
94	http://world.sanook.com/web/060310_4_380994.jpg	[3, 0, 0, 0, 0, 0, ..	2006-03-16
95	http://club.sanook.com/pictures/sm/5/sm0038841001.jpg	[5, 0, 0, 0, 0, 0, ..	2006-03-16
96	http://club.sanook.com/pictures/tn/3/tn0003710201.jpg	[31, 0, 0, 0, 0, 0, ..	2006-03-16
97	http://club.sanook.com/pictures/tn/3/tn0001611101.jpg	[2, 4, 2, 0, 0, 0, ..	2006-03-16
98	http://ecard.sanook.com/ecard_card/tn/tn_001123.jpg	[0, 0, 0, 0, 0, 0, ..	2006-03-16
99	http://world.sanook.com/web/ecard_20060207_1.jpg	[3, 0, 0, 0, 0, 0, ..	2006-03-16
100	http://win.sanook.com/story_picture/s/00383_001.jpg	[7, 0, 0, 0, 0, 0, ..	2006-03-16
101	http://win.sanook.com/story_picture/s/00378_001.jpg	[30, 1, 0, 0, 0, 0, 0, ..	2006-03-16
102	http://win.sanook.com/story_picture/s/00377_001.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
103	http://win.sanook.com/story_picture/s/00376_001.jpg	[5, 0, 0, 0, 0, 0, 0, ..	2006-03-16
104	http://win.sanook.com/story_picture/s/00371_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
105	http://win.sanook.com/story_picture/s/00362_001.jpg	[29, 1, 0, 0, 0, 0, 0, ..	2006-03-16
106	http://win.sanook.com/story_picture/s/00360_001.jpg	[44, 0, 0, 0, 0, 0, 0, ..	2006-03-16
107	http://win.sanook.com/story_picture/s/00350_001.jpg	[4, 3, 9, 6, 0, 0, 0, ..	2006-03-16
108	http://win.sanook.com/story_picture/s/00348_001.jpg	[8, 0, 0, 0, 0, 0, 0, ..	2006-03-16
109	http://win.sanook.com/story_picture/s/00382_001.jpg	[26, 0, 0, 0, 0, 0, 3, ..	2006-03-16
110	http://win.sanook.com/story_picture/s/00380_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
111	http://win.sanook.com/story_picture/s/00241_001.jpg	[34, 0, 0, 0, 0, 0, 0, ..	2006-03-16
112	http://win.sanook.com/story_picture/s/00329_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
113	http://win.sanook.com/story_picture/s/00324_001.jpg	[14, 0, 0, 0, 0, 0, 0, ..	2006-03-16
114	http://win.sanook.com/story_picture/s/00313_001.jpg	[2, 2, 0, 0, 0, 0, 0, ..	2006-03-16
115	http://win.sanook.com/story_picture/s/00312_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
116	http://win.sanook.com/story_picture/s/00290_001.jpg	[0, 2, 0, 0, 0, 0, 0, ..	2006-03-16
117	http://news.sanook.com/news_picture/9490000036419.jpg	[31, 0, 0, 0, 0, 0, 0, ..	2006-03-16
118	http://news.sanook.com/news_picture/9490000036359.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
119	http://talk.sanook.com/story_picture/m/05540_001.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
120	http://talk.sanook.com/story_picture/s/05372_001.jpg	[22, 0, 0, 0, 0, 0, 1, ..	2006-03-16
121	http://talk.sanook.com/story_picture/s/05324_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
122	http://talk.sanook.com/story_picture/s/05300_001.jpg	[9, 0, 0, 0, 0, 0, 0, ..	2006-03-16
123	http://talk.sanook.com/story_picture/s/05297_001.jpg	[12, 0, 0, 0, 0, 0, 0, ..	2006-03-16
124	http://talk.sanook.com/story_picture/s/05294_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16

125	http://talk.sanook.com/story_picture/s/05249_001.jpg	[3, 0, 0, 0, 0, 0, 0, ...	2006-03-16
126	http://talk.sanook.com/story_picture/s/05228_001.jpg	[10, 0, 0, 0, 0, 0, 0, ...	2006-03-16
127	http://talk.sanook.com/story_picture/s/05142_001.jpg	[0, 0, 0, 0, 0, 0, 1, ...	2006-03-16
128	http://talk.sanook.com/story_picture/s/05099_001.jpg	[18, 1, 0, 0, 0, 0, 0, ...	2006-03-16
129	http://talk.sanook.com/story_picture/s/05507_001.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
130	http://talk.sanook.com/story_picture/s/05381_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
131	http://travel.sanook.com/thm_butterfly/img_main_top.jpg	[0, 0, 0, 0, 0, 0, 47, ...	2006-03-16
132	http://travel.sanook.com/story_picture/m/08693_002.jpg	[13, 0, 0, 0, 0, 0, 0, ...	2006-03-16
133	http://travel.sanook.com/story_picture/s/08683_006.jpg	[4, 0, 0, 0, 0, 0, 0, ...	2006-03-16
134	http://travel.sanook.com/story_picture/s/08695_008.jpg	[3, 1, 0, 0, 0, 0, 0, ...	2006-03-16
135	http://travel.sanook.com/story_picture/s/08674_007.jpg	[2, 0, 0, 0, 0, 0, 0, ...	2006-03-16
136	http://travel.sanook.com/story_picture/s/08669_008.jpg	[5, 0, 0, 0, 0, 0, 0, ...	2006-03-16
137	http://travel.sanook.com/story_picture/s/01425_013.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
138	http://travel.sanook.com/story_picture/s/08691_001.jpg	[4, 1, 0, 0, 0, 0, 0, ...	2006-03-16
139	http://travel.sanook.com/story_picture/s/08610_023.jpg	[4, 1, 0, 0, 0, 0, 0, ...	2006-03-16
140	http://travel.sanook.com/story_picture/s/08653_001.jpg	[4, 0, 0, 0, 0, 0, 0, ...	2006-03-16
141	http://travel.sanook.com/images/rayong20060309.jpg	[8, 2, 0, 0, 0, 0, 0, ...	2006-03-16
142	http://travel.sanook.com/images/bkk.jpg	[4, 0, 0, 0, 0, 0, 0, ...	2006-03-16
143	http://travel.sanook.com/images/chan.jpg	[3, 0, 0, 0, 0, 0, 0, ...	2006-03-16
144	http://music.sanook.com/story_picture/s/11902_001.jpg	[27, 0, 0, 0, 0, 0, 0, ...	2006-03-16
145	http://music.sanook.com/story_picture/s/11895_001.jpg	[47, 1, 0, 0, 0, 0, 0, ...	2006-03-16
146	http://music.sanook.com/story_picture/s/11903_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
147	http://star.sanook.com/story_picture/s/11447_001.jpg	[5, 0, 0, 0, 0, 0, 0, ...	2006-03-16
148	http://star.sanook.com/story_picture/s/11446_001.jpg	[5, 0, 0, 0, 0, 0, 0, ...	2006-03-16
149	http://star.sanook.com/story_picture/s/11445_001.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
150	http://star.sanook.com/activityzone/man.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
151	http://star.sanook.com/story_picture/m/11552_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
152	http://star.sanook.com/story_picture/m/11550_001.jpg	[17, 2, 0, 0, 0, 0, 2, ...	2006-03-16
153	http://star.sanook.com/story_picture/m/11540_001.jpg	[13, 0, 0, 0, 0, 0, 1, ...	2006-03-16
154	http://star.sanook.com/story_picture/m/11541_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
155	http://star.sanook.com/story_picture/m/11538_001.jpg	[15, 0, 0, 0, 0, 0, 0, ...	2006-03-16
156	http://star.sanook.com/story_picture/m/11539_001.jpg	[10, 0, 0, 0, 0, 0, 0, ...	2006-03-16
157	http://star.sanook.com/story_picture/m/11025_001.jpg	[13, 0, 0, 0, 0, 0, 0, ...	2006-03-16
158	http://star.sanook.com/story_picture/m/11192_001.jpg	[10, 0, 0, 0, 0, 0, 0, ...	2006-03-16
159	http://star.sanook.com/story_picture/m/10927_001.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16
160	http://star.sanook.com/story_picture/m/11440_001.jpg	[1, 0, 0, 0, 0, 0, 0, ...	2006-03-16
161	http://star.sanook.com/story_picture/m/11549_001.jpg	[4, 0, 0, 0, 0, 0, 0, ...	2006-03-16
162	http://movie.sanook.com/story_picture/s/10938_001.jpg	[9, 1, 0, 0, 0, 0, 0, ...	2006-03-16
163	http://movie.sanook.com/story_picture/s/10921_001.jpg	[22, 0, 0, 0, 0, 0, 0, ...	2006-03-16
164	http://movie.sanook.com/story_picture/s/10917_001.jpg	[35, 0, 0, 0, 0, 0, 0, ...	2006-03-16
165	http://movie.sanook.com/story_picture/s/10931_001.jpg	[21, 1, 0, 0, 0, 0, 0, ...	2006-03-16
166	http://movie.sanook.com/story_picture/s/10817_001.jpg	[2, 0, 0, 0, 0, 0, 0, ...	2006-03-16
167	http://movie.sanook.com/story_picture/s/10884_001.jpg	[4, 0, 0, 0, 0, 0, 0, ...	2006-03-16
168	http://movie.sanook.com/story_picture/s/10843_001.jpg	[14, 1, 0, 0, 0, 0, 0, ...	2006-03-16
169	http://movie.sanook.com/story_picture/s/10840_001.jpg	[3, 0, 0, 0, 0, 0, 0, ...	2006-03-16
170	http://movie.sanook.com/story_picture/m/10659_001.jpg	[43, 3, 0, 0, 0, 0, 3, ...	2006-03-16
171	http://sex.sanook.com/story_picture/b/02321_003.jpg	[0, 0, 0, 0, 0, 0, 0, ...	2006-03-16

172	http://sex.sanook.com/story_picture/s/26248_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
173	http://sex.sanook.com/story_picture/s/22465_001.jpg	[7, 0, 0, 0, 0, 0, 0, ..	2006-03-16
174	http://sex.sanook.com/story_picture/s/22466_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
175	http://sex.sanook.com/story_picture/m/02323_001.jpg	[7, 0, 0, 0, 0, 0, 2, ..	2006-03-16
176	http://sex.sanook.com/story_picture/m/02324_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
177	http://sex.sanook.com/story_picture/m/02326_001.jpg	[4, 0, 0, 0, 0, 0, 0, ..	2006-03-16
178	http://sex.sanook.com/story_picture/s/20598_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
179	http://sex.sanook.com/story_picture/s/20587_001.jpg	[22, 0, 0, 0, 0, 0, 0, ..	2006-03-16
180	http://sex.sanook.com/story_picture/s/20602_001.jpg	[13, 0, 0, 0, 0, 0, 0, ..	2006-03-16
181	http://sex.sanook.com/story_picture/s/16780_001.jpg	[15, 0, 0, 0, 0, 0, 0, ..	2006-03-16
182	http://women.sanook.com/story_picture/m/16676_001.jpg	[25, 0, 0, 0, 0, 0, 0, ..	2006-03-16
183	http://women.sanook.com/story_picture/m/16729_001.jpg	[4, 0, 0, 0, 0, 0, 0, ..	2006-03-16
184	http://women.sanook.com/story_picture/s/17003_001.jpg	[57, 0, 0, 0, 0, 0, 0, ..	2006-03-16
185	http://women.sanook.com/story_picture/s/17185_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
186	http://women.sanook.com/story_picture/s/17177_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
187	http://women.sanook.com/story_picture/s/17137_001.jpg	[45, 0, 0, 0, 0, 0, 0, ..	2006-03-16
188	http://women.sanook.com/story_picture/s/16971_001.jpg	[18, 0, 0, 0, 0, 0, 0, ..	2006-03-16
189	http://women.sanook.com/story_picture/s/16862_001.jpg	[41, 0, 0, 0, 0, 0, 0, ..	2006-03-16
190	http://women.sanook.com/story_picture/s/17077_001.jpg	[7, 0, 0, 0, 0, 0, 0, ..	2006-03-16
191	http://women.sanook.com/story_picture/s/17178_001.jpg	[31, 0, 0, 0, 0, 0, 0, ..	2006-03-16
192	http://women.sanook.com/story_picture/s/16771_001.jpg	[15, 0, 0, 0, 0, 0, 0, ..	2006-03-16
193	http://women.sanook.com/story_picture/s/06062_001.jpg	[31, 0, 0, 0, 0, 0, 0, ..	2006-03-16
194	http://women.sanook.com/story_picture/s/16679_001.jpg	[15, 0, 0, 0, 0, 0, 0, ..	2006-03-16
195	http://women.sanook.com/story_picture/s/17099_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
196	http://women.sanook.com/story_picture/s/17095_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
197	http://women.sanook.com/story_picture/s/16704_001.jpg	[36, 0, 0, 0, 0, 0, 0, ..	2006-03-16
198	http://women.sanook.com/story_picture/s/16703_001.jpg	[17, 0, 0, 0, 0, 0, 0, ..	2006-03-16
199	http://women.sanook.com/story_picture/s/16711_001.jpg	[12, 0, 0, 0, 0, 0, 0, ..	2006-03-16
200	http://women.sanook.com/story_picture/s/16710_001.jpg	[32, 0, 0, 0, 0, 0, 0, ..	2006-03-16
201	http://women.sanook.com/story_picture/s/16709_001.jpg	[13, 0, 0, 0, 0, 0, 0, ..	2006-03-16
202	http://women.sanook.com/story_picture/s/16708_001.jpg	[21, 0, 0, 0, 0, 0, 0, ..	2006-03-16
203	http://women.sanook.com/story_picture/s/16707_001.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
204	http://women.sanook.com/story_picture/s/16706_001.jpg	[28, 0, 0, 0, 0, 0, 0, ..	2006-03-16
205	http://women.sanook.com/story_picture/s/16725_001.jpg	[17, 0, 0, 0, 0, 0, 0, ..	2006-03-16
206	http://women.sanook.com/story_picture/s/16705_001.jpg	[24, 1, 0, 0, 0, 0, 0, ..	2006-03-16
207	http://women.sanook.com/images/baby.jpg	[15, 0, 0, 0, 0, 0, 0, ..	0000-00-00
208	http://women.sanook.com/story_picture/m/16519_001.jpg	[4, 0, 0, 0, 0, 0, 0, ..	2006-03-16
209	http://women.sanook.com/story_picture/m/16484_001.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
210	http://women.sanook.com/story_picture/m/16414_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
211	http://women.sanook.com/images/baby11.jpg	[15, 0, 0, 0, 0, 0, 0, ..	0000-00-00
212	http://women.sanook.com/story_picture/s/16275_001.jpg	[16, 0, 0, 0, 0, 0, 0, ..	2006-03-16
213	http://women.sanook.com/story_picture/s/16414_001.jpg	[9, 0, 0, 0, 0, 0, 0, ..	2006-03-16
214	http://women.sanook.com/story_picture/s/16484_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
215	http://women.sanook.com/story_picture/s/16519_001.jpg	[12, 0, 0, 0, 0, 0, 1, ..	2006-03-16
216	http://women.sanook.com/story_picture/s/07047_001.jpg	[1, 0, 0, 0, 0, 0, 0, ..	2006-03-16
217	http://women.sanook.com/images/mom_g4.jpg	[15, 0, 0, 0, 0, 0, 0, ..	0000-00-00
218	http://game.sanook.com/images/head_hotgame.jpg	[14, 0, 0, 0, 0, 0, 0, ..	2006-03-16

219	http://game.sanook.com/story_picture/m/02111_001.jpg	[16, 1, 0, 0, 0, 0, 0, ..	2006-03-16
220	http://game.sanook.com/story_picture/s/02120_001.jpg	[17, 0, 0, 0, 0, 0, 0, ..	2006-03-16
221	http://game.sanook.com/story_picture/s/02119_001.jpg	[56, 3, 0, 0, 0, 0, 1, ..	2006-03-16
222	http://game.sanook.com/story_picture/s/02118_001.jpg	[6, 0, 0, 0, 0, 0, 1, ..	2006-03-16
223	http://game.sanook.com/story_picture/s/00023_001.jpg	[78, 0, 0, 0, 0, 0, 0, ..	2006-03-16
224	http://game.sanook.com/story_picture/prepaid.jpg	[36, 0, 0, 0, 0, 0, 0, ..	2006-03-16
225	http://game.sanook.com/story_picture/hardware.jpg	[32, 0, 0, 0, 0, 0, 0, ..	2006-03-16
226	http://game.sanook.com/story_picture/advertise.jpg	[4, 0, 0, 0, 0, 0, 1, ..	2006-03-16
227	http://formula1.sanook.com/images/f1_btimg.jpg	[9, 0, 0, 0, 0, 0, 0, ..	2006-03-16
228	http://formula1.sanook.com/story_picture/s/00392_001.jpg	[9, 1, 0, 0, 0, 0, 0, ..	2006-03-16
229	http://formula1.sanook.com/story_picture/s/00391_001.jpg	[5, 0, 0, 0, 0, 0, 0, ..	2006-03-16
230	http://formula1.sanook.com/story_picture/m/00390_001.jpg	[26, 6, 0, 0, 0, 0, 0, ..	2006-03-16
231	http://formula1.sanook.com/story_picture/m/00389_001.jpg	[1, 0, 0, 0, 0, 0, 0, ..	2006-03-16
232	http://formula1.sanook.com/story_picture/s/00058_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
233	http://formula1.sanook.com/story_picture/s/00061_001.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
234	http://formula1.sanook.com/images/f1_btimg_bt.jpg	[16, 4, 0, 0, 0, 0, 1, ..	2006-03-16
235	http://hitech.sanook.com/story_picture/m/06886_001.jpg	[21, 3, 0, 0, 0, 0, 0, ..	2006-03-16
236	http://hitech.sanook.com/story_picture/m/06891_001.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
237	http://hitech.sanook.com/story_picture/m/06895_001.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
238	http://hitech.sanook.com/story_picture/m/06898_001.jpg	[4, 3, 1, 0, 0, 0, 1, ..	2006-03-16
239	http://hitech.sanook.com/story_picture/m/06897_001.jpg	[9, 0, 0, 0, 0, 0, 0, ..	2006-03-16
240	http://hitech.sanook.com/story_picture/m/06894_001.jpg	[18, 2, 0, 0, 0, 0, 0, ..	2006-03-16
241	http://hitech.sanook.com/story_picture/m/06893_001.jpg	[23, 0, 0, 0, 0, 0, 0, ..	2006-03-16
242	http://education.sanook.com/story_picture/s/01569_001.jpg	[27, 0, 0, 0, 0, 0, 0, ..	2006-03-16
243	http://education.sanook.com/story_picture/s/01563_001.jpg	[20, 0, 0, 0, 0, 0, 0, ..	2006-03-16
244	http://education.sanook.com/story_picture/s/01562_001.jpg	[13, 0, 0, 0, 0, 0, 0, ..	2006-03-16
245	http://education.sanook.com/story_picture/b/01570_001.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
246	http://education.sanook.com/story_picture/s/01560_001.jpg	[7, 0, 0, 0, 0, 0, 0, ..	2006-03-16
247	http://education.sanook.com/story_picture/s/01558_001.jpg	[28, 2, 0, 0, 0, 0, 0, ..	2006-03-16
248	http://education.sanook.com/story_picture/s/01213_001.jpg	[33, 3, 0, 0, 0, 0, 0, ..	2006-03-16
249	http://education.sanook.com/story_picture/m/01175_001.jpg	[12, 0, 0, 0, 0, 0, 0, ..	2006-03-16
250	http://joke.sanook.com/picture/39342_0.jpg	[1, 0, 0, 0, 0, 0, 0, ..	2006-03-16
251	http://download.sanook.com/picture/qq-thai-3.0.jpg	[10, 0, 0, 0, 0, 0, 0, ..	2006-03-16
252	http://download.sanook.com/picture/mahjongsmall.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
253	http://download.sanook.com/picture/por-in-s.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
254	http://download.sanook.com/picture/temp-download.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16

255	http://download.sanook.com/picture/dekhor-s.jpg	[42, 5, 0, 0, 0, 0, 0, ..	2006-03-16
256	http://download.sanook.com/picture/maxthon-small.jpg	[3, 0, 0, 0, 0, 0, 0, ..	2006-03-16
257	http://club.sanook.com/pictures/sm/5/sm0038930402.jpg	[4, 0, 0, 0, 0, 0, 0, ..	2006-03-16
258	http://club.sanook.com/pictures/bg/5/bg0038930401.jpg	[5, 0, 0, 0, 0, 0, 0, ..	2006-03-16
259	http://qq.sanook.com/php/my_pic/7081057/my_2.jpg	[1, 0, 0, 0, 0, 0, 0, ..	2006-03-16
260	http://qq.sanook.com/php/my_pic/6992810/my_1.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
261	http://qq.sanook.com/php/my_pic/6914515/my_1.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
262	http://qq.sanook.com/php/my_pic/5608614/my_1.jpg	[5, 0, 0, 0, 0, 0, 6, ..	2006-03-16
263	http://qq.sanook.com/php/my_pic/7027440/my_1.jpg	[21, 0, 0, 0, 0, 0, 0, ..	2006-03-16
264	http://qq.sanook.com/php/my_pic/6714330/my_1.jpg	[1, 1, 0, 0, 0, 0, 0, ..	2006-03-16
265	http://diary.sanook.com/picture/daily/t/tony_nz@sanook.com_20060315.jpg	[1, 0, 0, 0, 0, 0, 0, ..	0000-00-00
266	http://blog.sanook.com/PortalPics/_zen/images/default/bb_b.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
267	http://blog.sanook.com/PortalPics/_zen/images/default/03.JPG	[10, 0, 0, 0, 0, 0, 0, ..	2006-03-16
268	http://blog.sanook.com/PortalPics/_zen/images/default/Woodland-view-in-fog-with-ferns-and-decaying-tree-trunk-Photographic-Print-C10241164.jpg	[1, 0, 0, 0, 0, 0, 0, ..	2006-03-16
269	http://albumfiles.sanook.com/MyAlbums/ImageStore/6e87ceac-1e76-4927-9a36-3071943b10be/11e98a8e-392d-44cb-9a21-fae63b1cb381_thumb.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
270	http://albumfiles.sanook.com/MyAlbums/ImageStore/6e87ceac-1e76-4927-9a36-3071943b10be/e895c6d2-f004-430e-9a3e-ceaafc3a3e61_thumb.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
271	http://albumfiles.sanook.com/MyAlbums/ImageStore/6e87ceac-1e76-4927-9a36-3071943b10be/fcd56cd8-323a-414e-bb64-d60d73f8b42c_thumb.jpg	[0, 0, 0, 0, 0, 0, 0, ..	2006-03-16
272	http://blog.sanook.com/images/no-photo.jpg	[43, 0, 0, 0, 0, 0, 0, ..	2006-03-16
273	http://talk.sanook.com/story_picture/b/05543_001.jpg	[13, 3, 0, 0, 0, 0, 0, ..	2006-03-16
274	http://talk.sanook.com/story_picture/source/source_003.jpg	[0, 0, 0, 0, 0, 0, 2, ..	2006-03-16
275	http://formula1.sanook.com/story_picture/m/00392_002.jpg	[42, 0, 0, 0, 0, 0, 6, ..	2006-03-16
276	http://formula1.sanook.com/story_picture/m/00392_003.jpg	[10, 0, 0, 0, 0, 0, 0, ..	2006-03-16
277	http://formula1.sanook.com/story_picture/m/00392_004.jpg	[6, 11, 0, 0, 0, 0, 0, ..	2006-03-16
278	http://formula1.sanook.com/story_picture/m/00392_005.jpg	[44, 0, 0, 0, 0, 0, 1, ..	2006-03-16
279	http://formula1.sanook.com/story_picture/m/00392_006.jpg	[21, 0, 0, 0, 0, 0, 0, ..	2006-03-16
280	http://formula1.sanook.com/story_picture/m/00392_007.jpg	[4, 0, 0, 0, 0, 0, 1, ..	2006-03-16
281	http://formula1.sanook.com/story_picture/m/00392_008.jpg	[26, 1, 0, 0, 0, 0, 0, ..	2006-03-16
282	http://formula1.sanook.com/story_picture/m/00392_009.jpg	[47, 0, 0, 0, 0, 0, 0, ..	2006-03-16
283	http://formula1.sanook.com/story_picture/m/00392_010.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
284	http://formula1.sanook.com/story_picture/m/00392_011.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16

285	http://formula1.sanook.com/story_picture/m/00392_012.jpg	[16, 2, 0, 0, 0, 0, 1, ..	2006-03-16
286	http://formula1.sanook.com/story_picture/m/00392_013.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
287	http://formula1.sanook.com/story_picture/m/00392_014.jpg	[12, 2, 0, 0, 0, 0, 0, ..	2006-03-16
288	http://formula1.sanook.com/story_picture/m/00392_015.jpg	[9, 0, 0, 0, 0, 0, 0, ..	2006-03-16
289	http://formula1.sanook.com/story_picture/m/00392_016.jpg	[27, 5, 0, 0, 0, 0, 0, ..	2006-03-16
290	http://formula1.sanook.com/story_picture/m/00392_017.jpg	[27, 0, 0, 0, 0, 0, 0, ..	2006-03-16
291	http://formula1.sanook.com/story_picture/m/00392_018.jpg	[11, 1, 0, 0, 0, 0, 0, ..	2006-03-16
292	http://formula1.sanook.com/story_picture/m/00392_019.jpg	[13, 0, 0, 0, 0, 0, 0, ..	2006-03-16
293	http://formula1.sanook.com/story_picture/m/00392_020.jpg	[11, 0, 0, 0, 0, 0, 0, ..	2006-03-16
294	http://formula1.sanook.com/story_picture/m/00392_021.jpg	[24, 0, 0, 0, 0, 0, 0, ..	2006-03-16
295	http://formula1.sanook.com/story_picture/m/00392_022.jpg	[6, 0, 0, 0, 0, 0, 0, ..	2006-03-16
296	http://formula1.sanook.com/story_picture/m/00392_023.jpg	[10, 0, 0, 0, 0, 0, 0, ..	2006-03-16
297	http://formula1.sanook.com/story_picture/m/00392_024.jpg	[7, 0, 0, 0, 0, 0, 0, ..	2006-03-16
298	http://formula1.sanook.com/story_picture/m/00392_025.jpg	[8, 0, 0, 0, 0, 0, 0, ..	2006-03-16
299	http://talk.sanook.com/story_picture/b/05549_001.jpg	[28, 1, 0, 0, 0, 0, 0, ..	2006-03-16
300	http://talk.sanook.com/story_picture/source/source_001.jpg	[9, 0, 0, 0, 1, 0, 0, ..	2006-03-16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก ข
ตัวอย่างรูปภาพในฐานข้อมูลจำนวน 300 ภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

ตัวอย่างรูปภาพในฐานะข้อมูลจำนวน 300 ภาพ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



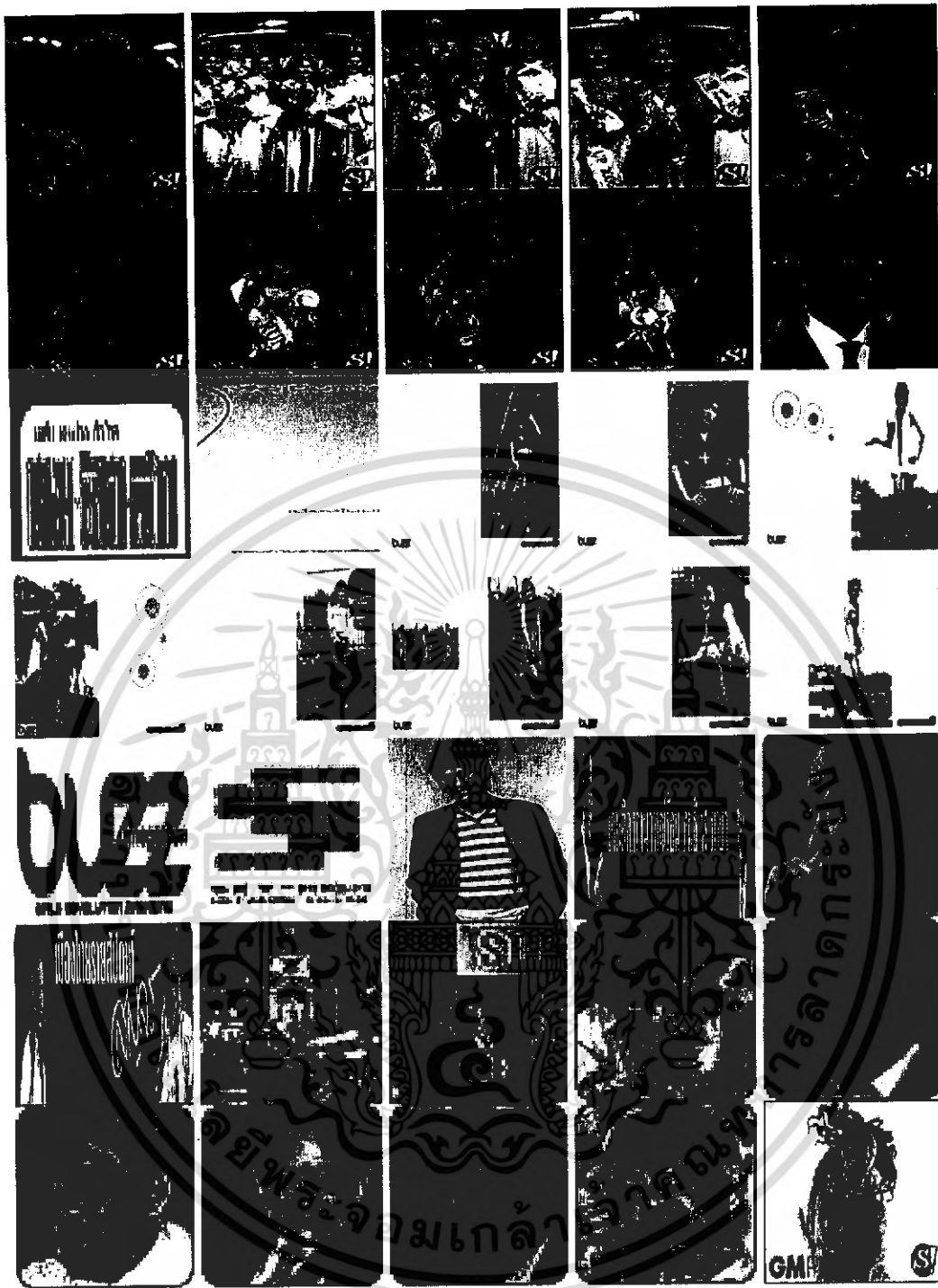
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้