

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

โปรแกรมการทำเหมืองข้อมูลการใช้เว็บโดยใช้เทคนิครูปแบบลำดับ
และการวิเคราะห์ทางสถิติ

WEB USAGE MINING TOOL USING SEQUENTIAL PATTERN
AND STATISTICAL ANALYSIS TECHNIQUES



อาจารย์ที่ปรึกษา
ผศ.ดร.ภัทรชัย ลลิตโรจน์วงศ์

วัน เดือน ปี.....	04 ส.ค. 2550
เลขทะเบียน.....	H003487
เลขเรียกหนังสือ.....	ศ.491จ 2549
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."	

6.11841266
11/1/51470

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานาน ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ภาคเรียนที่ 2 ปีการศึกษา 2549
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**WEB USAGE MINING TOOL USING SEQUENTIAL PATTERN
AND STATISTICAL ANALYSIS TECHNIQUES**



**A SYSTEM DEVELOPMENT PROJECT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อ ก 2/2006 เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2007

SCHOOL OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่ลงพิมพ์ในนามของสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เพื่อใช้ในการศึกษาเท่านั้น อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	โปรแกรมการทำเหมืองข้อมูลการใช้เว็บโดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ
นักศึกษา	นายศิริศักดิ์ แก้วเป็ย
รหัสนักศึกษา	47066436
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2549
อาจารย์ที่ปรึกษา	ผศ.ดร.ภัทรชัย ลลิตโรจน์วงศ์

บทคัดย่อ

โครงการฉบับนี้เสนอวิธีการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลการใช้งานเว็บไซต์ เพื่อค้นหารูปแบบการเยี่ยมชมเว็บไซต์จากเว็บเซิร์ฟเวอร์ล็อก โดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ เพื่อช่วยการทำเหมืองข้อมูลการใช้งานเว็บไซต์จะช่วยค้นหาความสัมพันธ์และแนวโน้มต่างๆทางธุรกิจ และยังช่วยให้เข้าใจลูกค้าและดำเนินธุรกิจอิเล็กทรอนิกส์ได้อย่างมีประสิทธิภาพ การทำเหมืองข้อมูลการใช้งานเว็บไซต์มี 3 เฟส คือ การเตรียมข้อมูลการใช้งาน ค้นหาแบบ และวิเคราะห์รูปแบบ

Title Web Usage Mining Tool Using Sequential Pattern
and Statistical Analysis Techniques

Student Mr. Sirisak Kaewpia

Student ID. 47066436

Degree Master of Science

Programme Information Science

Academic Year 2006

Advisor Asst. Prof. Dr. Pattarachai Lalitrojwong

ABSTRACT

This project applies web usage mining techniques to discover usage patterns from web server log. The project presents models for web usage mining that employs sequential pattern and statistical analysis techniques. Web usage mining enable us to discover meaningful business correlations and trends, and provides better understanding of customers and performing business efficiently. The process of web usage mining consists of three phases, namely usage data preprocessing, pattern discovery and pattern analysis.

กิตติกรรมประกาศ

โครงการฉบับนี้สำเร็จได้อย่างดี ด้วยความช่วยเหลือของอาจารย์ที่ปรึกษา ผศ.ดร.ภัทรชัย ลลิตโรจน์วงศ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่างๆ ข้าพเจ้ารู้สึกทราบบ้างในความอนุเคราะห์จากท่านอาจารย์และขอขอบพระคุณเป็นอย่างสูง

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกๆ เรื่อง ทำให้ข้าพเจ้าสามารถทำโครงการฉบับนี้สำเร็จลุล่วงด้วยดี

คุณค่าและประโยชน์อันพึงมาจากโครงการฉบับนี้ ข้าพเจ้าขอบแต่ผู้มีพระคุณทุกท่าน

ศิริศักดิ์ แก้วเป็ย



สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 ขอบเขตของโครงการ.....	1
1.4 ขั้นตอนการดำเนินการศึกษา.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.6 โครงสร้างของรายงาน.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 การทำเหมืองข้อมูลเว็บ	3
2.2 การทำเหมืองข้อมูลการใช้งานเว็บ.....	5
2.3 ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บ.....	8
2.4 การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ.....	17
2.5 รูปแบบลำดับ.....	19
2.6 วิธีการสร้างรูปแบบลำดับ.....	25
2.7 การอ้างอิงไปข้างหน้าไกลที่สุด.....	26
2.8 งานวิจัย WEBMINER.....	30
2.9 งานวิจัย SpeedTracer.....	31
บทที่ 3 การออกแบบโปรแกรมการทำเหมืองข้อมูลการใช้งานเว็บไซค์.....	32
3.1 รูปแบบลักษณะการดำเนินธุรกิจ.....	32
3.2 ความต้องการของระบบ.....	32
3.3 การวิเคราะห์และออกแบบระบบ.....	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.4 การออกแบบฐานข้อมูล	47
บทที่ 4 การทดลองและผลการทดลอง	54
4.1 สภาพแวดล้อมในการพัฒนาระบบ.....	54
4.2 วิธีการทดลอง.....	54
4.3 ผลการทดลอง.....	59
4.4 วิธีการนำรูปแบบลำดับไปใช้ปรับโครงสร้างลิงก์ของเว็บไซต์.....	69
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	86
5.1 สรุปผลโครงการ.....	86
5.2 ข้อเสนอแนะ.....	87
บรรณานุกรม.....	88
ภาคผนวก ก.....	89
ประวัติผู้เขียน.....	108

สารบัญตาราง

ตารางที่	หน้า
2.1 ฟิวด์ของล็อกไฟล์แบบคอมไบน์.....	7
2.2 นิยามคำศัพท์ที่ใช้ในการใช้งานเว็บ.....	7
2.3 กลไกการระบุตัวผู้ใช้.....	12
2.4 สรุปผลลัพธ์จากการทำกระบวนการเตรียมข้อมูล.....	14
2.5 ตัวอย่างการทำงานโดยใช้อัลกอริทึม MF.....	29
3.1 คำอธิบายยูสเคส Import Web Server Log.....	33
3.2 คำอธิบายยูสเคส Configure System.....	34
3.3 คำอธิบายยูสเคส Mine Web Server Log	35
3.4 คำอธิบายยูสเคส Mine Top Frequent Traversal Path.....	36
3.5 คำอธิบายยูสเคส Analyse Statistics.....	37
3.6 คำอธิบายยูสเคส Create Report.....	38
3.7 ตาราง Accounts.....	50
3.8 ตาราง Extensions.....	50
3.9 ตาราง Logfile.....	50
3.10 ตาราง Cleanedlog.....	50
3.11 ตาราง Cleanedlog_prep.....	51
3.12 ตาราง Url.....	51
3.13 ตาราง Agents.....	51
3.14 ตาราง HostAgent.....	51
3.15 ตาราง UserSequence.....	51
3.16 ตาราง Sessions.....	52
3.17 ตาราง Transactions.....	52
3.18 ตาราง Ck.....	52
3.19 ตาราง Pruned_Ck.....	52
3.20 ตาราง Lk.....	52
3.21 ตาราง Rk.....	53
3.22 ตาราง HourOfDay.....	53
3.23 ตาราง IpStats.....	53

เอกสารนี้เป็นทรัพย์สินของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
3.24 ตาราง NumberofVisits.....

การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านอื่น
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.1 ตัวอย่างรูปแบบลำดับที่ค้นพบ.....	70
4.2 เนื้อหาเกี่ยวกับเพลงต่างๆ ในตัวอย่างรูปแบบลำดับ.....	71



สารบัญรูป

รูปที่	หน้า
2.1 ประเภทของการทำเหมืองข้อมูลเว็บ.....	4
2.2 แหล่งที่มาของข้อมูลการใช้งานเว็บ.....	5
2.3 ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บอย่างง่าย	8
2.4 รายละเอียดขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บ.....	9
2.5 รายละเอียดการเตรียมข้อมูลการใช้งาน	9
2.6 ตัวอย่างเว็บเซิร์ฟเวอร์ล็อกไฟล์	12
2.7 การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ.....	17
2.8 ฐานข้อมูลต้นฉบับ.....	21
2.9 ฐานข้อมูลที่เรียงโดยรหัสลูกค้าและเวลาในทรานแซคชัน.....	21
2.10 ชุดลำดับลูกค้า.....	21
2.11 รูปแบบลำดับที่มีค่าสนับสนุนมากกว่า 25%.....	22
2.12 อัลกอริทึม AprioriAll	22
2.13 อัลกอริทึมการสร้างลำดับคู่แข่ง.....	23
2.14 ตัวอย่างการทำงานของอัลกอริทึม AprioriAll	24
2.15 อัลกอริทึมการสร้างรูปแบบลำดับ.....	26
2.16 ตัวอย่างของเส้นทางการเดินทางของผู้ใช้.....	27
2.17 อัลกอริทึม MF.....	28
2.18 สถาปัตยกรรมของ WEBMINER	30
2.19 โพลีไดอะแกรมสำหรับการพัฒนา SpeedTracer.....	31
3.1 ยูสเคสไดอะแกรมของระบบ.....	33
3.2 แผนภาพกิจกรรมของระบบ.....	39
3.3 คลาสไดอะแกรมของระบบ.....	41
3.4 ซีควเอนซ์ไดอะแกรมการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์.....	43
3.5 ซีควเอนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์โดยใช้เทคนิครูปแบบลำดับ.....	44
3.6 ซีควเอนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยการวิเคราะห์ทางสถิติ.....	45
3.7 ซีควเอนซ์ไดอะแกรมการกำหนดค่าและจัดการข้อมูลให้กับระบบ.....	46
3.8 ซีควเอนซ์ไดอะแกรมการออกรายงานของระบบ.....	46
3.9 ฐานข้อมูลของระบบ.....	49
4.1 หน้าจอล็อกอินเข้าสู่โปรแกรม.....	55

เอกสารนี้เป็นทรัพย์สินทางปัญญาของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.2 หน้าจอการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์.....	56
4.3 หน้าจอการล็อกไฟล์ที่ต้องการนำเข้า.....	56
4.4 หน้าจอการกำหนดค่าเวลานอก.....	57
4.5 หน้าจอการกำหนดค่าที่ใช้ในการค้นหาแบบลำดับ.....	57
4.6 หน้าจอการกำหนดชื่อเว็บไซต์และรายงานที่ต้องการ.....	58
4.7 หน้าจอการสั่งให้โปรแกรมประมวลผล.....	58
4.8 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 2 ตัว.....	59
4.9 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 3 ตัว.....	61
4.10 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 4 ตัว.....	63
4.11 รายงานทางสถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง.....	64
4.12 รายงานทางสถิติแยกตามไอพีแอดเดรสของผู้ใช้.....	65
4.13 รายงานทางสถิติแยกตามเอเจนท์.....	66
4.14 รายงานทางสถิติแยกตามจำนวนการเข้าชมเว็บไซต์.....	67
4.15 รายงานทางสถิติแยกตามเว็บเพจที่เข้าชม.....	68
4.16 ตัวอย่างเว็บเพจที่มีการปรับปรุงโครงสร้าง.....	69
4.17 เว็บเพจ http://www.differentasia.com/destinations/Chiang_Mai.php	74
4.18 เว็บเพจ http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php	76
4.19 เว็บเพจ http://www.differentasia.com/destinations/Koh_Chang_Surrounding_Islands.php	77
4.20 เว็บเพจ http://www.differentasia.com/destinations/Koh_Samui.php	78
4.21 เว็บเพจ http://www.differentasia.com/destinations/Pranburi.php	80
4.22 เว็บเพจ http://www.differentasia.com/index.php	81
4.23 เว็บเพจ http://www.differentasia.com/destinations/Cha_Am.php	83
4.24 เว็บเพจ http://www.differentasia.com/destinations/Krabi.php	84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการทำธุรกิจอิเล็กทรอนิกส์นั้น ปัจจัยหนึ่งที่มีความสำคัญอย่างมากต่อความสำเร็จของธุรกิจคือ ความเข้าใจในตัวลูกค้าหรือกลุ่มของลูกค้า ยิ่งรู้ข้อมูลมากก็ยิ่งทำให้เข้าใจลูกค้า โอกาสที่จะทำธุรกิจให้ตรงกับความต้องการของตลาดก็จะมีมากขึ้นไปด้วย การเจริญเติบโตของเว็บไซต์ที่เพิ่มขึ้นอย่างรวดเร็ว จึงทำให้ธุรกิจอิเล็กทรอนิกส์เพิ่มขึ้นอย่างรวดเร็ว ข้อมูลสำคัญซึ่งธุรกิจอิเล็กทรอนิกส์ควรทราบคือ มีใครเข้ามาเยี่ยมชมเว็บไซต์บ้าง มีจำนวนคนเข้ามาเยี่ยมชมมากน้อยเท่าไร เข้ามาเยี่ยมชมเมื่อใดและเข้ามาทำอะไรบนเว็บไซต์บ้าง ถ้าเราทราบข้อมูลเหล่านี้จะทำให้เราได้เปรียบคู่แข่ง การทำเหมืองข้อมูลการใช้งานเว็บไซต์สำหรับธุรกิจอิเล็กทรอนิกส์สามารถนำไปประยุกต์ใช้ในการจัดการความสัมพันธ์กับลูกค้า โดยการวิเคราะห์ล็อกไฟล์ซึ่งบันทึกกิจกรรมต่างๆที่เกิดขึ้นบนเว็บไซต์ ซึ่งโดยปกติแล้วแทร็กเกอร์โดยทั่วไปที่มีให้มาพร้อมกับโฮสต์แทบจะไม่มีการวิเคราะห์หาความสัมพันธ์ระหว่างเว็บเพจที่ถูกเยี่ยมชมและลำดับการเยี่ยมชมเว็บเพจ

การวิเคราะห์การเข้าใช้ โดยการทำเหมืองข้อมูลการใช้งานเว็บไซต์ จะช่วยหาความสัมพันธ์ระหว่างเว็บเพจและลำดับการเยี่ยมชม ซึ่งองค์ความรู้ที่ได้เหล่านี้เป็นประโยชน์ต่อการดำเนินธุรกิจอิเล็กทรอนิกส์

โครงการนี้จะมุ่งเน้นวิธีการสร้างเครื่องมือเพื่อหาความสัมพันธ์ระหว่างเว็บเพจและรูปแบบลำดับการเยี่ยมชมเว็บเพจ โดยการวิเคราะห์จากเว็บไซต์ล็อก

1.2 วัตถุประสงค์ของโครงการ

เพื่อพัฒนาเครื่องมือที่ใช้ค้นหารูปแบบลำดับการเยี่ยมชมเว็บไซต์และวิเคราะห์ทางสถิติการเข้าใช้งานเว็บไซต์ โดยทำการประมวลผลจากเว็บไซต์ล็อกของการเข้าถึงเว็บไซต์

1.3 ขอบเขตของโครงการ

1. วิเคราะห์รูปแบบลำดับการเยี่ยมชมเว็บเพจบนเครื่องเซิร์ฟเวอร์เพียงเครื่องเดียวเท่านั้น
2. นำเว็บไซต์ล็อกไฟล์ของการเข้าถึงเว็บไซต์แบบคอมไบนีหรือเอ็กซ์เทนชันเท่านั้นที่ใช้ประมวลผล

3. ค้นหารูปแบบลำดับในการเข้าถึงเว็บเพจจากเว็บเซิร์ฟเวอร์ล็อก โดยในระบบนี้จะใช้อัลกอริทึม AprioriAll
4. วิเคราะห์ข้อมูลทางสถิติการเข้าใช้งานเว็บ
5. ระบบจะประมวลผลกับเว็บเซิร์ฟเวอร์ล็อกที่บันทึกการเข้าใช้งานใน 1 วันเท่านั้น
6. ออกรายงานรูปแบบลำดับและรายงานทางสถิติการเข้าใช้งานเว็บ

1.4 ขั้นตอนการดำเนินการศึกษา

1. ศึกษางานวิจัยที่เกี่ยวข้อง
2. ศึกษาความต้องการของผู้ใช้เกี่ยวกับรูปแบบลำดับและข้อมูลของรายงานที่ต้องการ
3. ศึกษารูปแบบของเว็บเซิร์ฟเวอร์ล็อกไฟล์และเว็บเซิร์ฟเวอร์ของเว็บไซต์
4. ศึกษาทฤษฎีและแนวทางที่ใช้ในการประมวลผล
5. ออกแบบและพัฒนาวิธีการหารูปแบบลำดับการเยี่ยมชมเว็บไซต์และการวิเคราะห์ทางสถิติ
6. สรุปผลและจัดทำรายงานโครงการ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ช่วยแสดงสถิติการเข้าใช้งานเว็บ ทำให้สามารถวิเคราะห์ความนิยมของเว็บไซต์ได้
2. ช่วยแสดงลำดับการเยี่ยมชมเว็บเพจ
3. สามารถนำผลลัพธ์ที่ได้มาช่วยในการวิเคราะห์พฤติกรรมในการเข้าถึงเว็บของผู้ใช้เพื่อทำการเปลี่ยนโครงสร้างเว็บ เพื่อให้สามารถเข้าถึงได้ง่ายและทั่วถึงมากยิ่งขึ้น
4. สามารถนำเอาข้อมูลสารสนเทศที่มีอยู่แล้วในองค์กรมาใช้ให้เป็นประโยชน์มากขึ้น เพื่อช่วยในการตัดสินใจและเพิ่มความสามารถทางการดำเนินธุรกิจ

1.6 โครงสร้างของรายงาน

เนื้อหาของรายงานแบ่งออกเป็น 5 บทดังนี้ คือ บทที่ 1 บทนำ บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง เช่น การทำเหมืองข้อมูลเว็บ การทำเหมืองข้อมูลการใช้งานเว็บ การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ รูปแบบลำดับและงานวิจัยที่เกี่ยวข้อง เป็นต้น บทที่ 3 การออกแบบโปรแกรมการทำเหมืองข้อมูลการใช้งานเว็บ บทที่ 4 การทดลองและผลการทดลอง บทที่ 5 ซึ่งเป็นบทสุดท้าย จะเป็นการสรุปผลโครงการและข้อเสนอแนะในการพัฒนาขั้นตอนวิธีการค้นหารูปแบบลำดับการเข้าใช้งานเว็บ

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้องในการทำเหมืองข้อมูลการใช้งานเว็บ ประกอบไปด้วย พื้นฐานของการทำเหมืองข้อมูลเว็บ การทำเหมืองข้อมูลการใช้งานเว็บ ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บ การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ รูปแบบลำดับวิธีการสร้างรูปแบบลำดับ การอ้างอิงไปข้างหน้าไกลที่สุด และงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูลเว็บ

การทำเหมืองข้อมูลเว็บ (Web Mining) เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการค้นหาคำความรู้จากเนื้อหา โครงสร้าง การใช้งานเว็บไซต์ และคำบรรยายลักษณะผู้ใช้

2.1.1 ข้อมูลของเว็บ

ข้อมูลสำหรับการทำเหมืองข้อมูลเว็บ (Web Data) สามารถที่จะรวบรวมได้จากหลายแหล่งข้อมูล อาทิเช่น เซิร์ฟเวอร์ โคลเอนท์ หรือซี ฐานข้อมูลขององค์กร เป็นต้น ข้อมูลแต่ละแหล่งแตกต่างกันไปตามการจำแนกซึ่งแบ่งออกเป็นสี่กลุ่มดังนี้ (Srivastava, J. et al. 2000 : 12-13)

2.1.1.1 เนื้อหา

เนื้อหาของเว็บ (Content Data) คือกลุ่มของอ็อบเจกต์และความสัมพันธ์ที่ประกอบทางด้านข้อความ เอกสาร HTML รูปภาพ ไฟล์เสียง วิดีโอคลิป เป็นต้น

2.1.1.2 ข้อมูลโครงสร้าง

ข้อมูลโครงสร้าง (Structure Data) จะอธิบายโครงสร้างภายในเว็บเพจ (Intra-Page) ซึ่งการอธิบายทางด้านโครงสร้างในแต่ละเว็บเพจ เช่น คำสำคัญ คุณลักษณะของเอกสาร แท็กความหมาย หรือตัวแปร HTTP และไฮเปอร์ลิงก์ (Inter-Page) ที่ลิงก์ระหว่างเว็บเพจ

2.1.1.3 ข้อมูลการใช้งาน

ข้อมูลการใช้งาน (Usage Data) เป็นข้อมูลที่อธิบายการใช้งานเว็บเพจ เช่น ไอพี แอดเดรส วันและเวลาเข้าใช้งานเว็บ เป็นต้น ข้อมูลนี้จะได้มาจากเว็บเซิร์ฟเวอร์ล็อก

2.1.1.4 คำบรรยายลักษณะผู้ใช้

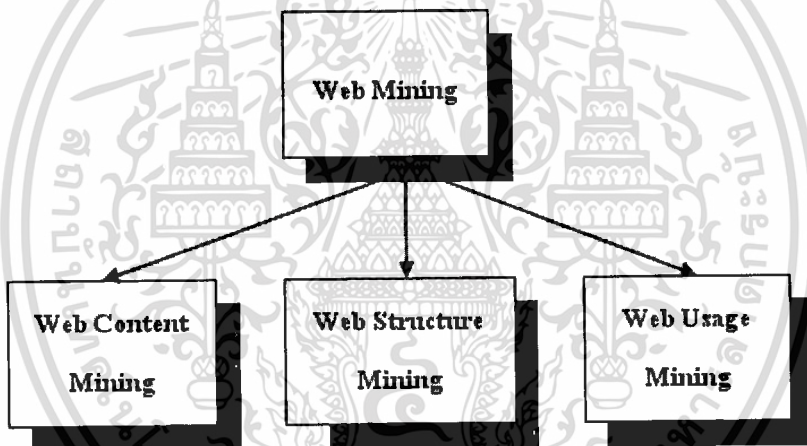
ข้อมูลคำบรรยายลักษณะผู้ใช้ (User Profile Data) มีการวิเคราะห์ออกมาใน 3 ลักษณะดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Demographics เป็นข้อมูลเกี่ยวกับที่อยู่ หรือสถานที่ของผู้ใช้ในขณะที่เข้าใช้บริการเว็บไซต์ หรือข้อมูลการลงทะเบียน
2. Psychographic เป็นข้อมูลซึ่งจะแสดงถึงพฤติกรรมหรือค่านิยมทางด้านต่างๆของผู้ใช้ โดยสามารถแยกกลุ่มผู้ใช้ตามข้อมูลการเข้าใช้บริการเว็บไซต์ ทั้งในแง่ของเวลาและเนื้อหา
3. Technological เป็นข้อมูลที่แสดงถึงระดับความรู้ หรือความสนใจในเทคโนโลยีด้านต่างๆ ของผู้เข้าร่วมถึงเครื่องคอมพิวเตอร์ที่ติดต่อเข้ามาด้วย

2.1.2 ประเภทของการทำเหมืองข้อมูลเว็บ

เหมืองข้อมูลเว็บ (Web Mining) เป็นการประยุกต์ใช้เทคนิคเหมืองข้อมูล (Data Mining) และการเรียนรู้ของเครื่อง (Machine Learning) ในการค้นหาองค์ความรู้จากเนื้อหา โครงสร้าง และ การใช้งานเว็บ การทำเหมืองข้อมูลเว็บแบ่งเป็น 3 ประเภท ดังรูปที่ 2.1



รูปที่ 2.1 ประเภทของการทำเหมืองข้อมูลเว็บ

1. การทำเหมืองข้อมูลเนื้อหาเว็บ (Web Content Mining) เป็นกระบวนการในการค้นหาองค์ความรู้ (Knowledge) และข้อมูลที่เป็นประโยชน์จากเนื้อหาของเว็บเพจ มี 2 วิธีในการทำเหมืองข้อมูลเนื้อหาเว็บไซด์ดังนี้

- Agent-Based Approach

ใช้ตัวแทนที่สามารถค้นหา กรอง และแยกประเภทข้อมูลได้ โดยใช้เทคนิคต่างๆ เช่น เทคนิคการแยกโดยใช้ลักษณะเฉพาะของโดเมน(Domain Characteristic) กำบรยายลักษณะผู้ใช้ เทคนิคการค้นคืนสารสนเทศ เป็นต้น วิธีที่ใช้ตัวแทนแบ่งออกเป็น 3 กลุ่มดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Intelligent Search Agent เป็นตัวแทนการค้นหาที่จะใช้คุณลักษณะของโดเมน และคำบรรยายลักษณะผู้ใช้เพื่อค้นหาสารสนเทศที่ผู้ใช้ต้องการตัวอย่างเช่น OCCAM, ParaSite, FAQ-Finder, Information Manifold

2. Information Filtering/Categorization เป็นตัวแทนที่ช่วยผู้ใช้ในการกรองสารสนเทศที่ไม่เกี่ยวข้องออก เลือกเอาเฉพาะสารสนเทศที่ตรงความต้องการผู้ใช้ โดยใช้ลักษณะไฮเปอร์เท็กซ์ของเนื้อหาเว็บไซต์ในการค้นคืนอัตโนมัติ การคัดสรรและการแบ่งแยกประเภท ตัวอย่างเช่น BO (Bookmark Organizer), HyPursuit

3. Personalized Web Agent เป็นตัวแทนที่มีความสามารถในการเรียนรู้สิ่งที่ผู้ใช้สนใจ และค้นคืนข้อมูลสารสนเทศที่ตรงกับความต้องการของผู้ใช้

- Database Approach

วิธีนี้จะเก็บสารสนเทศที่อยู่บนเว็บ ซึ่งมีลักษณะเป็นข้อมูลที่ไม่มีโครงสร้างมากกว่าข้อมูลที่มีลักษณะแบบมีโครงสร้าง มีการเก็บรวบรวมข้อมูลจากเว็บไซต์ลงฐานข้อมูลและวิเคราะห์ข้อมูลโดยเทคนิคการสอบถามที่เป็นมาตรฐานของฐานข้อมูลและใช้เทคนิคการทำเหมืองข้อมูล

2. การทำเหมืองข้อมูลโครงสร้างเว็บ (Web Structure Mining) เป็นกระบวนการในการค้นหาโครงสร้างข้อมูลของเว็บไซต์ที่อยู่ภายในเพจ (Intra-Page) หรือระหว่างเพจ (Inter-Page) ซึ่งก็คือการศึกษาในระดับไฮเปอร์ลิงก์ เรียกอีกอย่างว่า การวิเคราะห์ไฮเปอร์ลิงก์ (Hyperlink Analysis)

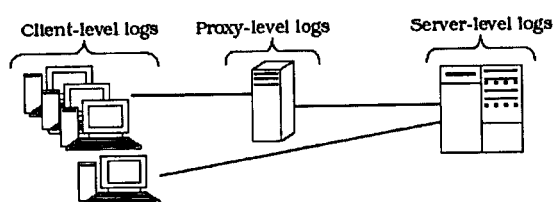
3. การทำเหมืองข้อมูลการใช้งานเว็บ (Web Usage Mining) เป็นกระบวนการในการศึกษาพฤติกรรมการใช้งานของผู้ใช้เพื่อค้นหารูปแบบลำดับการเข้าใช้งานเว็บของผู้ใช้

2.2 การทำเหมืองข้อมูลการใช้งานเว็บ

การทำเหมืองข้อมูลการใช้งานเว็บไซด์เป็นกระบวนการที่ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อค้นหารูปแบบลำดับการเข้าใช้งานจากข้อมูลเว็บ

2.2.1 แหล่งข้อมูล

แหล่งที่มาของข้อมูลการใช้งานเว็บ (Data Sources) แสดงดังรูปที่ 2.2 มาจาก 3 แหล่งดังต่อไปนี้ (Mortazavi-Asl. 2001 : 7)



รูปที่ 2.2 แหล่งที่มาของข้อมูลการใช้งานเว็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เฉพาะในวงจำกัดเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ล็อกระดับไคลเอนท์ (Client Level Log) จะใช้แอฟเฟล็ต จาวาสคริปต์ คุกกี้ หรือ เบราวี่เซอร์ที่มีการดัดแปร (Modified Browser) ในการเก็บข้อมูลการใช้งานเว็บทางฝั่งไคลเอนท์ ซึ่งต้องอาศัยความร่วมมือจากผู้ใช้
2. ล็อกระดับพร็อกซี (Proxy Level Log) จะเก็บหลายผู้ใช้จากหลายเว็บไซต์ โดยรับการร้องขอ HTTP จากผู้ใช้หลายคนเพื่อไปยังหลายๆเซิร์ฟเวอร์ ดังนั้น ล็อกไฟล์จะเก็บรายละเอียดของการร้องขอ HTTP จากไคลเอนท์หลายๆตัวไปหาเว็บเซิร์ฟเวอร์ต่างๆ จึงเหมาะสำหรับศึกษาพฤติกรรมของกลุ่มผู้ใช้ซึ่งใช้พร็อกซีเซิร์ฟเวอร์ตัวเดียวกัน
3. ล็อกระดับเซิร์ฟเวอร์ (Server Level Log) เป็นที่นิยมมากที่สุดในการทำเหมืองข้อมูลการใช้งานเว็บ โดยบันทึกการเข้ามาใช้งานเว็บของผู้ใช้ที่ฝั่งเซิร์ฟเวอร์

2.2.2 โครงสร้างข้อมูลของเว็บเซิร์ฟเวอร์ล็อก

ล็อกไฟล์ของการเข้าถึงเว็บเซิร์ฟเวอร์จะบันทึกกิจกรรมต่างๆที่เกิดขึ้นจากการที่ผู้ชมเข้าถึงไฟล์ที่อยู่บนเซิร์ฟเวอร์ NCSA ได้กำหนดมาตรฐานของรูปแบบล็อกไฟล์ไว้สองรูปแบบคือ แบบคอมมอนและแบบคอมไบน์หรือเอ็กซ์เทนดด์ ตัวอย่างที่ตามมานี้เป็นตัวอย่างของล็อกไฟล์แบบคอมมอน

```
125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0" 200 1043
```

ล็อกไฟล์แบบคอมไบน์หรือเอ็กซ์เทนดด์มีลักษณะเหมือนกับล็อกไฟล์แบบคอมมอน แต่มีฟิลด์ตำแหน่งที่ลิงก์มายังเว็บเพจและฟิลด์เบราเซอร์ แพลตฟอรม์ของผู้ชมเพิ่มขึ้นมาต่อท้าย ตัวอย่างที่ตามมานี้เป็นตัวอย่างของล็อกไฟล์แบบคอมไบน์

```
125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0":200 1043
"http://www.ibm.com/" "Mozilla/4.05 [en] (WinNT;I)"
```

ข้อมูลแต่ละฟิลด์ในล็อกไฟล์แบบคอมไบน์หรือเอ็กซ์เทนดด์ สามารถอธิบายดังตารางที่ 2.1

ตารางที่ 2.1 ฟิลด์ของล็อกไฟล์แบบคอมไบน์

ชื่อฟิลด์	คำอธิบาย
Remote host	ชื่อแม่ข่ายระยะไกล หรือไอพีแอดเดรส
Rfc931	ชื่อล็อกอินระยะไกลของไคลเอนท์
Auth. User	ชื่อที่ไคลเอนท์ใช้ในการพิสูจน์ตัวตนจริง
Date	วันและเวลาที่เยี่ยมชมเว็บไซต์
GMT Offset	จำนวนชั่วโมงแสดงระยะห่างจากเวลาสากล GMT
Method	วิธีดำเนินการ (ได้แก่ Get, Post, Head อื่นๆ) ต้องอยู่ในเครื่องหมายคำพูด
URI	เส้นทางของไฟล์ที่ถูกกระทำ
Protocol	โพรโทคอลที่ใช้/เวอร์ชัน
Status	โค้ดแสดงผลการตอบสนองของคำร้องขอ
Bytes	จำนวนไบต์ที่ส่ง
Referrer	ตำแหน่งที่ผู้ชมใช้ลิงก์มายังไซต์
Agent/Platform	เว็บเบราว์เซอร์และแพลตฟอร์มที่ใช้ในการเยี่ยมชมไซต์

2.2.3 นิยามคำศัพท์ที่ใช้ในการใช้งานเว็บ

คำศัพท์ที่ใช้ในการใช้งานเว็บอธิบายได้ดังตารางที่ 2.2

ตารางที่ 2.2 นิยามคำศัพท์ที่ใช้ในการใช้งานเว็บ

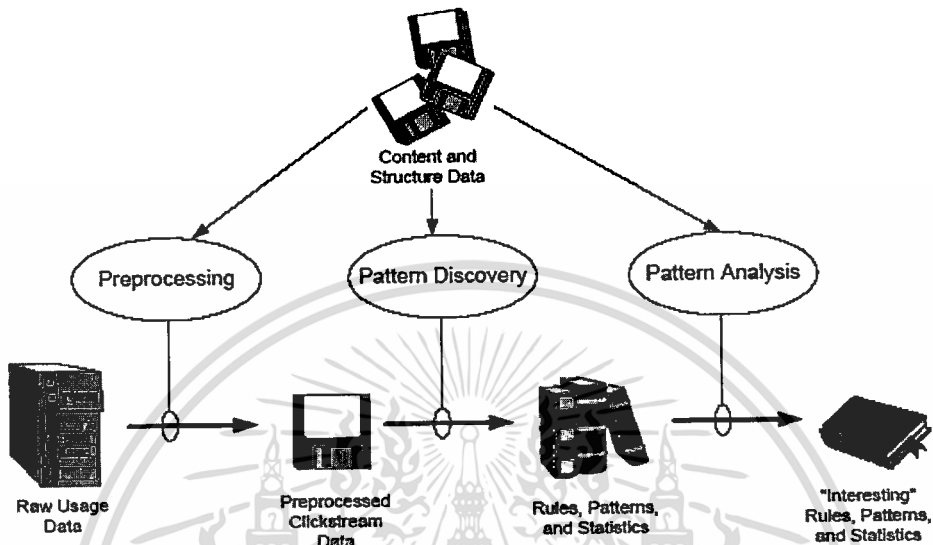
คำศัพท์	คำอธิบาย
Server	แอปพลิเคชันที่ให้บริการทรัพยากร
Proxy	แอปพลิเคชันที่อยู่ระหว่างเซิร์ฟเวอร์และไคลเอนท์
Client	แอปพลิเคชันที่ใช้ทรัพยากรจากเซิร์ฟเวอร์
User	ผู้ใช้งาน ไคลเอนท์แอปพลิเคชันเพื่อใช้ทรัพยากรจากเซิร์ฟเวอร์
User Session	เซตการคลิกของผู้ใช้ที่คลิกตั้งแต่หนึ่งหรือหลายเซิร์ฟเวอร์
Server session	เซตการคลิกของผู้ใช้ที่ถูกบันทึกโดยเซิร์ฟเวอร์เดียว
Episode	ลำดับเซตของการคลิกของผู้ใช้ที่เกี่ยวข้องกันใน User Session
Web page	กลุ่มของทรัพยากรที่มีการระบุโดย URI
Page View	เว็บเพจที่ไคลเอนท์แอปพลิเคชันเปิดขึ้นมา
Click Stream	เพจวิวที่เปิดโดยผู้ใช้อย่างเป็นไปตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเท่านั้น ไม่ควรนำออกเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บ

การทำเหมืองข้อมูลการใช้งานเว็บมีลักษณะที่คล้ายกับระเบียบวิธีในการทำเหมืองข้อมูลแสดงดังรูปที่ 2.3 ประกอบไปด้วย 3 ขั้นตอนคือ

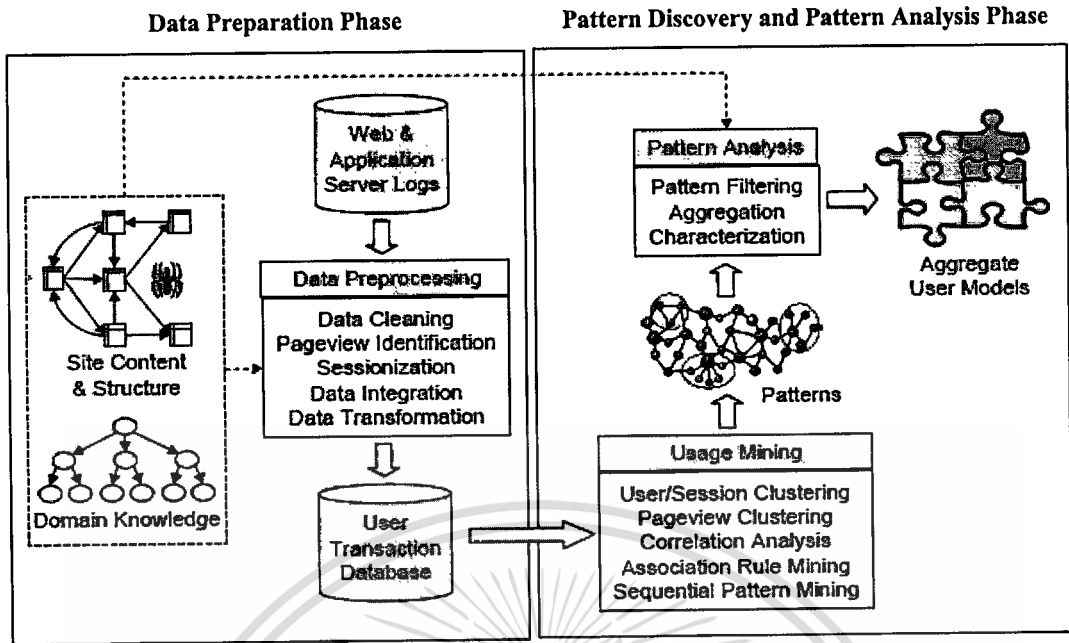


รูปที่ 2.3 ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บอย่างง่าย (Cooley, R. et al. 1999 : 6)

1. การเตรียมข้อมูลการใช้งาน (Usage Data Preparation หรือ Preprocessing) เป็นขั้นตอนเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับการทำเหมืองข้อมูลในขั้นตอนการค้นหาแบบ
2. การค้นหารูปแบบ (Pattern Discovery) จะนำข้อมูลที่ได้จากขั้นตอนการเตรียมข้อมูลการใช้งานมาค้นหารูปแบบโดยใช้เทคนิคทางเหมืองข้อมูล
3. การวิเคราะห์รูปแบบ (Pattern Analysis) เป็นขั้นตอนที่นำผลลัพธ์ที่ได้จากขั้นตอนการค้นหารูปแบบมาวิเคราะห์ แปรผล และนำไปประยุกต์ใช้

จากขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บทั้ง 3 ขั้นตอน สามารถอธิบายรายละเอียดได้ดังรูปที่ 2.4 โดยในขั้นตอนการเตรียมข้อมูลการใช้ประกอบไปด้วยขั้นตอนการแปลงข้อมูลการใช้งานเว็บ เนื้อหาและโครงสร้างของข้อมูลที่เกี่ยวข้องให้อยู่ในรูปแบบที่เหมาะสม เพื่อใช้การค้นหาแบบ โดยมีการจัดเก็บข้อมูลเหล่านี้ลงในฐานข้อมูล แล้วนำเข้าสู่ขั้นตอนการค้นหารูปแบบการใช้งาน ผลที่ได้จากขั้นตอนนี้คือรูปแบบ กฎ และสถิติ ต่างๆ เป็นต้น จากนั้นนำผลที่ได้นี้ไปวิเคราะห์ โดยในการวิเคราะห์จะเลือกเฉพาะรูปแบบ กฎ และสถิติที่น่าสนใจ แล้วรวบรวมสร้างเป็นโมเดลผู้ใช้

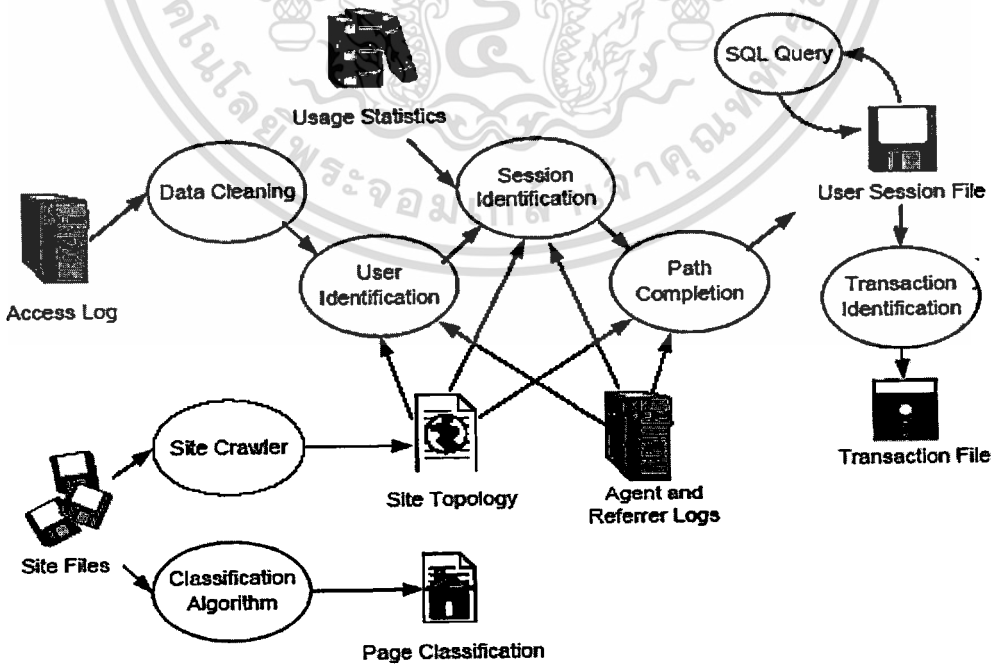
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 รายละเอียดขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บ

2.3.1 การเตรียมข้อมูลการใช้งาน

ขั้นตอนการเตรียมข้อมูลการใช้งานประกอบไปด้วยขั้นตอนการแปลงข้อมูลการใช้งานเว็บเนื้อหา และโครงสร้างของข้อมูลทีมาจากหลายๆ แหล่ง ให้เป็นข้อมูลที่เป็นประโยชน์ และใช้ในขั้นตอนการค้นหารูปแบบต่อไป การเตรียมข้อมูลการใช้งานประกอบด้วยขั้นตอนต่างๆ ดังรูปที่ 2.5



รูปที่ 2.5 รายละเอียดการเตรียมข้อมูลการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.1 การทำความสะอาดข้อมูล

การทำความสะอาดข้อมูล (Data Cleaning) หรือการกำจัดรายการที่ไม่เกี่ยวข้องออกจากเว็บเซิร์ฟเวอร์ล็อกไฟล์ โดยรายการที่ต้องกำจัดออกมีดังต่อไปนี้ (Lalani. 2003 : 14-16)

1. การกำจัดเว็บโรบอท ซึ่งเป็นตัวที่ทำหน้าที่เข้าสำรวจเว็บไซต์ต่างๆ ในเว็บเซิร์ฟเวอร์ล็อกไฟล์จะมีรายการที่เกิดขึ้นเนื่องจากการเข้าถึงของเว็บโรบอท รายการที่เว็บโรบอทเข้าไปคือ robot.txt ซึ่งในไฟล์นี้จะมีการจำกัดหรืออนุญาตให้เว็บโรบอทเข้ามาค้นคืนเอกสารได้ในระดับต่างๆ เช่น อนุญาตให้เว็บโรบอท สามารถตามลิงก์บางลิงก์ได้ หรือไม่สามารค้นคืนเอกสารบางอย่างได้ ภายในบางโคเร็กทอรีเว็บโรบอทจะถูกปฏิเสธด้วยข้อความที่ระบุใน robot.txt หรือ Embedded HTML Meta Tags ตัวอย่างของโรบอทมีดังนี้ Googlebot, Infoseek Robot 1.0, Infoseek Sidewinder, PerlCrawler 1.0, Scooter, SpiderBot, Site Searcher, Crawlers, Indexer, msnbot 1.0, Gigabot 2.0, Yahoo! Slurp, Exabot 2.0 (Zhu. 2001 : 16)

2. กรองภาพและข้อมูลรบกวน โดยเว็บเพจปกติจะมีรูปภาพ เสียงและวิดีโอไฟล์ไปพร้อมกับข้อมูล เว็บเซิร์ฟเวอร์จะบันทึกส่วนประกอบต่างๆของเว็บเพจซึ่งถูกร้องขอ โดยปกติส่วนประกอบต่างๆ อาทิเช่น รูปภาพ เสียงและวิดีโอ จะไม่ถูกนำมาใช้ในการวิเคราะห์การเข้าใช้งานเว็บไซต์ และการทำเหมืองข้อมูลการใช้งานเว็บไซต์ ซึ่งมีจุดประสงค์หลักในการเข้าใจพฤติกรรมของผู้ใช้ จึงไม่ต้องนำไฟล์ที่ไม่ได้ร้องขอจากผู้ใช้โดยตรงมาทำการวิเคราะห์ ดังนั้นการกำจัดรายการที่ไม่เกี่ยวข้องทำได้โดยตรวจสอบส่วนต่อท้าย (Suffix) ของ URL ที่ลงท้ายด้วย gif, jpeg, GIF, JPEG, jpg, JPG, map, cgi, WAV, wav, CLASS, class, TXT, txt, java script (JS), page style file (css) (Cooley, R. et al. 1999 : 12)

3. บางครั้งผู้ใช้อ้างอิงเว็บเพจที่ไม่มีอยู่ เว็บเซิร์ฟเวอร์จะสร้างรายการขึ้นมาพร้อมด้วยโค้ดที่เหมาะสม ซึ่งรายการล่อนี้จะไม่นำมาใช้ในการวิเคราะห์

4. เว็บเซิร์ฟเวอร์บันทึกการเข้าถึงไฟล์ proxy.pac ซึ่งก็คือ ไฟล์โครงแบบอัตโนมัติ (Auto Configuration) ที่ใช้ในการกำหนดโครงแบบเว็บเบราว์เซอร์ของไคลเอนท์ รายการนี้จะถูกกรองออกไปเพื่อไม่ให้มีผลต่อการค้นหารูปแบบ

เมื่อกำจัดรายการที่ไม่เกี่ยวข้องออกไปจากเว็บเซิร์ฟเวอร์ล็อกไฟล์แล้ว ขั้นตอนต่อไปจะระบุตัวผู้ใช้และแบ่งออกเป็นเซสชันต่างๆของผู้ใช้แต่ละคน

2.3.1.2 การระบุตัวผู้ใช้

การระบุตัวผู้ใช้ (User Identification) มีสมมุติฐานว่าผู้ใช้ทุกคนมีไอพีแอดเดรสเฉพาะที่แตกต่างกัน และใช้งานเบราว์เซอร์เพียงชนิดเดียว แต่จะมีบางกรณีที่ทำให้สมมุติฐานนี้ไม่เป็นจริง ซึ่งปัญหาที่เกิดขึ้นมีดังต่อไปนี้ (Srivastava, J. et al. 2000 : 14)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- หนึ่งไอพีแอดเดรส/หลายเซิร์ฟเวอร์เซสชัน (Single IP Address/Multiple Server Sessions) บางบริษัท บางหน่วยงาน บริษัทผู้ให้บริการอินเทอร์เน็ต มีการใช้หรือให้บริการเชื่อมต่ออินเทอร์เน็ตโดยอาศัยพรีอ็อกซีเซิร์ฟเวอร์ ทำให้ไอพีแอดเดรสของผู้เยี่ยมชมที่แสดงนั้นเป็น ไอพีแอดเดรสของพรีอ็อกซีเซิร์ฟเวอร์เดียวกันหมด ทั้งที่มีจำนวนผู้เยี่ยมชมอยู่หลายคนไม่ได้เก็บไอพีแอดเดรสที่แท้จริงของเครื่องผู้ใช้ แต่จะเก็บไอพีแอดเดรสของพรีอ็อกซีเซิร์ฟเวอร์

- หลายไอพีแอดเดรส/หนึ่งเซิร์ฟเวอร์เซสชัน (Multiple IP Address/Single Server Session) บางไอเอสพีหรือเครื่องมือทางด้านความปลอดภัยจะกำหนดแต่ละการร้องขอเว็บเพจมาจากหลายๆไอพีแอดเดรส ในกรณีนี้หนึ่งเซิร์ฟเวอร์เซสชันสามารถมีได้หลายไอพีแอดเดรส เช่น การทำ NAT (Network Address Translation)

- หลายไอพีแอดเดรส/ผู้ใช้งานเดียว (Multiple IP Address/Single User) ผู้ใช้คนเดียวกัน แต่เข้าเว็บไซต์ด้วยเครื่องคอมพิวเตอร์คนละเครื่องกัน ก็จะมีไอพีแอดเดรสที่ต่างกัน จึงเป็นการยากที่จะระบุได้ว่าเป็นผู้ใช้งานเดียวกัน

- หลายเอเจนต์/ผู้ใช้งานเดียว (Multiple Agent/Single User) ผู้ใช้คนเดียวกันใช้งานเว็บเบราว์เซอร์หลายตัวบนเครื่องเดียวกัน จะปรากฏเป็นผู้ใช้หลายคน

สมมุติฐานที่ว่าผู้ใช้ทุกคนมีไอพีแอดเดรสเฉพาะที่แตกต่างกัน และใช้งานเบราว์เซอร์เพียงชนิดเดียว ซึ่งไม่เป็นจริงข้างต้น เราสามารถแก้ปัญหาโดยอาศัยกลไกการระบุตัวผู้ใช้แบบต่างๆ ดังตารางที่ 2.3

ตัวอย่าง การระบุตัวผู้ใช้โดยใช้ล็อกไฟล์รูปที่ 2.6 จากรูปจะเห็นว่ารายการ (Entry) ทั้งหมดมี ไอพีแอดเดรสเดียวกัน และไม่มีการบันทึก User ID ใดๆก็ตาม รายการลำดับที่ 5, 6, 8 และ 10 เป็นการเข้าถึงโดยใช้เอเจนต์ที่แตกต่างจากในรายการอื่นๆ อาจบอกได้ว่าล็อกนี้เป็นการแสดงการเข้าใช้งานอย่างน้อย 2 เซสชันผู้ใช้

ตัว Heuristic ถัดไปสำหรับการระบุตัวผู้ใช้ที่ใช้ร่วมกันคือล็อกแสดงตำแหน่งที่ผู้ชมใช้ลิงก์มายังเว็บไซต์ (Referrer Log) และโครงสร้างเว็บไซต์เพื่อสร้างเส้นทางการเดินทางของผู้ใช้แต่ละคน ถ้าเพจที่ถูกร้องขอไม่สามารถเข้าถึงได้โดยตรงจากเพจอื่นที่ผู้ใช้เยี่ยมชม เราจะถือว่าผู้ใช้คนอื่นที่มีไอพีแอดเดรสเดียวกัน จากรูปที่ 2.6 รายการที่ 3 (เพจ L) ไม่ได้ถูกเข้าถึงจากเพจ A หรือเพจ B รายการลำดับที่ 7 (เพจ R) ถูกเข้าถึงจากเพจ L แต่ไม่ใช่จากรายการอื่นๆก่อนหน้า เราจึงสรุปว่ามีผู้ใช้ 3 คนแต่มีไอพีแอดเดรสเดียวกัน ดังนั้น หลังจากที่ทำการระบุตัวผู้ใช้แล้ว จะได้เส้นทางของผู้ใช้แต่ละคนดังนี้ A-B-F-O-G-A-D, A-B-C-J และ L-R ตามลำดับ (Cooley, R. et al. 1999 : I3)

#	IP Address	Userid	Time	Method/ URL/ Protocol	Status	Size	Referred	Agent
1	123.456.78.9	-	[25/Apr/1998:03:04:41 -0500]	"GET A.html HTTP/1.0"	200	3290	-	Mozilla/3.04 (Win95, I)
2	123.456.78.9	-	[25/Apr/1998:03:05:34 -0500]	"GET B.html HTTP/1.0"	200	2050	A.html	Mozilla/3.04 (Win95, I)
3	123.456.78.9	-	[25/Apr/1998:03:05:39 -0500]	"GET L.html HTTP/1.0"	200	4130	-	Mozilla/3.04 (Win95, I)
4	123.456.78.9	-	[25/Apr/1998:03:06:02 -0500]	"GET F.html HTTP/1.0"	200	5096	B.html	Mozilla/3.04 (Win95, I)
5	123.456.78.9	-	[25/Apr/1998:03:06:58 -0500]	"GET A.html HTTP/1.0"	200	3290	-	Mozilla/3.01 (X11, I, IRIX6.2, IP22)
6	123.456.78.9	-	[25/Apr/1998:03:07:42 -0500]	"GET B.html HTTP/1.0"	200	2050	A.html	Mozilla/3.01 (X11, I, IRIX6.2, IP22)
7	123.456.78.9	-	[25/Apr/1998:03:07:55 -0500]	"GET R.html HTTP/1.0"	200	8140	L.html	Mozilla/3.04 (Win95, I)
8	123.456.78.9	-	[25/Apr/1998:03:09:50 -0500]	"GET C.html HTTP/1.0"	200	1820	A.html	Mozilla/3.01 (X11, I, IRIX6.2, IP22)
9	123.456.78.9	-	[25/Apr/1998:03:10:02 -0500]	"GET O.html HTTP/1.0"	200	2270	F.html	Mozilla/3.04 (Win95, I)
10	123.456.78.9	-	[25/Apr/1998:03:10:45 -0500]	"GET J.html HTTP/1.0"	200	8430	C.html	Mozilla/3.01 (X11, I, IRIX6.2, IP22)
11	123.456.78.9	-	[25/Apr/1998:03:12:23 -0500]	"GET G.html HTTP/1.0"	200	7220	B.html	Mozilla/3.04 (Win95, I)
12	123.456.78.9	-	[25/Apr/1998:05:05:22 -0500]	"GET A.html HTTP/1.0"	200	3290	-	Mozilla/3.04 (Win95, I)
13	123.456.78.9	-	[25/Apr/1998:05:06:03 -0500]	"GET D.html HTTP/1.0"	200	1680	A.html	Mozilla/3.04 (Win95, I)

รูปที่ 2.6 ตัวอย่างเว็บเซิร์ฟเวอร์ล็อกไฟล์

ตารางที่ 2.3 กลไกการระบุตัวผู้ใช้ (Berendt, B. et al. 2002 : 12)

วิธีการ	การอธิบาย	ความเป็นส่วนตัว	ข้อดี	ข้อเสีย
ไอพีแอดเดรส + เอเจนต์	กำหนดไอพีแอดเดรสและ เอเจนต์เป็นคู่ในการระบุผู้ใช้แต่ละคน	ต่ำ	มีอยู่สามารใช้ได้ประโยชน์ได้ไม่ต้องใช้เทคโนโลยีเพิ่มเติม	ไม่รับประกันความเป็นหนึ่งจะล้มเหลวเมื่อมีการสับเปลี่ยนหมุนเวียนไอพี
เซสชันไอดี	ใช้เซสชันไอดีที่เก็บทางฝั่งเซิร์ฟเวอร์	ต่ำถึงกลาง	มีอยู่สามารใช้ได้ประโยชน์ได้ไม่ขึ้นกับไอพีแอดเดรส	ไม่สามารถตรวจจับการเข้ามาใช้งานซ้ำและมีการใช้เวลาเพิ่มเติมสำหรับไดนามิกเพจ
การลงทะเบียน	ผู้ใช้เข้าสู่ระบบด้วยการล็อกอิน หลังจากที่ได้ลงทะเบียน	กลาง	สามารถติดตามการใช้งานในแต่ละบุคคลเมื่อเป็นผู้ใช้คนเดียวกันแต่ใช้เว็บเบราว์เซอร์หลายอัน	คนส่วนใหญ่จะไม่ลงทะเบียน ทำให้ไม่สามารถใช้งานได้ก่อนการลงทะเบียน
คุกกี้	บันทึกไอดีบนเครื่องไคลเอนต์	กลางถึงสูง	สามารถติดตามการเข้าใช้ซ้ำจากเบราว์เซอร์ตัวเดียวกัน	สามารถถูกยกเลิกการใช้คุกกี้โดยผู้ใช้
ซอฟต์แวร์เอเจนต์	โปรแกรมที่โหลดลงในเบราว์เซอร์และส่งข้อมูลการใช้เว็บกลับมา	สูง	มีความถูกต้องในการใช้งานสำหรับเว็บไซต์เดียว	มีแนวโน้มที่จะถูกปฏิเสธ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.3 การระบุเซสชัน

ในการระบุเซสชันของผู้ใช้ (Session Identification) แต่ละคนมีการจับเวลาออก (Time Out) ถ้าคาบเวลาระหว่าง 2 เเพจที่เยี่ยมชมต่อเนื่องกันมีค่ามากกว่าค่าที่กำหนดไว้ 2 เซสชันนั้น จะถูกพิจารณาเป็นการสิ้นสุดและเริ่มต้นของเซสชันตามลำดับ ระยะเวลาออกที่นิยมใช้คือ 30 นาที การระบุเซสชันผู้ใช้ จะใช้เวลาออกเป็นตัวแบ่งแต่ละเซสชันของผู้ใช้ออกจากคลิกสตรีม (Click Stream) ที่ปรากฏอยู่ในเว็บเซิร์ฟเวอร์ล็อกไฟล์ คลิกสตรีมหนึ่งสามารถพิจารณาออกได้เป็นหนึ่งเซสชันหรือเซตของหลายเซสชันที่ถูกเข้าถึง โดยผู้ใช้จากเครื่องเดียวกันในหน้าต่างเวลาที่กำหนด (Defined Time Window)

เซสชันที่ถูกระบุด้วยกลุ่มของเพจที่มีการร้องขอโดยผู้ใช้คนเดียวกัน สามารถเก็บเซตของเซสชันอยู่ในรูปของ (session-id, {page-id, time}) โดย session-id เป็นเลขระบุเฉพาะ (Unique ID) ที่กำหนดให้กับเซสชัน เช่น (s-id, p0, 20, p1, 30, p2, 58) บอกว่าผู้ใช้คนหนึ่งใช้เวลา 20 นาทีบนเพจ p0 30 นาที บนเพจ p1 และ 58 นาทีบนเพจ p2

ตัวอย่างจากรูปที่ 2.6 เส้นทางของผู้ใช้คนที่ 1 คือ A-B-F-O-G-A-D ถ้ากำหนดค่าเวลาออกเป็น 30 นาที จะแตกออกมาเป็น 2 เซสชันผู้ใช้ เนื่องจาก 2 เเพจสุดท้ายใช้เวลาเกินกว่าชั่วโมง จากการอ้างอิง 5 อันแรก ดังนั้น ขั้นตอนนี้จะได้เซสชันออกมาเป็น A-B-F-O-G, A-D, A-B-C-J และ L-R

2.3.1.4 การทำเส้นทางให้สมบูรณ์

เว็บเบราว์เซอร์ส่วนมากจะเก็บเพจที่มีการร้องขอแล้ว เป็นผลให้เมื่อผู้ใช้กดปุ่มย้อนกลับ (Back) หรือมีการร้องขอเพจเดิม เพจที่ถูกเก็บ (Cached Page) นั้นจะถูกนำมาแสดง โดยไม่ต้องส่งการร้องขอไปที่เว็บเบราว์เซอร์อีก และพร็อกซีเซิร์ฟเวอร์ก็จะเก็บเพจที่มีการเยี่ยมชมบ่อยๆ เพื่อลดการจราจรในเครือข่ายและเพิ่มประสิทธิภาพของเซิร์ฟเวอร์ ดังนั้น จึงมีแนวคิดของการทำเส้นทางให้สมบูรณ์ (Path Completion) เพิ่มเพจที่ขาดหายไปเนื่องจากการใช้แคชของเบราว์เซอร์และเว็บพร็อกซี

การทำเส้นทางให้สมบูรณ์ (Cooley, R. et al. 1999 : 15) เป็นกระบวนการระบุเพจที่หายไปในการเข้าใช้งานเว็บไซต์ ซึ่งไม่ได้บันทึกลงในเซิร์ฟเวอร์ล็อกไฟล์ มีการแนะนำให้ใช้โครงสร้างเว็บไซต์ (Site Topology) และล็อกแสดงตำแหน่งที่ผู้ชมใช้ลิงก์มายังเว็บไซต์ เพื่อช่วยในการทำเส้นทางให้สมบูรณ์ หรือทำ Cache Busting ซึ่งใช้ในการขจัดวางเว็บเบราว์เซอร์จากการใช้เพจที่มีการเก็บในแคช บังคับให้ต้องไปดาวน์โหลดเพจใหม่จากเว็บเซิร์ฟเวอร์ทุกครั้งที่มีการร้องขอ แต่ Cache Busting นั้นจะทำให้ประสิทธิภาพลดลง ตัวอย่างจากรูปที่ 2.6 รายการที่ 11 (เพจ G) ไม่ได้เข้าถึงโดยตรงจากเพจ O ล็อกแสดงตำแหน่งที่ผู้ชมใช้ลิงก์มายังเว็บไซต์สำหรับเพจ G ได้ร้องขอเพจ B จึงบอกได้ว่าผู้ใช้คนที่ 1 มีการย้อนกลับมายังเพจ B โดยใช้ปุ่มย้อนกลับก่อนที่จะมาร้องขอเพจ G

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น เพจ F และ B ควรที่จะถูกเพิ่มลงไปในเซสชันของผู้ใช้คนที่ 1 มันอาจเป็นไปได้ว่าผู้ใช้รู้ URL แล้วอาจพิมพ์ URL เข้าไปโดยตรง ซึ่งมันก็ไม่น่าจะเกิดขึ้น และไม่น่าจะเกิดบ่อยมากพอที่จะกระทบต่ออัลกอริทึมทางการทำเหมืองข้อมูล เมื่อเสร็จสิ้นขั้นตอนการทำเส้นทางให้สมบูรณ์จะได้เส้นทางดังนี้ A-B-F-O-F-B-G, A-D, A-B-A-C-J และ L-R ผลลัพธ์จากการทำกระบวนการเตรียมข้อมูลสรุปดังตารางที่ 2.4

ตารางที่ 2.4 สรุปผลลัพธ์จากการทำกระบวนการเตรียมข้อมูล

Task	Result
Clean Log	<ul style="list-style-type: none"> • A-B-L-F-A-B-R-C-O-J-G-A-D
User Identification	<ul style="list-style-type: none"> • A-B-F-O-G-A-D • A-B-C-J • L-R
Session Identification	<ul style="list-style-type: none"> • A-B-F-O-G • A-D • A-B-C-J • L-R
Path Completion	<ul style="list-style-type: none"> • A-B-F-O-F-B-G • A-D • A-B-A-C-J • L-R

2.3.1.5 การจัดรูปแบบ

ขั้นตอนสุดท้ายของการเตรียมข้อมูลการใช้งานคือ การระบุแอตทริบิวต์ที่เกี่ยวข้องและลดมิติของข้อมูลลง โดยการกำจัดรายการแอตทริบิวต์ที่ไม่เกี่ยวข้องของเซสชันหรือทรานแซคชันออก เพื่อให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมกับการค้นหาในรูปแบบ เรียกว่าการจัดคุณลักษณะ (Feature Formatting) ของข้อมูล ซึ่งต้องจัดให้เหมาะสมกับอัลกอริทึมสำหรับการทำเหมืองข้อมูลที่จะใช้

2.3.1.6 การระบุทรานแซคชัน

ทรานแซคชันเป็นเซตย่อยหรือสับเซตของเซสชันผู้ใช้ เมื่อระบุเซสชันผู้ใช้แล้ว จะมีการแยกเซสชันผู้ใช้ออกเป็นสับเซต ที่เรียกว่า Episode การระบุทรานแซคชัน (Episode หรือ Transaction Identification) มี 3 วิธีดังนี้

1. ความยาวของการอ้างอิง (Reference Length) เป็นการระบุทรานแซคชันโดยใช้วิธีการอ้างอิงที่อยู่บนพื้นฐานของสมมติฐานที่ว่า ระยะเวลาเฉลี่ยที่ผู้ใช้ใช้ในการดูเว็บเพจซึ่งมีความสัมพันธ์กับเพจซึ่งถูกจำแนกออกเป็นเพจช่วย (Auxiliary Page) ซึ่งจะใส่เข้าไปยังเพจเนื้อหาหรือเพจเนื้อหา (Content Page) โดยมีการคาดว่า ถ้าเป็นเพจช่วยผู้ใช้จะใช้เวลาบนหน้านั้นสั้น แต่ถ้าเป็นเพจเนื้อหาผู้ใช้จะใช้เวลาบนหน้านั้นนาน ด้วยสมมติฐานที่ว่าผู้ใช้จะใช้เวลาบนเพจทั้งสองประเภทต่างกัน

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์สำหรับการใช้ในแวดวงวิชาการเท่านั้น เมื่ออนุญาตเห็นาเบ้ขอประยิขณต่านการค้

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แนวความคิดพื้นฐานในการประมาณการแจกแจงความยาวของการอ้างอิงของเพจทั้งหมด ด้วยการแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) โดยจุดตัด (Cut-off Point) t สำหรับความยาวของการอ้างอิงกำหนดได้ดังนี้

$$t = \frac{-\ln(1-\gamma)}{\lambda} \quad (2.1)$$

โดยที่ $\gamma = \%$ ของการอ้างอิงเพจช่วย

$\lambda =$ ส่วนกลับของค่าเฉลี่ยของความยาวของการอ้างอิง

2. การอ้างอิงไปข้างหน้าไกลที่สุด (Maximum Forward Reference) เป็นการกำหนดเขตของเพจวิที่จะสิ้นสุด เมื่อใดก็ตามที่มีการอ้างอิงย้อนหลัง (Backward Reference) ไปยังเพจที่เยี่ยมชมผ่านมาแล้ว การอ้างอิงไปข้างหน้าไกลที่สุดนั้นจะสิ้นสุดทันที และจะเก็บลงฐานข้อมูลและเริ่มใหม่เนื่องจากขณะเยี่ยมชมนั้นผู้ใช้นิยมย้อนกลับไปยังเพจก่อนหน้าที่ได้เยี่ยมชมไปแล้ว จึงใช้การอ้างอิงไปข้างหน้าไกลที่สุด เพื่อใช้ในการแตกเซชันผู้ใช้ออกเป็นทรานแซกชันย่อยๆ สำหรับการทำเหมืองข้อมูลการเข้าใช้งานเว็บไซต์

3. หน้าต่างเวลา (Time Window) ใช้ระบุทรานแซกชันจากเซชันผู้ใช้ โดยแบ่งตามหน้าต่างเวลา ซึ่งต้องไม่มากกว่าพารามิเตอร์ที่กำหนดไว้ เช่น เพจ I_m และ I_l อยู่ในเซชันเดียวกัน ถ้า $I_m.time - I_l.time \leq W$ โดย W คือ ความยาวของหน้าต่างเวลาเราสามารถประยุกต์ใช้หน้าต่างเวลากับวิธีการแบ่งแยกอื่นๆ ได้

2.3.1.7 การระบุลำดับผู้ใช้

มีการระบุลำดับผู้ใช้เพื่อในการค้นหารูปแบบการใช้งานเว็บ โดยจะนำทรานแซกชันที่ได้จากการระบุทรานแซกชันมาสร้างเป็นลำดับผู้ใช้ รายละเอียดเรื่องลำดับผู้ใช้ศึกษาได้จากหัวข้อที่ 2.5

2.3.2 การค้นหารูปแบบ

เฟสการค้นหารูปแบบ (Pattern Discovery) เป็นการค้นหารูปแบบการใช้งานเว็บ มีเทคนิคดังนี้ (Cooley, R. et al. 1997 ; Srivastava, J. et al. 2000)

เทคนิคทางสถิติ (Statistic Analysis) เป็นวิธีที่ใช้วิเคราะห์ค่าทางสถิติ เช่น ค่าความถี่ ฐานนิยม มัชยฐาน เป็นต้น ตัวอย่างการใช้เทคนิคทางสถิติ เช่น ใช้หาเพจที่มีการเข้าถึงบ่อยที่สุด เวลาโดยเฉลี่ยที่ใช้ต่อเว็บเพจหน้าใดๆ

1. กฎความสัมพันธ์ (Association Rule) ใช้อธิบายความสัมพันธ์ระหว่างเพจต่างๆ ที่มักจะมีการใช้งานร่วมกันในหนึ่งเซิร์ฟเวอร์เซชัน กฎความสัมพันธ์หาได้จากทรานแซกชันของผู้ชมโดย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่แต่ละทรานแซกชันประกอบไปด้วยเซตของ URL ที่ผู้ชมคนหนึ่งเข้าชมเว็บไซต์แต่ละครั้ง ตัวอย่างเช่น

- 40% ของโคลเอนท์ที่เข้าถึงเว็บเพจที่มี URL เป็น /company/products/ product1.html จะเข้าถึงเว็บเพจ /company/products/product2.html ด้วย
- 30% ของโคลเอนท์ที่เข้าถึงเว็บเพจที่มี URL เป็น /company/products/ product1.html จะเข้าถึงเว็บเพจ /company/products/product5.html ด้วย

2. เทคนิคในการรวมกลุ่ม (Clustering) โดยรวมสิ่งที่มีคุณลักษณะคล้ายคลึงกันเข้าด้วยกัน เมื่อเราจัดกลุ่มลูกค้าได้ เราก็จะให้บริการได้ตรงความต้องการทำให้เกิดความพึงพอใจ และซื่อสัตย์ต่อบริษัท ในการทำเหมืองข้อมูลการใช้งานเว็บไซต์ มีการรวมกลุ่ม (Clustering) อยู่ 2 แบบคือ

- การจัดกลุ่มการใช้ (Usage Clustering) จัดกลุ่มที่มีรูปแบบการใช้งานที่คล้ายกัน
- การจัดกลุ่มของเว็บเพจ (Page Clustering) จัดกลุ่มที่มีเนื้อหาเกี่ยวข้องกัน

3. การจำแนก (Classification) รายการข้อมูล ให้อยู่ในคลาสที่กำหนดไว้ อัลกอริทึมที่ใช้ เช่น Decision Tree, Naïve Bayesian Classifier, K-Nearest Neighbor Classifier, Support Vector Machine ตัวอย่างเช่น

- 50% ของผู้ที่ทำรายการสั่งซื้อใน /company/product2 เป็นผู้ที่มียุทธระหว่าง 20 – 25 ปี และอาศัยอยู่ในชายฝั่งตะวันตก

4. รูปแบบลำดับ (Sequential Pattern) การเยี่ยมชมเว็บเพจที่จัดเรียงตามเวลาที่ระบุในทรานแซกชัน โดยแต่ละทรานแซกชันจะประกอบไปด้วย URL และเวลาการเยี่ยมชมแต่ละครั้ง ตัวอย่างเช่น

- 60% ของลูกค้าที่สั่งซื้อสินค้าใน /company/products/product1.html แล้วยังทำการสั่งซื้อสินค้าใน /company/products/product4.html ภายใน 15 วันด้วย
- 30% ของลูกค้าซึ่งเข้าชมเว็บที่มี URL /product/software/ ได้ค้นคว้า software ที่ yahoo มาก่อน

5. โมเดลการขึ้นต่อกัน (Dependency Model) เป็น โมเดลเพื่อใช้ค้นหาพฤติกรรมของผู้ใช้ เทคนิคที่ใช้ เช่น Hidden Markov Model และ Bayesian Belief Networks

2.3.3 การวิเคราะห์รูปแบบ

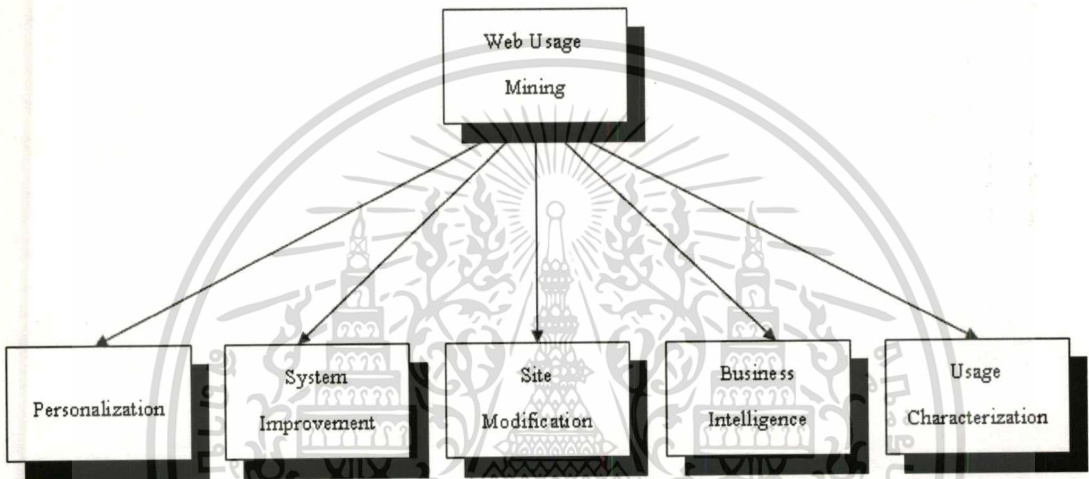
ขั้นตอนสุดท้ายของรูปแบบลำดับที่ค้นพบในเฟสการค้นหารูปแบบ แล้วนำมาวิเคราะห์เพื่อนำไปประยุกต์ใช้งาน ตัวอย่างในการวิเคราะห์รูปแบบ (Pattern Analysis) เช่น การไหลของข้อมูลไปเป็นดาต้าคิวบ์ เพื่อทำ OLAP (Online Analytical Processing) ซึ่งจะช่วยให้สามารถแสดงมุมมองในหลายมิติของข้อมูล โดยจะมีการหมุนแกนเปลี่ยนแกน (Rotate/Pivot) และการเจาะลึกข้อมูล (Drill Up/Drill

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Down) วิเคราะห์และสรุปความหมายของผลลัพธ์ที่ได้ ซึ่งเป็นองค์ความรู้นำไปเป็นสารสนเทศที่ช่วยในการตัดสินใจ

2.4 การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ

การทำเหมืองข้อมูลการใช้งานเว็บสามารถนำมาประยุกต์ใช้ในการทำให้เป็นส่วนบุคคล การปรับปรุงระบบ การปรับปรุงโครงสร้างเว็บไซต์ ระบบธุรกิจอัจฉริยะ และจำแนกลักษณะการใช้งาน แสดงดังรูปที่ 2.7



รูปที่ 2.7 การประยุกต์ใช้การทำเหมืองข้อมูลการใช้งานเว็บ

2.4.1 การทำให้เป็นส่วนบุคคล

การทำให้เป็นส่วนบุคคล (Personalization) เป็นการวิเคราะห์พฤติกรรมการใช้งานเว็บของผู้ใช้ เพื่อที่จะนำเสนอหรือแนะนำสิ่งที่ผู้เข้ามาใช้งานเว็บสนใจ เช่น

- วิเคราะห์บันทึกการเข้าใช้ เพื่อช่วยในการโฆษณาสินค้าหรือบริการต่างๆ ให้แก่ลูกค้า
- เพิ่มความเข้าใจอย่างลึกซึ้งและสร้างความสัมพันธ์ให้แน่นแฟ้นยิ่งขึ้น โดยใช้วิธี Cross-Sell/Up-Sell Products
- เสริมสร้างการทำการตลาดอย่างต่อเนื่องเพื่อสร้างความประทับใจให้เป็นลูกค้าประจำ
- สามารถวัดพฤติกรรมของลูกค้าและติดตามดูการตอบสนองต่อข้อเสนอทางการตลาดของลูกค้า

การทำให้เป็นส่วนบุคคลมีวิธีมาตรฐานที่ใช้ทำการระบุตัวบุคคลที่ใช้งานเว็บดังนี้

1. Rule-Based Filtering เป็นการจัดเนื้อหาให้กับผู้ใช้งานพื้นฐานของกฎที่ได้ระบุไว้ เช่น ถ้าผู้ใช้คลิกบน A และผู้ใช้รหัสไปรษณีย์เป็น 90210 ให้เพิ่มลิงก์ไปที่ C
 2. Collaborative Filtering เป็นการให้คำแนะนำแก่ผู้ใช้งานพื้นฐานของการตอบสนองและจัดอันดับของผู้ใช้คนอื่นที่มีความคล้ายคลึงกัน เป็นวิธีที่แพร่หลายและประสบความสำเร็จ แต่จะมีปัญหาเมื่อมีรายการและผู้ใช้มากๆ
 3. Content-Based Filtering จะติดตามเว็บเพจที่ผู้ใช้เยี่ยมชมและให้คำแนะนำเว็บเพจอื่นที่มีเนื้อหาคล้ายคลึงกัน แต่จะขาดความสัมพันธ์ระหว่างอ็อบเจกต์ที่ไม่ได้อยู่บนเนื้อหาที่คล้ายคลึงกัน
- Hybrid Methods วิธีนี้เป็นการผสมทั้งวิธีการ Content-Based และ Collaborative Filtering

2.4.2 การปรับปรุงระบบ

การปรับปรุงระบบช่วยให้เข้าใจการจราจรที่เกิดขึ้นบนเว็บไซต์ ซึ่งจะถูกใช้ในการสร้างนโยบายสำหรับการทำเว็บแคม การจัดการทางด้านเน็ตเวิร์ค การแบ่งงานที่เข้ามาให้กระจายไปหลายๆเครื่อง และช่วยในการตรวจสอบการบุกรุก

2.4.3 การปรับปรุงโครงสร้างเว็บไซต์

เมื่อศึกษาการใช้พฤติกรรมการใช้งาน ถ้าโครงสร้างของเว็บมีการจัดการที่ไม่ดี จะทำให้การเข้าถึงเว็บเพจต่างๆเป็นไปได้ยาก การทำเหมืองข้อมูลการใช้งานเว็บจะช่วยในการตัดสินใจว่าเว็บเพจใดควรจะลิงก์กันโดยตรงหรืออยู่ในกลุ่มเดียวกัน จากนั้นทำการออกแบบโครงสร้างเว็บไซต์ใหม่ เพื่อให้ง่ายต่อการเข้าถึงข้อมูล

2.4.4 ระบบธุรกิจอัจฉริยะ

องค์ความรู้ที่ได้จากการทำเหมืองข้อมูลการใช้งานเว็บจะช่วยในด้านการตัดสินใจและเพิ่มความสามารถทางด้านธุรกิจออนไลน์ และเป็นการรวบรวมข้อมูลข่าวสารด้านตลาด ข้อมูลลูกค้า และคู่แข่งกัน

2.4.5 จำแนกลักษณะการใช้งาน

ศึกษาการเยี่ยมชมเว็บเพจภายในเว็บไซต์แยกตามลักษณะการใช้งานของผู้ใช้ (Web Usage Characterization) เช่น กลยุทธ์ในการคลิก สถิติเกี่ยวกับการคลิกปุ่มไปข้างหน้าและย้อนกลับ (Forward / Back) การบันทึกเว็บเพจ การค้นหา เป็นต้น

2.5 รูปแบบลำดับ

2.5.1 ที่มาของปัญหา

การทำเหมืองข้อมูลรูปแบบลำดับ (Mining Sequential Pattern) เกิดจากการที่ประสบปัญหาในการสนับสนุนการตัดสินใจที่เผชิญอย่างมากในองค์กรที่ทำการค้าปลีกขนาดใหญ่ ความก้าวหน้าทางเทคโนโลยีบาร์โค้ดทำให้องค์กรสามารถที่จะเก็บรวบรวมข้อมูลทางการขายขนาดใหญ่ได้ และอ้างอิงเป็นข้อมูลตระกร้า (Basket Data) ภายในระยะเวลาประกอบด้วยวันที่ เวลาและรายการที่ซื้อในทรานแซกชัน บ่อยครั้งที่ข้อมูลเหล่านี้จะเก็บรหัสลูกค้าด้วย เมื่อมีการซื้อโดยใช้บัตรเครดิตหรือบัตรลูกค้าประจำ โดยบันทึกในบัญชีรายชื่อยังเก็บข้อมูลการซื้อโดยลำดับ ตัวอย่างของการทำเหมืองข้อมูลรูปแบบลำดับ เช่น รูปแบบการเช่าหนัง โดยลูกค้ามักจะเช่าเรื่อง “Star Wars” จากนั้นตามด้วยการเช่าเรื่อง “Empire Strike Back” และตามด้วย “Return of the Jedi” การเช่านั้นไม่จำเป็นที่จะต้องต่อเนื่องกันตามลำดับ และจำนวนสมาชิกในรูปแบบลำดับไม่จำเป็นต้องเป็นรายการเดียว เช่น การซื้อผ้าปูที่นอนและปลอกหมอน ต่อมาจะซื้อผ้าขนวม จากนั้นตามด้วยผ้าม่านจีบและผ้าระบายตกแต่ง (Agrawal, R. et al. 1995)

นิยาม 2.1 D เป็นเซตของทรานแซกชัน

กำหนดให้ D เป็นเซตของทรานแซกชัน แต่ละทรานแซกชันประกอบด้วยฟิลด์ดังต่อไปนี้ รหัสลูกค้า (customer-id) เวลาในทรานแซกชัน (transaction-time) และรายการ (items) ที่ซื้อในทรานแซกชัน ไม่มีลูกค้าคนใดที่จะมีทรานแซกชันมากกว่าหนึ่งทรานแซกชันในเวลาเดียวกัน

นิยาม 2.2 เซตของรายการ (Itemset)

กำหนดให้เซตของรายการ I เป็น $\{i_1, i_2, \dots, i_m\}$ โดยที่ i_j เป็นรายการหนึ่ง ซึ่งเซตของรายการนี้จะถูกทำให้เป็นเซตของตัวเลขจำนวนเต็ม

นิยาม 2.3 ลำดับ (Sequence) คือ ลิสต์ของเซตของรายการ

กำหนดให้ ลำดับ (Sequence) คือ ลิสต์ของเซตของรายการที่เรียงตามลำดับ กำหนดลำดับ S เป็น $\langle s_1, s_2, \dots, s_m \rangle$ โดยที่ s_j คือ เซตของรายการ กำหนดให้ความยาว (Length) ของลำดับเป็นจำนวนของรายการที่อยู่ในลำดับ โดยที่ลำดับที่มีความยาวเป็น k จะเรียกว่า k-sequence ลำดับนั้นสร้างขึ้นมาจากการนำ 2 ลำดับ x และ y มาเชื่อมต่อกันกำหนดเป็น x.y

นิยาม 2.4 ลำดับ $\langle a_1, a_2, \dots, a_n \rangle$ เป็นลำดับย่อยของ $\langle b_1, b_2, \dots, b_m \rangle$ ได้ ถ้ามีตัวเลขจำนวนเต็ม $i_1 < i_2 < \dots < i_n$ โดยที่ $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, a_n \subseteq b_{i_n}$ เราใช้สัญลักษณ์ “ \prec ” เพื่อแสดงถึงความสัมพันธ์แบบ “is contained”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างเช่น ลำดับ $\langle(3)(4\ 5)(8)\rangle \prec \langle(7)(3\ 8)(9)(4\ 5\ 6)(8)\rangle$ เนื่องจาก $(3) \subseteq (3\ 8)$, $(4\ 5) \subseteq (4\ 5\ 6)$ และ $(8) \subseteq (8)$ อย่างไรก็ตามลำดับ $\langle(3)(5)\rangle$ ไม่ได้อยู่ใน $\langle(3\ 5)\rangle$ และในทางกลับกัน $\langle(3\ 5)\rangle$ ก็ไม่ได้อยู่ใน $\langle(3)(5)\rangle$ โดยที่ $\langle(3\ 5)\rangle$ แสดงว่าเราซื้อรายการที่ 3 และ 5 ไปพร้อมกัน แต่ $\langle(3)(5)\rangle$ จะแสดงว่าเราซื้อรายการที่ 3 ก่อนแล้วซื้อรายการที่ 5

นิยาม 2.5 ลำดับลูกค้า (Customer Sequence)

กำหนดให้ลำดับลูกค้า คือ ทรานแซกชันทั้งหมดของลูกค้าคนหนึ่งซึ่งจัดเรียงเป็นลำดับ โดยแต่ละทรานแซกชันคือเซตของรายการและลำดับของทรานแซกชันถูกเรียงตามเวลาในทรานแซกชัน เป็น T_1, T_2, \dots, T_n และกำหนดให้เซตของรายการใน T_i เป็น $\text{itemset}(T_i)$ แล้ว จะได้ว่าลำดับของลูกค้าคือลำดับ $\langle \text{itemset}(T_1) \text{itemset}(T_2) \dots \text{itemset}(T_n) \rangle$

นิยาม 2.6 ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence)

รูปแบบลำดับเป็นลำดับที่อยู่ในรูปของ $X \xrightarrow{S,C} Y$ โดยที่ X และ Y เป็นเซตของรายการ

- ค่าสนับสนุนของรูปแบบลำดับเท่ากับ S ก็ต่อเมื่อ อัตราส่วนระหว่างจำนวนทรานแซกชันที่มี $X.Y$ เป็นส่วนประกอบ กับ จำนวนทรานแซกชันทั้งหมด
- ค่าความเชื่อมั่นของรูปแบบลำดับเท่ากับ C ก็ต่อเมื่อ อัตราส่วนระหว่างจำนวนทรานแซกชันที่มี $X.Y$ เป็นส่วนประกอบ กับ จำนวนทรานแซกชันที่มี X เป็นส่วนประกอบ

การค้นหารูปแบบลำดับจะพิจารณาค่าทางสถิติ 2 ค่าคือ

1. ค่าสนับสนุนขั้นต่ำ (Minimum Support) ซึ่งค่านี้จะเป็นตัวกำหนดว่ารูปแบบลำดับที่ได้จะต้องมาจากเซตของรายการที่มีการเกิดอย่างน้อยเท่ากับค่าสนับสนุนขั้นต่ำ เซตของรายการที่ผ่านค่าสนับสนุนขั้นต่ำตามที่กำหนด จะเรียกว่าเป็นเซตของรายการขนาดใหญ่ (Large Itemset) หรือ litemset

2. ค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) ซึ่งค่านี้จะเป็นตัวกำหนดว่ารูปแบบลำดับที่ได้มีความน่าเชื่อถือและเป็นจริงขั้นต่ำอย่างน้อยเท่ากับค่าความเชื่อมั่นขั้นต่ำ

ลำดับซึ่งมีค่าสนับสนุนไม่น้อยกว่าค่าสนับสนุนขั้นต่ำนี้เรียกว่า ลำดับขนาดใหญ่ (Large Sequence) ซึ่งจะเป็นลิสต์ของเซตของรายการขนาดใหญ่ เราเรียกแต่ละลำดับนี้ว่ารูปแบบลำดับ (Sequential Pattern)

ตัวอย่าง พิจารณาฐานข้อมูลในรูปที่ 2.8 และรูปที่ 2.9 แสดงให้เห็นว่าฐานข้อมูลมีการจัดเรียงตามรหัสลูกค้าและเวลาในทรานแซกชัน ในรูปที่ 2.10 แสดงเซตของลำดับลูกค้า และในรูปที่ 2.11 แสดงฐานข้อมูลเซตลำดับลูกค้า ด้วยค่าสนับสนุนขั้นต่ำที่ตั้งไว้ที่ 25% ตัวอย่างเช่น สองลำดับ $\langle(30)(90)\rangle$ และ $\langle(30)(40\ 70)\rangle$ เป็นลำดับที่ยาวที่สุดซึ่งตอบสนองต่อเงื่อนไขของค่าสนับสนุนขั้นต่ำที่ตั้งไว้ และเป็นรูปแบบลำดับที่เราต้องการ รูปแบบลำดับ $\langle(30)(90)\rangle$ ได้รับการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สนับสนุนจากลูกค้าคนที่ 1 และ 4 โดยลูกค้าคนที่ 4 ซื้อรายการ (40 70) ในระหว่างรายการที่ 30 และ 90 แต่ยังคงสนับสนุนรูปแบบ $\langle(30)(90)\rangle$ เนื่องจากเรากำลังมองหา รูปแบบซึ่งไม่จำเป็นที่จะต้องอยู่ติดกัน รูปแบบลำดับ $\langle(30)(40\ 70)\rangle$ ได้รับการสนับสนุนจากลูกค้าคนที่ 2 และ 4 โดยลูกค้าคนที่ 2 ซื้อรายการที่ 60 พร้อมกับรายการที่ 40 และ 70 แต่ยังคงได้รับการสนับสนุนรูปแบบลำดับนี้เนื่องจาก (40 70) เป็นสับเซตของ (40 60 70)

ตัวอย่างลำดับซึ่งที่ไม่ได้รับการสนับสนุนคือ $\langle(10\ 20)(30)\rangle$ ซึ่งได้รับการสนับสนุนจากลูกค้าคนที่ 2 เท่านั้น รวมทั้งลำดับ $\langle(30)\rangle$, $\langle(40)\rangle$, $\langle(70)\rangle$, $\langle(90)\rangle$, $\langle(30)(40)\rangle$, $\langle(30)(70)\rangle$ และ $\langle(40)(70)\rangle$

Transaction Time	Customer Id	Items Bought
June 10 '93	2	10, 20
June 12 '93	5	90
June 15 '93	2	30
June 20 '93	2	40, 60, 70
June 25 '93	4	30
June 25 '93	3	30, 50, 70
June 25 '93	1	30
June 30 '93	1	90
June 30 '93	4	40, 70
July 25 '93	4	90

รูปที่ 2.8 ฐานข้อมูลต้นฉบับ

Customer Id	Transaction Time	Items Bought
1	June 25 '93	30
1	June 30 '93	90
2	June 10 '93	10, 20
2	June 15 '93	30
2	June 20 '93	40, 60, 70
3	June 25 '93	30, 50, 70
4	June 25 '93	30
4	June 30 '93	40, 70
4	July 25 '93	90
5	June 12 '93	90

รูปที่ 2.9 ฐานข้อมูลที่เรียงโดยรหัสลูกค้าและเวลาในทรานแซกชัน

Customer Id	Customer Sequence
1	$\langle(30)\ (90)\rangle$
2	$\langle(10\ 20)\ (30)\ (40\ 60\ 70)\rangle$
3	$\langle(30\ 50\ 70)\rangle$
4	$\langle(30)\ (40\ 70)\ (90)\rangle$
5	$\langle(90)\rangle$

รูปที่ 2.10 ชุดลำดับลูกค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sequential Patterns with support > 25%
{ (30) (90) }
((30) (40 70))

รูปที่ 2.11 รูปแบบลำดับที่มีค่าสนับสนุนมากกว่า 25 %

2.5.2 ขั้นตอนวิธีการค้นหาแบบลำดับ

ในการค้นหาแบบลำดับจะนำเสนออัลกอริทึม AprioriAll ซึ่งโครงสร้างโดยทั่วไป อัลกอริทึมนี้คือ มันจะมีจำนวนรอบการอ่านผ่านข้อมูลหลายรอบ ในแต่ละรอบเราจะเริ่มด้วยเซตตั้งต้นของลำดับขนาดใหญ่ เราใช้เซตนี้เพื่อสร้างลำดับขนาดใหญ่ ซึ่งจะเรียกลำดับนี้ว่า ลำดับคู่แข่ง จากนั้นค้นหาค่าสนับสนุนสำหรับลำดับคู่แข่งเหล่านี้ระหว่างที่มีการอ่านผ่านข้อมูล ในตอนท้ายของแต่ละรอบ จะพิจารณาว่าลำดับคู่แข่งอันไหนที่มีขนาดใหญ่ ลำดับคู่แข่งที่มีขนาดใหญ่เหล่านี้จะกลายมาเป็นตัวตั้งต้นในรอบถัดไป

2.5.2.1 อัลกอริทึม AprioriAll

อัลกอริทึมที่ชื่อว่า AprioriAll มีพื้นฐานมาจากอัลกอริทึม Apriori ที่ใช้สำหรับการเรียนรู้รูปแบบลำดับ ซึ่งจะหาลำดับทั้งหมดที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำ กำหนดให้ L_k คือเซตของ k-sequence ขนาดใหญ่ และ C_k เป็นเซตของ k-sequence คู่แข่ง

```

 $L_1 = \{\text{large 1-sequences}\};$  // Result of the itemset phase
for (  $k = 2; L_{k-1} \neq \emptyset; k++$  ) do
  begin
     $C_k =$  New candidates generated from  $L_{k-1}$  (see Section 3.1.1).
    foreach customer-sequence  $c$  in the database do
      Increment the count of all candidates in  $C_k$  that are contained in  $c$ .
     $L_k =$  Candidates in  $C_k$  with minimum support.
  end
Answer = Maximal Sequences in  $\bigcup_k L_k$ ;

```

รูปที่ 2.12 อัลกอริทึม AprioriAll

จากรูปที่ 2.12 ในแต่ละรอบเราใช้ลำดับขนาดใหญ่จากรอบก่อนหน้ามาสร้างลำดับคู่แข่ง และวัดค่าสนับสนุนโดยอ่านผ่านข้อมูลในฐานข้อมูล โดยรอบสุดท้ายค่าสนับสนุนของลำดับคู่แข่ง จะถูกใช้พิจารณาหาลำดับขนาดใหญ่ ยกเว้นในรอบแรกที่เซตของลำดับขนาดใหญ่ที่มีสมาชิก 1 ตัว (1- Large Sequence) จะหาได้โดยนับค่าสนับสนุนของแต่ละเซตของรายการจากในฐานข้อมูลแล้ว เปรียบเทียบค่าสนับสนุนของแต่ละเซตของรายการกับค่าสนับสนุนขั้นต่ำ

2.5.2.2 การสร้างลำดับคู่แข่ง

การสร้างลำดับคู่แข่ง (Apriori Candidate Generation) จะใช้ฟังก์ชัน apriori-generate โดยการนำ L_{k-1} ซึ่งเป็นเซตของ $(k-1)$ -sequence มาเป็นอาร์กิวเมนต์และคืนค่าลำดับคู่แข่งที่มีสมาชิก k ตัว ที่เป็นไปได้ทั้งหมด ฟังก์ชันมีการทำงานดังนี้ โดยขั้นแรกจะทำการประสาน (Join Step) L_{k-1} ด้วย L_{k-1} ซึ่งมีสมาชิก $k-2$ ตัวแรกเหมือนกันเข้าด้วยกันเพื่อสร้างลำดับคู่แข่งที่มีสมาชิก k ตัว และต่อมาจะทำการลบลำดับทั้งหมด $c \in C_k$ ซึ่งมีลำดับย่อยบางตัวของ c ที่มีสมาชิก $k-1$ ตัวที่ไม่อยู่ใน C_{k-1} เรียกขั้นตอนนี้ว่าขั้นตอนการตัดเล็ม (Prune Step) วิธีการสร้างลำดับคู่แข่งนี้จะเป็นการจัดลำดับเซตของรายการซึ่งถือว่าลำดับก่อนหลังมีความแตกต่างกัน อัลกอริทึมการสร้างลำดับคู่แข่งแสดงดังรูปที่ 2.13

```

insert into  $C_k$ 
select  $p.litemset_1, p.litemset_2, \dots, p.litemset_{k-1}, q.litemset_{k-1}$ 
from  $L_{k-1} p, L_{k-1} q$ 
where  $p.litemset_1 = q.litemset_1, \dots, p.litemset_{k-2} = q.litemset_{k-2};$ 

```

Next, delete all sequences $c \in C_k$ such that some $(k-1)$ -subsequence of c is not in L_{k-1} :

```

forall sequences  $c \in C_k$  do
  forall  $(k-1)$ -subsequences  $s$  of  $c$  do
    if  $(s \notin L_{k-1})$  then
      delete  $c$  from  $C_k;$ 

```

รูปที่ 2.13 อัลกอริทึมการสร้างลำดับคู่แข่ง

ขั้นตอนวิธีการค้นหารูปแบบลำดับโดยสรุปจะใช้อัลกอริทึม AprioriAll ซึ่งจะมีขั้นตอนการทำงานคล้ายกับอัลกอริทึม Apriori แต่จะมีความแตกต่างตรงที่วิธีการสร้างลำดับคู่แข่ง โดยจะถือว่าลำดับก่อนหลังมีความแตกต่างกัน หลังจากที่ได้ลำดับขนาดใหญ่ทั้งหมดที่เป็นไปได้แล้วจะนำลำดับขนาดใหญ่เหล่านี้ไปสร้างเป็นรูปแบบลำดับโดยใช้อัลกอริทึม genpattern โดยรูปแบบลำดับที่สร้างขึ้นจากอัลกอริทึมนี้ จะนำมาวิเคราะห์ แปลผล และนำไปประยุกต์ใช้

2.5.2.3 ตัวอย่างการทำงานของอัลกอริทึม AprioriAll

D	
TID	Sequences
1	$\langle\{1\}\{5\}\rangle$
2	$\langle\{1\}\{2,3,4\}\rangle$
3	$\langle\{1,3\}\rangle$
4	$\langle\{1\}\{2,3,4\}\{5\}\rangle$
5	$\langle\{5\}\rangle$

L_1	
1-Sequence	Count
$\langle 1 \rangle$	4
$\langle 2 \rangle$	2
$\langle 3 \rangle$	4
$\langle 4 \rangle$	4
$\langle 5 \rangle$	4

L_2	
2-Sequences	Count
$\langle 1 2 \rangle$	2
$\langle 1 3 \rangle$	4
$\langle 1 4 \rangle$	3
$\langle 1 5 \rangle$	3
$\langle 2 3 \rangle$	2
$\langle 2 4 \rangle$	2
$\langle 3 4 \rangle$	3
$\langle 3 5 \rangle$	2
$\langle 4 5 \rangle$	2

L_3	
3-Sequences	Count
$\langle 1 2 3 \rangle$	2
$\langle 1 2 4 \rangle$	2
$\langle 1 3 4 \rangle$	3
$\langle 1 3 5 \rangle$	2
$\langle 2 3 4 \rangle$	2

L_4	
4-Sequences	Count
$\langle 1 2 3 4 \rangle$	2

C_4		
Large 3 -Sequences	Candidate 4 -Sequences (after joining)	Candidate 4-Sequences (after pruning)
$\langle 1 2 3 \rangle$	$\langle 1 2 3 4 \rangle$	$\langle 1 2 3 4 \rangle$
$\langle 1 2 4 \rangle$	$\langle 1 2 4 3 \rangle$	
$\langle 1 3 4 \rangle$	$\langle 1 3 4 5 \rangle$	
$\langle 1 3 5 \rangle$	$\langle 1 3 5 4 \rangle$	
$\langle 2 3 4 \rangle$		

รูปที่ 2.14 ตัวอย่างการทำงานของอัลกอริทึม AprioriAll

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.14 แสดงตัวอย่างการทำงานของอัลกอริทึม AprioriAll ซึ่งสามารถอธิบายได้ดังนี้ ขั้นแรกจะนับค่าสนับสนุนของแต่ละเซตของรายการจากทรานแซกชันทั้งหมดในตารางฐานข้อมูล D จากนั้น พิจารณาเซตรายการที่มีค่าสนับสนุนไม่ต่ำกว่าค่าสนับสนุนขั้นต่ำ ซึ่งกำหนดไว้เท่ากับ 2 ทำให้ได้ลำดับขนาดใหญ่ที่มีสมาชิก 1 ตัว จำนวน 5 ลำดับคือ $\langle 1 \rangle$ $\langle 2 \rangle$ $\langle 3 \rangle$ $\langle 4 \rangle$ และ $\langle 5 \rangle$ แสดงดังตาราง L_1 ในรอบต่อมานำลำดับขนาดใหญ่ L_1 มาสร้างเป็นลำดับคู่แข่งที่มีสมาชิก 2 ตัว และจะกำหนดลำดับขนาดใหญ่ โดยพิจารณาจากค่าสนับสนุนขั้นต่ำ ซึ่งทำให้ได้ลำดับขนาดใหญ่ที่มีสมาชิก 2 ตัว จำนวน 9 ลำดับ แสดงดังตาราง L_2 ในรอบต่อมานำลำดับขนาดใหญ่จากใน L_2 มาสร้างเป็นลำดับคู่แข่งที่มีสมาชิก 3 ตัว และกำหนดลำดับขนาดใหญ่โดยพิจารณาจากค่าสนับสนุนขั้นต่ำ ซึ่งทำให้ได้ลำดับขนาดใหญ่ที่มีสมาชิก 3 ตัวดังแสดงในตาราง L_3 ในรอบต่อมาในขั้นตอนการประสานจะนำลำดับขนาดใหญ่จาก L_3 มาสร้างเป็นลำดับคู่แข่งที่มีสมาชิก 4 ตัว จำนวน 4 ลำดับ แสดงดังสมการที่ 2 ของตาราง C_4 ต่อมาในขั้นตอนการเลือกจะกำจัดลำดับคู่แข่งทุกตัวที่มีลำดับย่อยที่มีสมาชิก 3 ตัวที่ไม่อยู่ใน L_3 ออกไปซึ่งทำให้เหลือลำดับคู่แข่งเพียง 1 ลำดับ ดังแสดงในสมการที่ 3 ของตารางที่ C_4 และจะพิจารณาว่าเป็นลำดับคู่แข่งนี้ว่าเป็นลำดับขนาดใหญ่หรือไม่ โดยพิจารณาจากค่าสนับสนุนซึ่งต้องไม่ต่ำกว่าค่าสนับสนุนขั้นต่ำ ทำให้ได้ลำดับขนาดใหญ่ที่มีสมาชิก 4 ตัว จำนวน 1 ลำดับ ดังแสดงในตาราง L_4 จากนั้นนำลำดับขนาดใหญ่นี้มาสร้างเป็นลำดับคู่แข่งที่มีสมาชิก 5 ตัว แต่เนื่องจากไม่สามารถสร้างลำดับคู่แข่งที่มีสมาชิก 5 ตัวจาก L_4 ได้ จึงทำให้ไม่สามารถสร้างลำดับขนาดใหญ่ที่มีสมาชิก 5 ตัวได้ อัลกอริทึม AprioriAll จึงหยุดการทำงาน

2.6 วิธีการสร้างรูปแบบลำดับ

ในการสร้างรูปแบบลำดับจะทำการหาลำดับย่อยทั้งหมดที่มีสมาชิกมากกว่า 1 ตัวของลำดับขนาดใหญ่ s ทุกตัว ดังนั้น สำหรับลำดับย่อย a ทุกๆตัวของ s จะสร้างรูปแบบลำดับที่อยู่ในรูป $a \Rightarrow (s-a)$ ถ้าอัตราส่วนของค่าสนับสนุนของ s ต่อค่าสนับสนุนของ a มีค่าไม่ต่ำกว่าค่าความเชื่อมั่นขั้นต่ำ

เราสามารถปรับปรุงกระบวนการข้างบนได้โดยสร้างลำดับย่อยของลำดับขนาดใหญ่โดยทำซ้ำไปในเชิงลึกก่อน ตัวอย่างเช่น เมื่อกำหนดลำดับ ABCD มาให้ ขั้นแรกเราจะพิจารณาลำดับย่อย ABC และลำดับย่อยตัวต่อไปคือ AB เป็นต้น และถ้าลำดับย่อย a ของลำดับขนาดใหญ่ s ไม่สามารถสร้างรูปแบบลำดับได้ก็ไม่จำเป็นที่จะต้องพิจารณาลำดับย่อยอื่นของ a เพื่อมาใช้ในการสร้างรูปแบบลำดับ

ตัวอย่างเช่น ถ้า $ABC \Rightarrow D$ มีค่าความเชื่อมั่นไม่ผ่านเกณฑ์ที่เรากำหนดแล้ว ก็ไม่จำเป็นต้องตรวจสอบ $AB \Rightarrow CD$ อีก วิธีการนี้จะไม่ทำให้พลาดรูปแบบลำดับใด เนื่องจากค่าสนับสนุนของลำดับย่อย a' ของ a จะมีค่ามากกว่า a ซึ่งจะทำให้ค่าความเชื่อมั่นของรูปแบบลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$a' \Rightarrow (s-a')$ มีค่าไม่มากไปกว่าค่าความเชื่อมั่นของ $a \Rightarrow (s-a)$ ซึ่งหมายความว่า ถ้า a ไม่สามารถสร้างรูปแบบลำดับจาก s ได้ a' ก็ไม่สามารถสร้างได้เช่นเดียวกัน วิธีการสร้างรูปแบบลำดับใช้อัลกอริทึมแสดงดังรูปที่ 2.15

For all large sequence $s_k, k \geq 2$ do

Call genpattern(s_k);

Procedure genpattern(s_k : large k -sequence)

$i = k-1$;

while $i \geq 1$ begin

$a =$ subsequence of s_k that is first i elements of s_k

$conf = \text{support}(s_k) / \text{support}(a)$;

if ($conf \geq \text{minconf}$) then

output the rule $a \Rightarrow (s_k - a)$, with confidence = $conf$ and support = $\text{support}(s_k)$

else

exit loop

$i = i - 1$;

end

end

รูปที่ 2.15 อัลกอริทึมการสร้างรูปแบบลำดับ

2.7 การอ้างอิงไปข้างหน้าไกลที่สุด

2.7.1 ที่มาของปัญหา

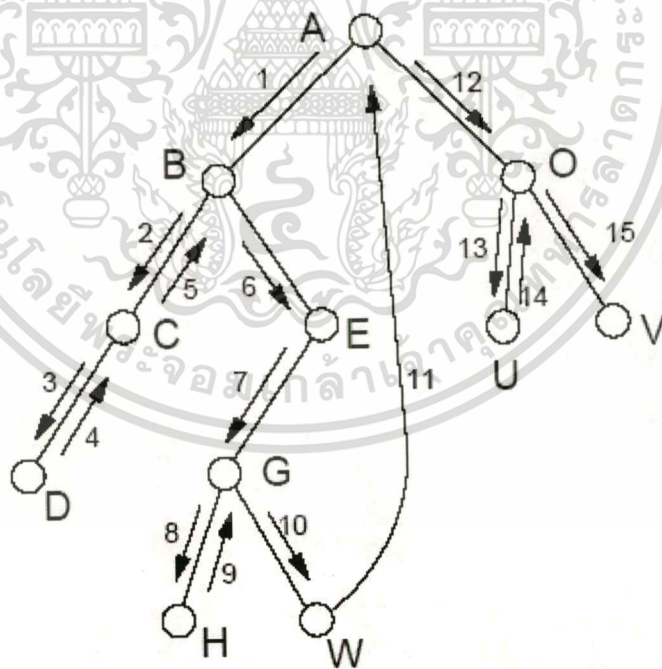
สารสนเทศได้จัดไว้ในสภาพแวดล้อมที่ซึ่งอ็อบเจกต์ต่างๆ ถูกลิงก์เข้าไว้ด้วยกัน ผู้ใช้มักจะเดินทางไปยังอ็อบเจกต์ไปข้างหน้าและย้อนกลับตามที่ได้มีการจัดวางลิงก์และไอคอนไว้ ดังนั้น บางโหนดอาจจะถูกเยี่ยมชมซ้ำเนื่องจากตำแหน่งที่ตั้งของมันเองมากกว่าที่จะเป็นเนื้อหาของมัน ตัวอย่างเช่น ในสภาพแวดล้อมของเว็ลด์ไวด์เว็บ การที่เข้าถึงโหนดพี่น้อง (Sibling Node) ผู้ใช้มีแนวโน้มที่จะใช้ไอคอนย้อนกลับ (Backward Icon) จากนั้นเลือกเดินทางไปข้างหน้าแทนที่จะเปิด URL ใหม่ ดังนั้น การที่จะดึงเอารูปแบบการเข้าถึงของผู้ใช้ที่มีความสำคัญจากเว็บเซิร์ฟเวอร์ล็อก เราต้องพิจารณาผลกระทบที่เกิดจากการเดินทางไปแบบย้อนกลับและค้นหารูปแบบการเข้าถึงจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เราตั้งสมมุติฐานว่าการอ้างอิงย้อนกลับเกิดขึ้นได้ง่ายสำหรับการเยี่ยมชม แต่ไม่ใช่สำหรับการค้นหา และให้ความสนใจกับรูปแบบการอ้างอิงไปข้างหน้า (Ming-Syan, C. et al. 1998 : 209-221)

การอ้างอิงย้อนกลับ (Backward Reference) หมายถึงการเยี่ยมชมอ็อบเจกต์ที่ผู้ใช้เพิ่งเยี่ยมชมไปก่อนหน้านี้โดยผู้ใช้คนเดียวกัน เมื่อเกิดการอ้างอิงย้อนกลับ การอ้างอิงไปข้างหน้าถือเป็นสิ้นสุด ผลลัพธ์ของเส้นทางการอ้างอิงไปข้างหน้ากำหนดไว้ในรูปของการอ้างอิงไปข้างหน้าไกลที่สุด (Maximal Forward Reference) หลังจากที่เราได้การอ้างอิงไปข้างหน้าไกลที่สุดแล้ว เราจะย้อนกลับไปยังจุดที่จะเริ่มต้นการอ้างอิงไปข้างหน้า และดำเนินเส้นทางการอ้างอิงไปข้างหน้าต่อ นอกจากนี้การมีอยู่ของโหนดว่าง (Null Node) ยังเป็นตัวบ่งชี้การสิ้นสุดของเส้นทางการอ้างอิงไปข้างหน้าและเป็นจุดเริ่มต้นเส้นทางใหม่

อัลกอริทึมการอ้างอิงไปข้างหน้าไกลที่สุดจะกล่าวถึงในหัวข้อถัดไป ตัวอย่างต่อไปนี้แสดงการอ้างอิงไปข้างหน้าไกลที่สุด สมมุติการเดินทางของผู้ใช้คนหนึ่งเป็นดังต่อไปนี้ {A, B, C, D, C, B, E, G, H, W, A, O, U, O, V} ดังแสดงในรูปที่ 2.16 จากนั้นใช้อัลกอริทึมนี้หาชุดของทรานแซคชันย่อยที่อยู่ในแต่ละเซสชันของผู้ใช้จะได้ผลลัพธ์ดังต่อไปนี้คือ {ABCD, ABEGH, ABEGW, AOU, AOV}



รูปที่ 2.16 ตัวอย่างของเส้นทางการเดินทางของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 อัลกอริทึมสำหรับการอ้างอิงไปข้างหน้าไกลที่สุด

อัลกอริทึมการอ้างอิงไปข้างหน้าไกลที่สุด (MF) จะแปลงลำดับการเดินทางไปในเว็บไซต์ของผู้ใช้ให้เป็นชุดของการอ้างอิงไปข้างหน้าไกลที่สุด ฐานข้อมูลสื่อประกอบด้วยคู่ของ (ต้นทาง, ปลายทาง) ของแต่ละการเดินทางที่ลิงก์เข้าหากัน สำหรับการเริ่มต้นของการเดินทางใหม่ที่ซึ่งไม่ได้ลิงก์กันกับการเดินทางก่อนหน้านี้ โหนดเริ่มต้นจะว่าง กำหนดให้ลำดับการเดินทางของผู้ใช้เป็น $\{(s_1, d_1), (s_2, d_2), \dots, (s_n, d_n)\}$ จากนั้นจะจัดเป็นหลายชุดลำดับ ซึ่งแต่ละชุดลำดับจะแสดงถึงการอ้างอิงไปข้างหน้าไกลที่สุด อัลกอริทึมสำหรับหาการอ้างอิงไปข้างหน้าไกลที่สุดแสดงดังรูปด้านล่าง เริ่มจากฐานข้อมูลสื่อการเดินทางจะถูกเรียงลำดับโดยใช้ User ID ได้ออกมาเป็น $\{(s_1, d_1), (s_2, d_2), \dots, (s_n, d_n)\}$ สำหรับผู้ใช้แต่ละคนเรียงตามลำดับเวลา จากนั้นนำอัลกอริทึม MF นำมาประยุกต์ใช้เพื่อการอ้างอิงไปข้างหน้าไกลที่สุด กำหนดให้ D_F แสดงถึงฐานข้อมูลที่ใส่เก็บทุกๆ การอ้างอิงไปข้างหน้าไกลที่สุด

Algorithm MF: An algorithm to find maximal forward references.

Step 1: Set $i = 1$ and string Y to null for initialization, where string Y is used to store the current forward reference path. Also, set the flag $F = 1$ to indicate a forward traversal.

Step 2: Let $A = s_i$ and $B = d_i$.

If A is equal to null then

/* this is the beginning of a new traversal */

begin

Write out the current string Y (if not null) to the database D_F ;

Set string $Y = B$;

Go to Step 5.

end

Step 3: If B is equal to some reference (say the j -th reference) in string Y then

/* this is a cross-referencing back to a previous reference */

begin

If F is equal to 1 then write out string Y to database D_F ;

Discard all the references after the j -th one in string Y ;

$F = 0$;

Go to Step 5.

end

Step 4: Otherwise, append B to the end of string Y .

/* we are continuing a forward traversal */

If F is equal to 0, set $F = 1$.

Step 5: Set $i = i + 1$. If the sequence is not completed scanned then go to Step 2.

รูปที่ 2.17 อัลกอริทึม MF

จากรูปที่ 2.16 สามารถตรวจสอบได้ว่าการเข้าอ้างอิงย้อนกลับเกิดขึ้นครั้งแรกในการเคลื่อนที่ครั้งที่ 4 (จาก D ไปยัง C) ที่จุดนั้น การอ้างอิงไปข้างหน้า ABCD ถูกเขียนลงใน D_F (โดยขั้นตอนที่ 3)

ในการเคลื่อนที่ถัดมา (จาก C ไป B) แม้ว่าเงื่อนไขแรกในขั้นตอนที่ 3 จะเป็นจริงอีกครั้ง แต่ไม่มีการเขียนอะไรลงใน D_F เนื่องจากค่าแฟลก $F = 0$ ซึ่งหมายความว่าอยู่ในการเดินทางย้อนกลับ การอ้างอิงไปข้างหน้าอันถัดไปจะเก็บ ABEGH ลงสตริงที่ Y ซึ่งต่อไปจะถูกเขียนลงใน D_F เมื่อพบการอ้างอิงย้อนกลับ (จาก H ไปยัง G) การทำงานโดยอัลกอริทึม MF โดยมีอินพุตในรูปแบบที่ 2.16 แสดงดังตารางที่ 2.5

ตารางที่ 2.5 ตัวอย่างการทำงานโดยใช้อัลกอริทึม MF

move	string Y	output to D_F
1	AB	-
2	ABC	-
3	ABCD	-
4	ABC	ABCD
5	AB	-
6	ABE	-
7	ABEG	-
8	ABEGH	-
9	ABEG	ABEGH
10	ABEGW	-
11	A	ABEGW
12	AO	-
13	AOU	-
14	AO	AOU
15	AOV	AOV (end)

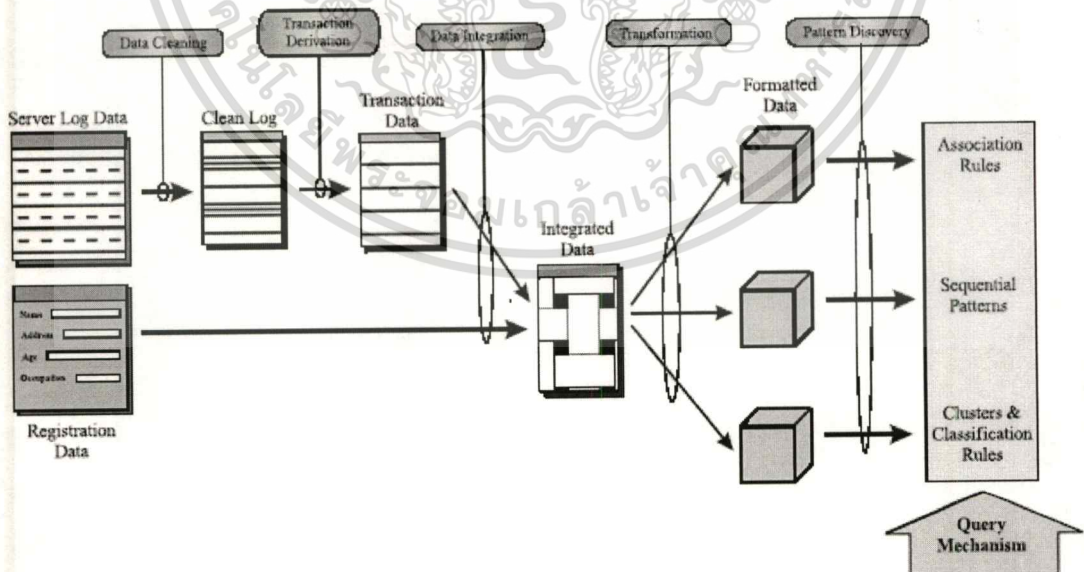
ในบางกรณีที่อยู่ในระเบียบของล๊อคเก็บเฉพาะปลายทางแทนที่จะเป็นคู่ลำดับ เช่น $\{d_1, d_2, \dots, d_n\}$ ด้วยอินพุตลักษณะนี้ เรายังสามารถที่จะแปลงไปเป็นเซตของการอ้างอิงไปข้างหน้า 'ไกลที่สุดได้' สิ่งที่เป็นข้อแตกต่างมีเพียงว่าในกรณีนี้ เราไม่สามารถระบุจุดพัก (Breakpoint) ซึ่งจะบอกว่าผู้ใช้ได้เริ่มต้นการเดินทางเส้นใหม่ ตัวอย่างเช่น ABEH และ WXYZ เป็นชุดการอ้างอิงไปข้างหน้าไกลที่สุด 2 อันที่ต่อเนื่องกัน แต่เรามองมันเป็นเส้นทางเพียงเส้นเดียวคือ ABEHWXYZ ถ้าไม่มีโหนดต้นทางแล้ว จะเพิ่มความซับซ้อนในการคำนวณ เพราะเส้นทางที่ใช้พิจารณาจะมีความยาวมากขึ้น อย่างไรก็ตาม ข้อจำกัดนี้มีผลกระทบเล็กน้อยในการระบุการอ้างอิงลำดับถัดไป เนื่องจากไม่มีลิงก์เชิงตรรกะระหว่าง H ละ W ดังนั้น ลำดับการอ้างอิงถัดไปที่มี HW ไม่น่าจะปรากฏขึ้นบ่อย อัลกอริทึม MF สามารถใช้ในกรณีที่โหนดต้นทางไม่มีอยู่ได้

2.8 งานวิจัย WEBMINER

WEBMINER เป็นระบบสำหรับการทำเหมืองข้อมูลการใช้งานเว็บพัฒนาโดย Bamshad Mobasher, Namit Jain, Eui-Hong (Sam) Han, Jaideep Srivastava (Cooley, R. et al. 1999 : 5)

แรกเริ่มได้มีการนำเสนอกรอบการทำงานสำหรับการทำเหมืองข้อมูลเว็บ ซึ่งเป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลและการค้นหาความรู้ในข้อมูลทรานแซกชันที่ได้จากเว็บไซต์ ในส่วนของสถาปัตยกรรมของระบบที่นำเสนอ จะแบ่งกระบวนการทางการทำเหมืองข้อมูลเว็บ ออกเป็น 2 ส่วนหลักๆ คือ ส่วนแรกเกี่ยวข้องกับกระบวนการในการแปลงข้อมูลเว็บให้เป็นรูปแบบของทรานแซกชันที่เหมาะสม ส่วนที่สองเกี่ยวข้องกับการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล ได้แก่ เทคนิคการค้นหาความสัมพันธ์และรูปแบบลำดับ ภาพรวมของสถาปัตยกรรมแสดงดังรูปที่ 2.18

จากรูปที่ 2.18 อธิบายดังนี้ เริ่มจากการนำเข้าสู่ข้อมูลเซิร์ฟเวอร์ล็อกแล้วนำไปกำจัดรายการที่ไม่เกี่ยวข้องออกได้ออกมาเป็น Clean Log จากนั้น นำมาหาทรานแซกชัน จากรูปไม่เพียงแต่ข้อมูลจาก ล็อกของการเข้าถึง เราสามารถนำข้อมูลการลงทะเบียนของผู้ใช้มารวมกันกับทรานแซกชันที่ได้ แล้วนำจัดรูปแบบให้เหมาะสมเพื่อนำเข้าสู่กระบวนการค้นหาแบบ ซึ่งWEBMINER นำเสนอกลไกภาษาที่มีโครงสร้างเหมือน SQL ในการสอบถาม (SQL-Like Query) เพื่อค้นหาองค์ความรู้ในรูปแบบของความสัมพันธ์และรูปแบบลำดับ ส่วนการรวมกลุ่มและการจำแนกเป็นงานที่กำลังพัฒนาเพิ่มในระบบ WEBMINER



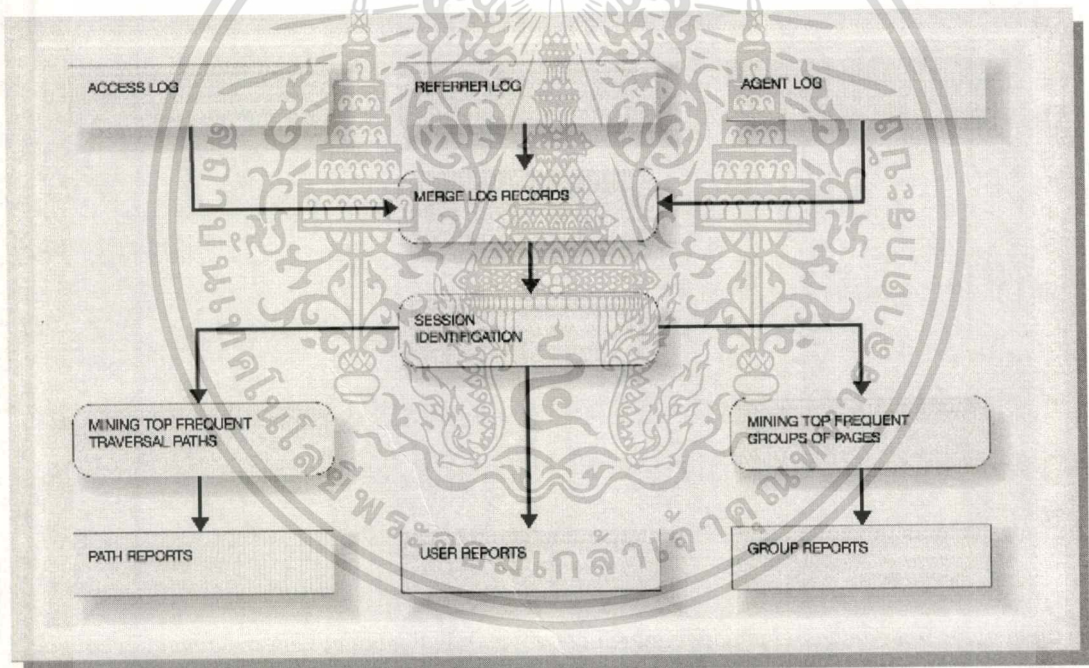
รูปที่ 2.18 สถาปัตยกรรมของ WEBMINER

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9 งานวิจัย SpeedTracer

SpeedTracer เป็นเครื่องมือการทำเหมืองข้อมูลการใช้งานเว็บที่สร้างขึ้นเพื่อทำความเข้าใจพฤติกรรมการใช้งานของผู้ใช้เว็บไซต์ โดยใช้เว็บเซิร์ฟเวอร์ล็อกไฟล์กับเทคนิคการทำเหมืองข้อมูลพัฒนาโดย Kun-Lung Wu, Philip S. Yu, Allen Ballman

SpeedTracer จะเริ่มต้นการระบุเซสชันผู้ใช้ด้วยการสร้างเส้นทางการเดินทางของผู้ใช้ (User Traversal Path) โดยไม่ใช่ลูกก็หรือการลงทะเบียนเพื่อใช้ในการระบุเซสชันผู้ใช้ ความเป็นส่วนตัวของผู้ใช้ได้รับการปกป้อง เมื่อเซสชันผู้ใช้ได้ถูกระบุแล้ว อัลกอริทึมการทำเหมืองข้อมูลถูกนำมาประยุกต์ใช้เพื่อค้นหาเส้นทางการเดินทางที่พบร่วมกันมากที่สุดและกลุ่มของเพจที่มีการเยี่ยมชมด้วยกันบ่อยครั้ง รูปแบบการเข้าถึงที่สำคัญช่วยให้เข้าใจพฤติกรรมทางท่องไปในเว็บไซต์ของผู้ใช้ SpeedTracer สามารถออกรายงานได้ 3 ประเภทด้วยกัน ได้แก่ รายงานผู้ใช้ รายงานเส้นทางการเข้าถึง และรายงานกลุ่มของเส้นทางการเข้าถึง SpeedTracer มีการทำงานดังไดอะแกรมดังต่อไปนี้



รูปที่ 2.19 โพลว์ไดอะแกรมสำหรับการพัฒนา SpeedTracer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบโปรแกรมการทำเหมืองข้อมูลการใช้งานเว็บ

3.1 รูปแบบลักษณะการดำเนินธุรกิจ

กรณีศึกษาเว็บ <http://www.differentasia.com> เป็นธุรกิจตัวแทนให้บริการจองที่พักโรงแรมต่างๆ ซึ่งสามารถจองผ่านทางเว็บไซต์และชำระเงินผ่านบัตรเครดิต บริษัททำให้เป็นที่รู้จักผ่านทางสื่อสิ่งพิมพ์โดยลงโฆษณาที่นิตยสารท่องเที่ยว เปิดบูธในงานท่องเที่ยว จัดทำเว็บไซต์ <http://www.differentasia.com> และมีการส่งจดหมายข่าว เพื่อเสนอให้ลูกค้าหรือกลุ่มของลูกค้าที่เคยใช้บริการ และลูกค้าที่ยังไม่เคยใช้บริการแต่มีแนวโน้มที่จะมาใช้บริการพิจารณา

3.2 ขอบเขตความต้องการของระบบ

ระบบสามารถวิเคราะห์รูปแบบลำดับการเยี่ยมชมเว็บเพจบนเครื่องเซิร์ฟเวอร์เพียงเครื่องเดียวเท่านั้นและเว็บเซิร์ฟเวอร์ที่นำเข้ามาเป็นแบบแบบคอมไบนีหรือเอ็กซ์เทนดที่นั่น ในเฟสการค้นหารูปแบบ ระบบสามารถค้นหารูปแบบลำดับการใช้งานเว็บไซต์ของผู้ใช้โดยใช้เทคนิครูปแบบลำดับ และการวิเคราะห์ทางสถิติ ซึ่งในส่วนของ การค้นหารูปแบบลำดับในการเข้าถึงเว็บเพจจากเว็บเซิร์ฟเวอร์ล็อกในระบบจะใช้อัลกอริทึม AprioriAll ในส่วนการวิเคราะห์ทางสถิติจะประกอบไปด้วย สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง สถิติแยกตามไอพีแอดเดรสของผู้ใช้ สถิติแยกตามเอเจนต์ สถิติแยกตามจำนวนการเข้าชมเว็บไซต์ และสถิติแยกตามเว็บเพจที่เข้าชม

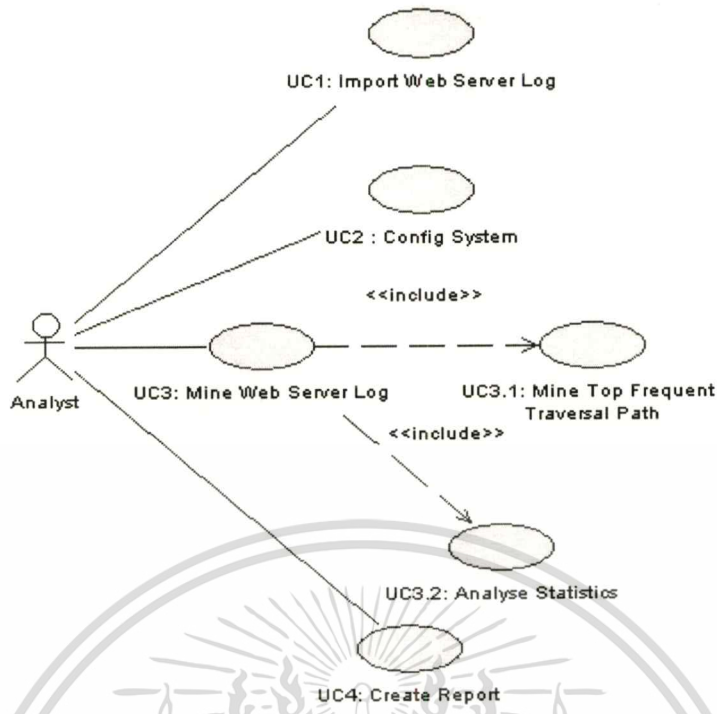
องค์ความรู้เหล่านี้สามารถนำไปประยุกต์ใช้กับการปรับปรุงโครงสร้างเว็บไซต์เพื่อช่วยให้เข้าถึงได้ง่ายและทั่วถึงมากยิ่งขึ้น

3.3 การวิเคราะห์และออกแบบระบบ

3.3.1 ยูสเคสระบบ

ภาพรวมของระบบประกอบไปด้วย นักวิเคราะห์ (Analyst) ผู้ซึ่งนำเข้าเว็บเซิร์ฟเวอร์ล็อกและกำหนดค่าและจัดการข้อมูลต่างๆที่ระบบต้องใช้ในการประมวลผล จากนั้นนักวิเคราะห์สั่งให้ระบบทำเหมืองข้อมูลการเข้าใช้งานเว็บไซต์ เมื่อระบบประมวลผลเสร็จเรียบร้อยแล้ว นักวิเคราะห์สามารถสั่งให้ระบบออกรายงานได้ ผลลัพธ์จากการวิเคราะห์ที่ได้จากระบบ ผู้พัฒนาเว็บ (Web Developer) จะนำไปใช้ปรับโครงสร้างเว็บไซต์ สำหรับยูสเคสของระบบแสดงดังรูปที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 ยูสเคสโคอะแกรมของระบบ

ตารางที่ 3.1 คำอธิบายยูสเคส Import Web Server Log

UC1: Import Web Server Log	
Brief Description:	ยูสเคสนี้ใช้อธิบายว่า Analyst จะนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์ได้อย่างไร โดยการเลือกเมนู File -> Import Log หรือปุ่ม Import จากนั้นเลือกไฟล์ที่ต้องการนำเข้า เมื่อระบบตรวจสอบว่านำเข้าไฟล์ถูกต้อง จึงทำการนำเข้าไฟล์นั้นสู่ระบบเพื่อเตรียมพร้อมสำหรับการประมวลผล
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	Analyst ต้องล็อกอินเข้ามาในระบบเรียบร้อยแล้ว
Basic Flows:	<ol style="list-style-type: none"> 1) Analyst เลือกเมนู File -> Import Log หรือปุ่ม Import 2) Analyst นำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์ที่ต้องการ 3) ระบบตรวจสอบการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์ว่าถูกต้อง
Alternative Flows:	2a) ถ้าAnalyst ยกเลิกการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์ ระบบจะแสดงข้อความเตือน
Post Conditions:	ระบบสามารถนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเท่านั้น ไม่ควรเผยแพร่ไปยังบุคคลภายนอกโดยไม่ได้รับอนุญาตจากคณะผู้บริหาร

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 คำอธิบายยูสเคส Configure System

UC2: Configure System	
Brief Description:	ยูสเคสนี้ใช้ธิบายว่า Analyst เข้ามากำหนดค่าต่างๆและจัดการข้อมูลที่ใช้ในระบบได้อย่างไร โดยการเลือกเมนู Configure -> Option หรือปุ่ม Configure จากนั้นเลือกกำหนดค่าและจัดการข้อมูลต่างๆ ระบบจะบันทึกค่าต่างๆเหล่านี้เพื่อเตรียมพร้อมสำหรับการประมวลผล
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	Analyst ต้องล็อกอินเข้ามาในระบบเรียบร้อยแล้ว
Basic Flows:	<ol style="list-style-type: none"> 1) Analyst เลือกเมนู Configure -> Option หรือปุ่ม Configure 2) ระบบจะแสดงหน้าจอสำหรับการกำหนดค่าและจัดการข้อมูล 3) Analyst เลือกรายการที่ต้องการกำหนดค่าและจัดการข้อมูลซึ่งประกอบไปด้วยหน้าจอดังต่อไปนี้ <ul style="list-style-type: none"> ● Session ใช้กำหนดค่าเซสชัน ● Pattern Discovery ใช้กำหนดค่า Minimum Support, Minimum Confidence และ Maximum k-itemset ● Suffix Filtering ใช้จัดการข้อมูลที่ใช้ในการกรอง ● Report Setting ใช้กำหนดรายงานที่ต้องการแสดง ● User Account ใช้จัดการรายชื่อผู้ใช้ 4) Analyst ยืนยันการกำหนดค่าและจัดการข้อมูลต่างๆ 5) ระบบบันทึกรายการเปลี่ยนแปลงที่เกิดขึ้น
Alternative Flows:	4a) ถ้า Analyst ยกเลิกการกำหนดค่าและการจัดการข้อมูล ระบบจะยกเลิกบันทึกการเปลี่ยนแปลงที่เกิดขึ้น
Post Conditions:	ระบบสามารถกำหนดค่าและจัดการข้อมูลที่ใช้ในระบบได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 คำอธิบายยูสเคส Mine Web Server Log

UC3: Mine Web Server Log	
Brief Description:	ยูสเคสนี้ใช้อธิบายว่า Analyst จะสั่งให้ระบบทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกได้อย่างไร โดยการเลือกเมนู Run -> Execute หรือปุ่ม Run จากนั้นระบบจะทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกแล้วจึงแสดงรายงาน
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	<ol style="list-style-type: none"> 1) Analyst ต้องล็อกอินเข้ามาในระบบเรียบร้อยแล้ว 2) Analyst ต้องนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์เรียบร้อยแล้ว
Basic Flows:	<ol style="list-style-type: none"> 1) Analyst เลือกเมนู Run -> Execute หรือปุ่ม Run 2) ระบบจะนำเว็บเซิร์ฟเวอร์ล็อกที่ได้นำเข้ามาทำความสะอาด เพื่อใช้ในขั้นตอนการทำเหมืองข้อมูล 3) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ได้ทำความสะอาดเรียบร้อยแล้ว มาทำเหมืองข้อมูล ตามที่ได้กำหนดไว้ใน UC2 4) Analyst เลือกรายงานที่ต้องการดูในหน้าจอแสดงผลรายงาน 5) ระบบแสดงผลรายงานที่ Analyst เลือกในเชิงของตาราง
Alternative Flows:	<ol style="list-style-type: none"> 3a) ถ้า Analyst ยังไม่ได้กำหนดค่าและจัดการข้อมูลต่างๆ ที่ใช้ในระบบ ระบบจะใช้ค่าที่ติดตั้งโดยปริยาย 3b) ถ้า Analyst กำหนดค่าและจัดการข้อมูลต่างๆ ที่ใช้ในระบบเป็น <ol style="list-style-type: none"> 3b1) รายงานรูปแบบลำดับ : ระบบทำ UC3.1 3b2) รายงานการวิเคราะห์ทางสถิติ : ระบบทำ UC3.2
Post Conditions:	ระบบสามารถทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกได้

ตารางที่ 3.4 คำอธิบายยูสเคส Mine Top Frequent Traversal Path

UC3.1: Mine Top Frequent Traversal Path	
Brief Description:	ยูสเคสนี้ใช้อธิบายว่าระบบวิเคราะห์เหมืองข้อมูลเว็บเซิร์ฟเวอร์ ล็อกโดยใช้เทคนิครูปแบบลำดับได้อย่างไร
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	Analyst กำหนดรายงานที่ต้องการเป็นรายงานรูปแบบลำดับ
Basic Flows:	<ol style="list-style-type: none"> 1) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการทำความสะอาดมา ระบุตัวผู้ใช้ 2) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุตัวผู้ใช้ มาระบุ เซสชันผู้ใช้ 3) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุเซสชันผู้ใช้ มา ระบุทรานแซคชันผู้ใช้ 4) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุทรานแซคชัน ผู้ใช้มาระบุลำดับผู้ใช้ 5) ระบบนำลำดับผู้ใช้มาค้นหาในรูปแบบลำดับ 6) ระบบบันทึกรูปแบบลำดับที่ได้ลงฐานข้อมูล
Alternative Flows:	-
Post Conditions:	ระบบสามารถทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้เทคนิค รูปแบบลำดับได้

ตารางที่ 3.5 คำอธิบายยูสเคส Analyze Statistics

UC3.2: Analyze Statistics	
Brief Description:	ยูสเคสนี้ใช้อธิบายระบบวิเคราะห์เหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อก โดยใช้การวิเคราะห์ทางสถิติได้อย่างไร
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	Analyst กำหนดรายงานที่ต้องการเป็นรายงานทางสถิติ
Basic Flows:	<ol style="list-style-type: none"> 1) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการทำความสะอาดมาระบุตัวผู้ใช้ 2) ระบบนำเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุตัวผู้ใช้มาระบุเซสชันผู้ใช้ 3) ระบบนำเว็บเซิร์ฟเวอร์ล็อกจากข้อ 1 และ 2 มาสร้างรายงานทางสถิติ ซึ่งประกอบไปด้วย <ul style="list-style-type: none"> ● สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง ● สถิติแยกตามไอพีแอดเดรสของผู้ใช้ ● สถิติแยกตามเอเจนต์ ● สถิติแยกตามจำนวนการเข้าชมเว็บไซต์ ● สถิติแยกตามเว็บเพจที่เข้าชม 4) ระบบบันทึกสถิติที่ได้ลงฐานข้อมูล
Alternative Flows:	-
Post Conditions:	ระบบสามารถทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้การวิเคราะห์ทางสถิติได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

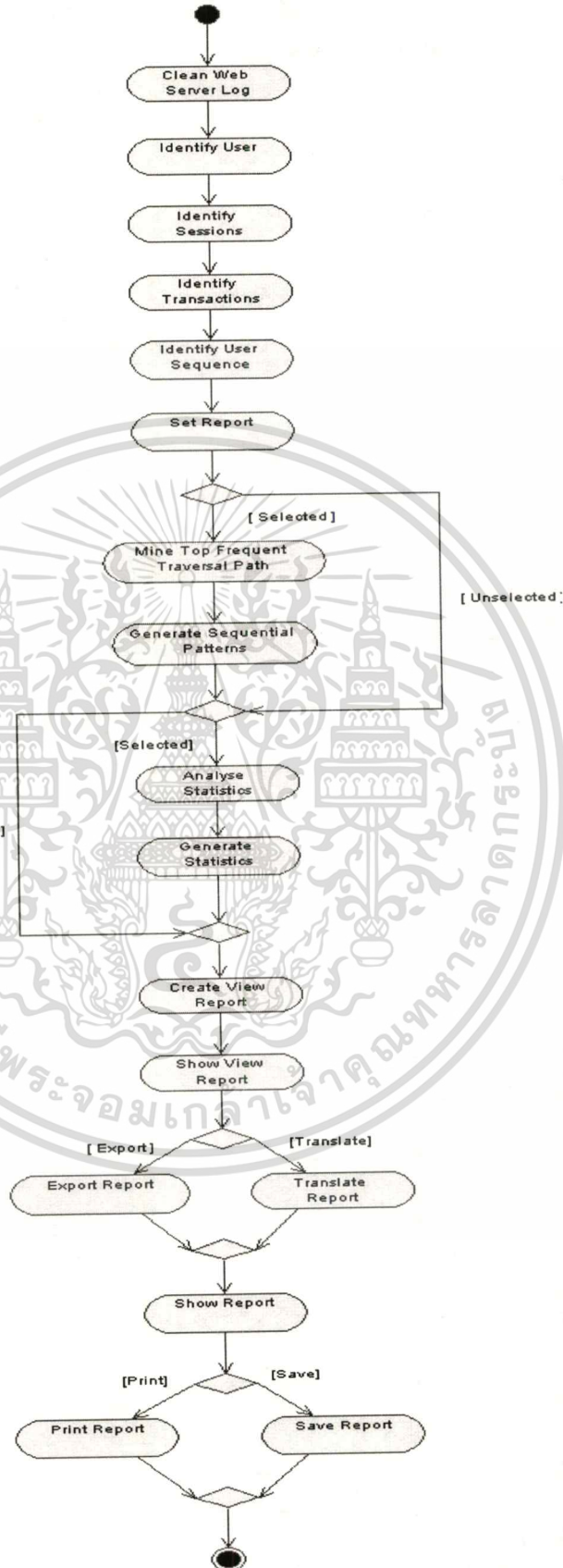
ตารางที่ 3.6 คำอธิบายยูสเคส Create Report

UC4: Create Report	
Brief Description:	ยูสเคสนี้ใช้อธิบายว่า Analyst ตั้งให้ระบบออกรายงานได้อย่างไร รายงานที่ออกนั้นมี 2 แบบ คือ รายงานรูปแบบลำดับและรายงานทางสถิติ ในการออกรายงานนั้น Analyst จะคลิกที่ชื่อรายงานที่ต้องการ จากนั้นเลือกเมนู Report แล้วเลือกวิธีการออกรายงานซึ่งมีอยู่ 2 วิธี คือ Export และ Translate จากนั้นระบบจะออกรายงานให้
Primary Actors:	Analyst
Stakeholder:	Web Developer
Preconditions:	ระบบต้องทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ลึกลับเรียบร้อยแล้ว
Basic Flows:	<ol style="list-style-type: none"> 1) Analyst เลือกรายงานที่ต้องการจากหน้าจอแสดงรายงาน 2) ระบบแสดงรายงานที่ Analyst เลือกในเชิงตาราง 3) Analyst เลือกอย่างใดอย่างหนึ่งระหว่างเลือกเมนู Report -> Export หรือปุ่ม Export กับเลือกเมนู Report -> Translate หรือปุ่ม Translate 4) ระบบออกรายงาน
Alternative Flows:	<ol style="list-style-type: none"> 3a) ถ้า Analyst เลือกเมนู Report -> Export หรือปุ่ม Export <ol style="list-style-type: none"> 3a1) กรณีรายงานที่เลือกเป็นรายงานรูปแบบลำดับ ระบบจะออกรายงานรูปแบบลำดับ โดยรายงานนี้จะแสดงรูปแบบลำดับในเชิงตัวเลข 3a2) กรณีรายงานที่เลือกเป็นรายงานทางสถิติ ระบบจะออกรายงานทางสถิติให้ 3b1) กรณีรายงานที่เลือกเป็นรายงานรูปแบบลำดับ ระบบจะออกรายงานรูปแบบลำดับ โดยรายงานนี้จะมีการแสดงรูปแบบลำดับในรูปของ URL 3b) กรณีรายงานที่เลือกเป็นรายงานทางสถิติ ระบบแสดงข้อความเตือน 4a) ถ้า Analyst เลือกรายงานที่ไม่ได้กำหนดไว้ใน UC2 ระบบแสดงข้อความเตือน
Post Conditions:	ระบบสามารถออกรายงานได้ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่ควรเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 ขั้นตอนการทำงานของระบบ



รูปที่ 3.2 แผนภาพกิจกรรมของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากแผนภาพกิจกรรมรูปที่ 3.2 ระบบนำเข้าเว็บเซิร์ฟเวอร์ล็อกที่ยังไม่ผ่านการทำความสะอาดมาทำความสะอาดโดยการกำจัดรายการที่ไม่เกี่ยวข้องออก ซึ่งรายการที่ต้องกำจัดออกมีดังต่อไปนี้

1. กำจัดเว็บ ไรบอท ซึ่งเป็นตัวทำหน้าที่สำรวจเว็บไซต์ต่างๆ การเข้าถึงของเว็บ ไรบอทนั้น จะทำให้มีรายการ robot.txt ใน URI ของเส้นทางของไฟล์ที่ถูกกระทำ ซึ่งเราจะไม่ใช้รายการนี้ในการทำเหมืองข้อมูลการใช้งานเว็บ
2. กรองภาพและข้อมูลรบกวน โดยเว็บเพจปกติจะมีรูปภาพ เสียงและวิดีโอไฟล์ไปพร้อมกับข้อมูล โดยตรวจสอบส่วนต่อท้าย (Suffix) ของ URL ที่ลงท้ายด้วย gif, jpeg, GIF, JPEG, jpg, JPG, map, cgi, WAV, wav, CLASS, class, TXT, txt, java script (JS), page style file (css) (Cooley, R. et al. 1999 : 12)
3. บางครั้งผู้ร้องขอเว็บเพจที่ไม่มีอยู่ เว็บเซิร์ฟเวอร์จะสร้างรายการขึ้นมาพร้อมด้วยโค้ดที่เหมาะสม ซึ่งรายการล็อกนี้จะไม่นำมาใช้ในการวิเคราะห์
4. เว็บเซิร์ฟเวอร์บันทึกการเข้าถึงไฟล์ proxy.pac ซึ่งก็คือ ไฟล์โครงแบบอัตโนมัติ (Auto Configuration) ที่ใช้ในการกำหนดโครงแบบเว็บเบราว์เซอร์ของไคลเอนท์ รายการนี้จะถูกกรองออกไปเพื่อไม่ให้มีผลต่อการค้นหารูปแบบ

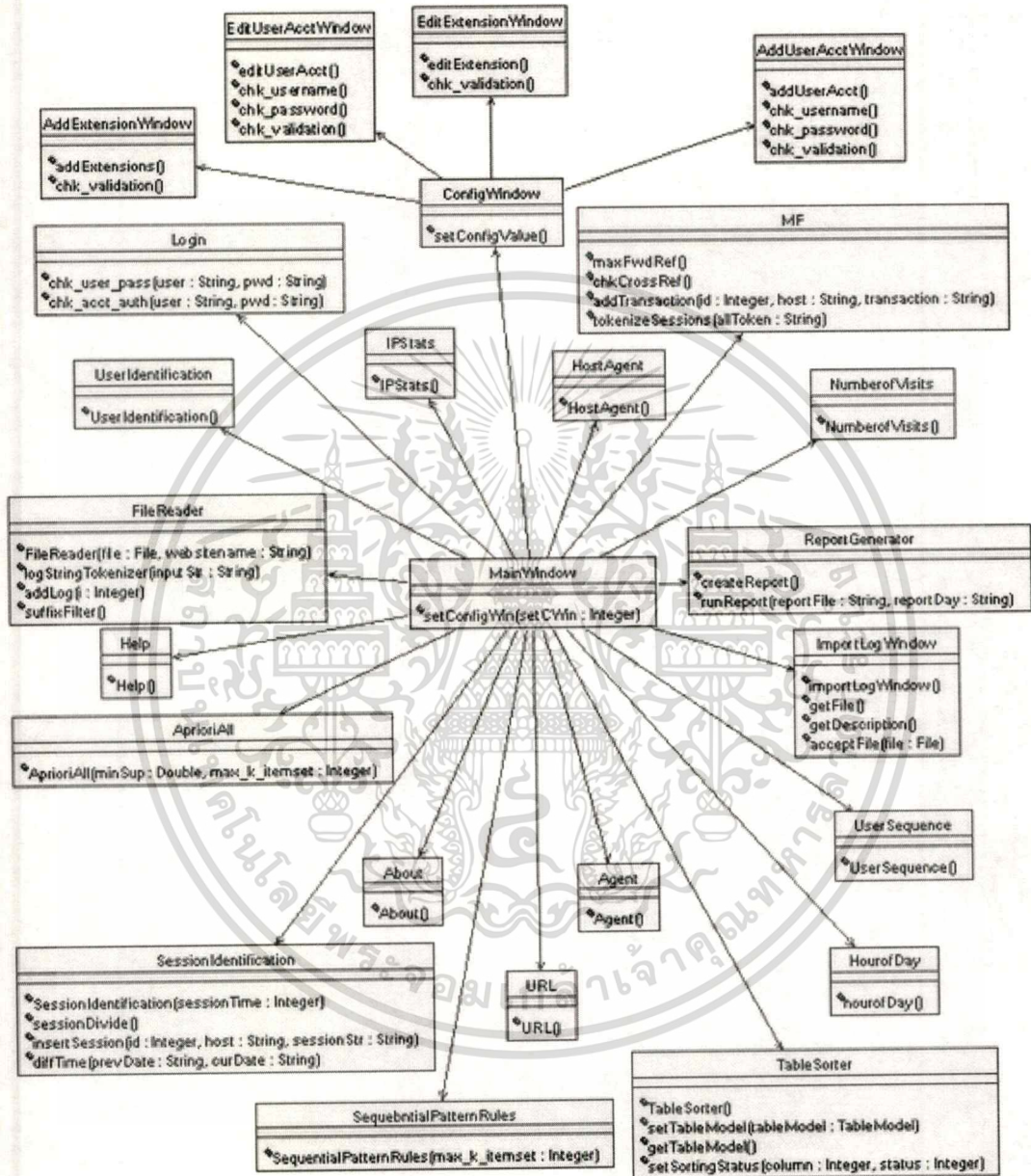
เมื่อทำความสะอาดเสร็จเรียบร้อยแล้ว จึงนำเว็บเซิร์ฟเวอร์ล็อกนี้มาระบุตัวผู้ใช้ ระบุเซชันผู้ใช้ โดยใช้เวลานอกที่ 30 นาที แล้วทำการระบุทรานแซกชัน โดยใช้อัลกอริทึมการอ้างอิงไปข้างหน้าไกลที่สุด เพื่อช่วยแก้ปัญหาผลกระทบที่เกิดจากการเดินทางย้อนกลับของผู้ใช้ โดยเมื่อเกิดการอ้างอิงย้อนกลับ การอ้างอิงไปข้างหน้าจะถือเป็นสิ้นสุด เมื่อระบุทรานแซกชันเสร็จเรียบร้อยแล้ว จะมาระบุลำดับผู้ใช้ โดยทรานแซกชันทั้งหมดของลูกค้านั่งซึ่งจัดเรียงเป็นลำดับ โดยแต่ละทรานแซกชันคือเซตของรายการและลำดับของทรานแซกชันถูกเรียงตามเวลาในทรานแซกชันเป็น T_1, T_2, \dots, T_n และกำหนดให้เซตของรายการใน T_i เป็น $itemset(T_i)$ แล้ว จะได้ว่าลำดับของผู้ใช้คือลำดับ $\langle itemset(T_1) \ itemset(T_2) \ \dots \ itemset(T_n) \rangle$

จากนั้นจะนำข้อมูลที่ได้จากแต่ละขั้นตอนข้างต้นมาทำเหมืองข้อมูล โดยในการทำเหมืองข้อมูลจะใช้ 2 วิธีคือ เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ เมื่อระบบทำเหมืองข้อมูลเสร็จ ผู้ใช้สามารถเลือกให้ระบบแสดงรายงานให้กับผู้ใช้ได้ โดยในการแสดงรายงานนี้ ระบบจะไปดึงรายงานที่ได้สร้างไว้ในฐานข้อมูล จากนั้นจึงแสดงรายงานในรูปแบบของตาราง ซึ่งรายงานเหล่านี้ ถ้านักวิเคราะห์ต้องการจัดเก็บก็สามารถทำได้ โดยใช้วิธีการออกรายงาน ซึ่งระบบสามารถออกรายงานได้ 2 วิธีคือ Export และ Translate จากนั้นระบบจะออกรายงานที่ต้องการ โดยที่นักวิเคราะห์สามารถเลือกพิมพ์รายงานหรือบันทึกรายงานได้ สำหรับการบันทึกรายงานนั้นสามารถบันทึกได้ในหลายรูปแบบด้วยกัน ได้แก่ PDF, RTF, HTML, XLS, CSV and XML

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3 คลาสไดอะแกรมของระบบ

หลังจากวิเคราะห์ยูสเคสไดอะแกรมแล้ว ขั้นตอนนี้จะวิเคราะห์รายละเอียดของข้อมูลที่ใช้ในระบบ โดยจะสร้างเป็นคลาสไดอะแกรมแสดงดังรูปที่ 3.3



รูปที่ 3.3 คลาสไดอะแกรมของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำอธิบายหน้าที่รับผิดชอบและการทำงานร่วมกันของแต่ละคลาส

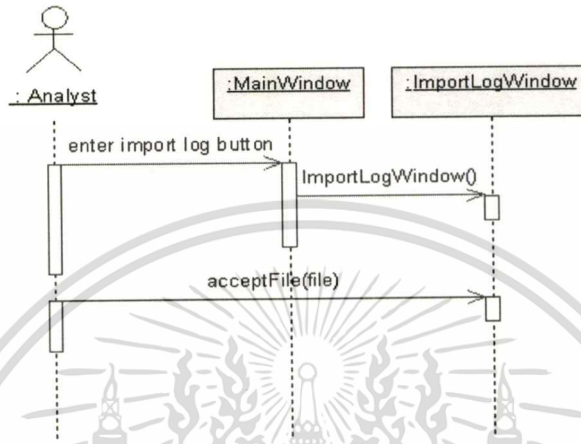
- AddExtensionWindow เป็นส่วนต่อประสานทำหน้าที่เพิ่มรายการ Extensions หรือ Suffix ที่ใช้ในการทำความสะอาดเว็บเบราว์เซอร์ลือก
- EditExtensionWindow เป็นส่วนต่อประสานทำหน้าที่แก้ไขรายการ Extensions หรือ Suffix ที่ใช้ในการทำความสะอาด
- AddUserAcctWindow เป็นส่วนต่อประสานทำหน้าที่เพิ่มรายการนักวิเคราะห์เหมืองข้อมูลในระบบ
- EditUserAcctWindow เป็นส่วนต่อประสานทำหน้าที่แก้ไขรายการนักวิเคราะห์เหมืองข้อมูลในระบบ
- Login เป็นส่วนต่อประสานทำหน้าที่รองรับการล็อกอินเข้าสู่ระบบ
- ImportLogWindow เป็นส่วนต่อประสานทำหน้าที่นำเข้าเว็บเบราว์เซอร์ลือกไฟล์ที่ใช้ในการทำเหมืองข้อมูล
- ConfigWindow เป็นส่วนต่อประสานทำหน้าที่กำหนดค่าและจัดการข้อมูลที่ใช้ในระบบ
- FileReader ทำหน้าที่อ่านเว็บเบราว์เซอร์ลือกเข้าสู่ระบบ
- UserIdentification ทำหน้าที่ระบุตัวผู้ใช้โดยใช้โฮสต์และเอเจนต์เป็นตัวระบุ
- SessionIdentification ทำหน้าที่ระบุเซสชันผู้ใช้โดยอาศัยเวลาออก
- MF ทำหน้าที่ระบุทรานแซกชันผู้ใช้โดยใช้อัลกอริทึมการอ้างอิงไปยังหน้าไกลที่สุด
- AprioriAll ทำหน้าที่ใช้ค้นหารูปแบบลำดับ
- TableSorter ทำหน้าที่เรียงลำดับข้อมูลที่นำมาแสดงในตารางที่อยู่ใน MainWindow
- SequentialPatternRules ทำหน้าที่สร้างกฎรูปแบบลำดับ
- ReportGenerator ทำหน้าที่สร้างรายงานที่ได้จากการทำเหมืองข้อมูลเว็บเบราว์เซอร์ลือก
- UserSequence ทำหน้าที่สร้างลำดับผู้ใช้เพื่อใช้ค้นหารูปแบบลำดับในอัลกอริทึม AprioriAll
- Agent ทำหน้าที่สร้างเอเจนต์ผู้ใช้
- HostAgent ทำหน้าที่สร้างตาราง HostAgent เพื่อใช้ในการระบุตัวผู้ใช้
- URL ทำหน้าที่สร้างตาราง Url ที่มีการเชื่อมโยงในเว็บไซค์
- NumberofVisits ทำหน้าที่สร้างสถิติแยกตามจำนวนผู้เยี่ยมชมเว็บไซค์
- HourofDay ทำหน้าที่สร้างสถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง
- Help เป็นส่วนต่อประสานทำหน้าที่ช่วยให้คำแนะนำในการใช้งานระบบ
- About เป็นส่วนต่อประสานทำหน้าที่แสดงคำอธิบายเกี่ยวกับระบบ
- MainWindow เป็นส่วนต่อประสานทำหน้าที่เป็นหน้าจอหลักที่ใช้ในการทำเหมืองข้อมูล

เอกสารนี้เป็นเว็บเบราว์เซอร์ลือกหรือการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.4 ซีควেনซ์ไดอะแกรมของระบบ

ซีควেনซ์ไดอะแกรมอธิบายลำดับการทำงานของอ็อบเจกต์ของคลาสภายในคลาสไดอะแกรมรูปที่ 3.4 แสดงดังต่อไปนี้

3.3.4.1 ซีควেনซ์ไดอะแกรมการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์

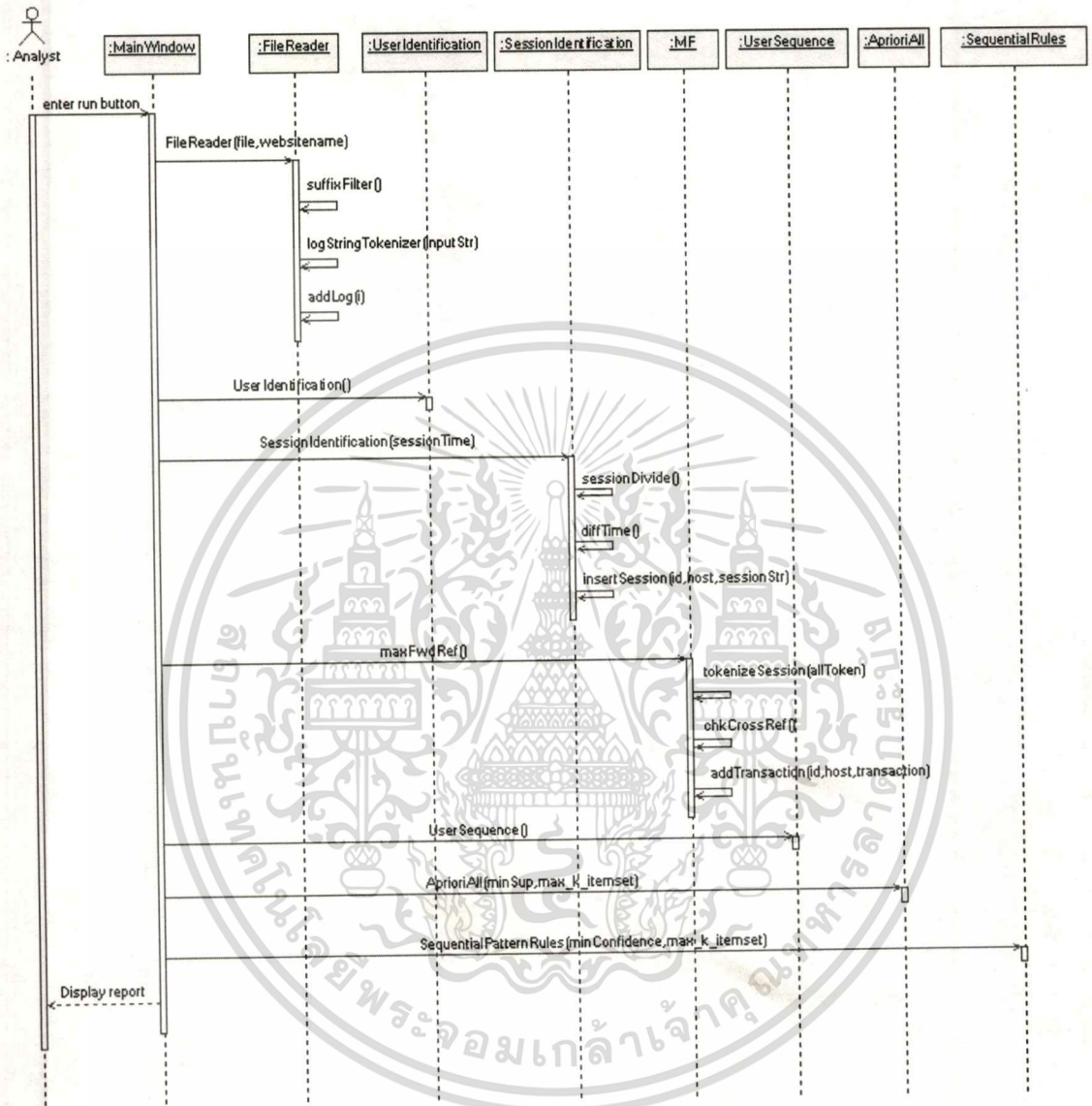


รูปที่ 3.4 ซีควেনซ์ไดอะแกรมการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์

จากรูปที่ 3.4 นักวิเคราะห์นำเข้าเว็บเซิร์ฟเวอร์ล็อกที่ยังไม่ผ่านการทำความสะอาด โดยการเลือกที่เมนูหรือปุ่ม Import Log ที่หน้าจอ MainWindow จากนั้นอ็อบเจกต์ MainWindow จะไปเรียกอ็อบเจกต์ ImportLogWindow ให้แสดงหน้าจอการนำเข้า นักวิเคราะห์จึงนำเข้าเว็บเซิร์ฟเวอร์ล็อก

ระบบจะรองรับเว็บเซิร์ฟเวอร์ล็อกไฟล์ของการเข้าถึงเว็บเซิร์ฟเวอร์แบบคอมไบน์หรือเอ็กซ์เทนด์เท่านั้นที่ใช้ประมวลผล ระบบจะประมวลผลกับเว็บเซิร์ฟเวอร์ล็อกที่บันทึกการเข้าใช้งานใน 1 วันเท่านั้น

3.3.4.2 ซีควนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์โดยใช้เทคนิค รูปแบบลำดับ



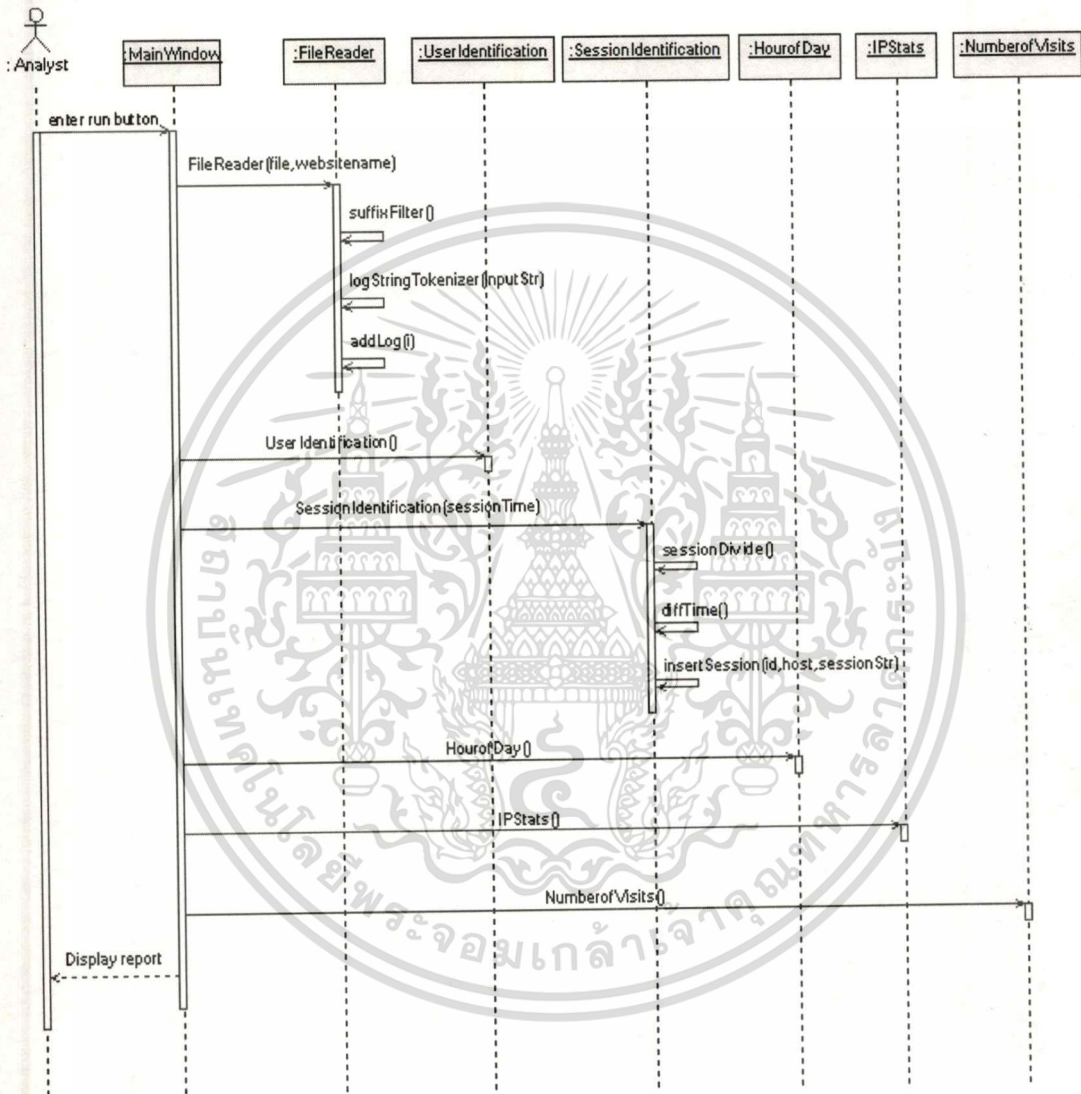
รูปที่ 3.5 ซีควนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์โดยใช้เทคนิครูปแบบลำดับ

จากรูปที่ 3.5 หลังจากที่นักวิเคราะห์นำเข้าเว็บเซิร์ฟเวอร์ล็อกเรียบร้อยแล้ว นักวิเคราะห์จะสั่งให้ระบบดำเนินการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้เทคนิครูปแบบลำดับ โดยเลือกที่เมนู Run -> Execute หรือปุ่ม Run ที่หน้าจอ MainWindow จากนั้นระบบจึงนำไฟล์ที่นำเข้ามาทำ ความสะอาดโดยอ็อบเจกต์ FileReader เมื่อทำความสะอาดข้อมูลเสร็จ ก็จะเอาข้อมูลที่ได้นำมาระบุตัว ผู้ใช้ เซสชัน ทรานแซกชัน และลำดับผู้ใช้ โดยใช้อ็อบเจกต์ UserIdentification อ็อบเจกต์ SessionIdentification อ็อบเจกต์ MF และอ็อบเจกต์ UserSequence ตามลำดับ จากนั้นจะนำลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้เข้ามาค้นหาในรูปแบบลำดับโดยใช้โออบเจกต์ AprioriAll สุดท้ายจะนำรูปแบบลำดับที่ได้มาสร้างเป็นรูปแบบลำดับโดยใช้โออบเจกต์ SequentialRules

3.3.4.3 ซีควนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยการวิเคราะห์ทางสถิติ

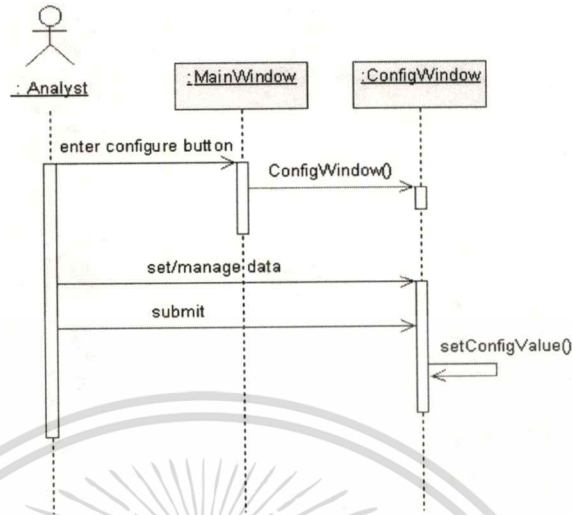


รูปที่ 3.6 ซีควนซ์ไดอะแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยการวิเคราะห์ทางสถิติ

จากรูปที่ 3.6 ในการวิเคราะห์ทางสถิติจะใช้ข้อมูลที่ได้จากการทำความสะอาด ข้อมูลที่ได้จากการระบุตัวผู้ใช้ และข้อมูลที่ได้จากการระบุเซชันผู้ใช้ มาวิเคราะห์ทางสถิติ ดังนั้น ปฏิสัมพันธ์ระหว่างโออบเจกต์ต่างๆในรูปนี้จะคล้ายกับซีควนซ์ไดอะแกรมในรูปที่ 3.5 ในส่วนของการนำเข้าเว็บเซิร์ฟเวอร์ล็อก การระบุตัวผู้ใช้ และการระบุเซชัน จากนั้นก็จะเป็นการวิเคราะห์ทางสถิติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

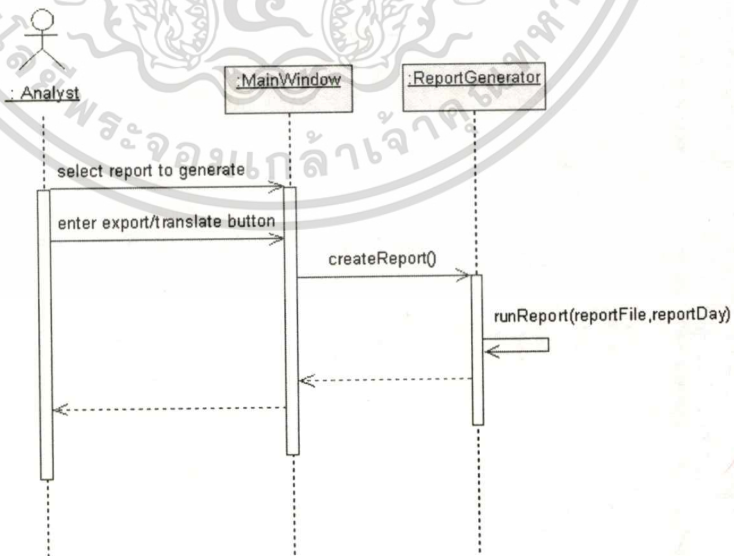
3.3.4.4 ซีควเอนซ์ไดอะแกรมการกำหนดค่าและจัดการข้อมูลให้กับระบบ



รูปที่ 3.7 ซีควเอนซ์ไดอะแกรมการกำหนดค่าและจัดการข้อมูลให้กับระบบ

จากรูปที่ 3.7 นักวิเคราะห์กำหนดค่าและจัดการข้อมูลให้กับระบบที่หน้าจอ ConfigWindow ซึ่งข้อมูลเหล่านี้จะถูกบันทึกไว้ โดยอ็อบเจกต์ ConfigWindow เพื่อใช้ในการทำเหมืองข้อมูล

3.3.4.5 ซีควเอนซ์ไดอะแกรมการออกรายงานของระบบ



รูปที่ 3.8 ซีควเอนซ์ไดอะแกรมการออกรายงานของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.8 นักวิเคราะห์ออกรายงาน โดยการเลือกรายงานที่ต้องการในหน้าจอ MainWindow และเลือกวิธีการออกรายงาน จากนั้นอ็อบเจกต์ของคลาส ReportGenerator จะทำหน้าที่สร้างและแสดงรายงาน

3.4 การออกแบบฐานข้อมูล

ฐานข้อมูลของระบบที่ใช้ในการวิเคราะห์ข้อมูลต่างๆ ที่จำเป็นต้องใช้ในการเข้ามาทำเหมืองข้อมูลการใช้งานประกอบไปด้วยข้อมูลหลักดังนี้

- รายชื่อสำหรับการล็อกอินเข้าสู่ระบบ
- ส่วนต่อท้ายที่ใช้ในการทำความสะอาดเว็บเซิร์ฟเวอร์ล็อกโดยการกรอกรายการที่ไม่เกี่ยวข้องออก
- เว็บเซิร์ฟเวอร์ล็อกที่ผ่านการกำจัดรายการที่ไม่เกี่ยวข้องออกแล้ว
- เว็บเซิร์ฟเวอร์ล็อกที่มีการจัดรูปเพื่อเตรียมพร้อมสำหรับการระบุตัวผู้ใช้
- ผู้ใช้แต่ละคนซึ่งได้จากการระบุตัวผู้ใช้จากเว็บเซิร์ฟเวอร์ล็อกที่มีการจัดรูปเพื่อเตรียมพร้อมสำหรับการระบุตัวผู้ใช้
- เซสชันผู้ใช้ซึ่งได้จากการระบุเซสชันผู้ใช้จากเว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุตัวผู้ใช้แล้ว
- ทรานแซกชันผู้ใช้ซึ่งได้จากการระบุทรานแซกชันผู้ใช้จากเซสชันผู้ใช้
- ลำดับผู้ใช้ซึ่งได้จากการระบุลำดับผู้ใช้จากทรานแซกชันผู้ใช้
- รูปแบบลำดับซึ่งได้จากการค้นหารูปแบบลำดับจากลำดับผู้ใช้
- สถิติต่างๆที่ได้จากการวิเคราะห์

ดังนั้นเพื่อให้รองรับการทำงานของระบบจึงต้องการเอนทิตีทั้งหมด 18 เอนทิตี ดังต่อไปนี้

1. เอนทิตี Accounts คือ นักวิเคราะห์เหมืองข้อมูลในระบบ
2. เอนทิตี Extensions คือ Suffix ที่ใช้ในการทำความสะอาดเว็บเซิร์ฟเวอร์ล็อก
3. เอนทิตี Logfile คือ เว็บเซิร์ฟเวอร์ล็อกที่ได้ผ่านการกรองแล้ว
4. เอนทิตี Cleanedlog คือ เว็บเซิร์ฟเวอร์ล็อกที่ได้ผ่านการระบุตัวผู้ใช้แล้ว
5. เอนทิตี Cleanedlog_prep คือ เว็บเซิร์ฟเวอร์ล็อกที่ผ่านการระบุตัวผู้ใช้แล้ว และได้จัดรูปแบบให้เหมาะสม เพื่อใช้ในการระบุเซสชันผู้ใช้
6. เอนทิตี URL คือ เส้นทางของไฟล์ที่ถูกกระทำ
7. เอนทิตี Agents คือ เอเจนต์ซึ่งประกอบด้วยเว็บเบราว์เซอร์ และแพลตฟอร์มที่ใช้ในการเยี่ยมชม
8. เอนทิตี Hostagent คือ ผู้ใช้แต่ละคนที่ระบุด้วยไอพีแอดเดรสและเอเจนต์
9. เอนทิตี Sessions คือ รายการเซสชันของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

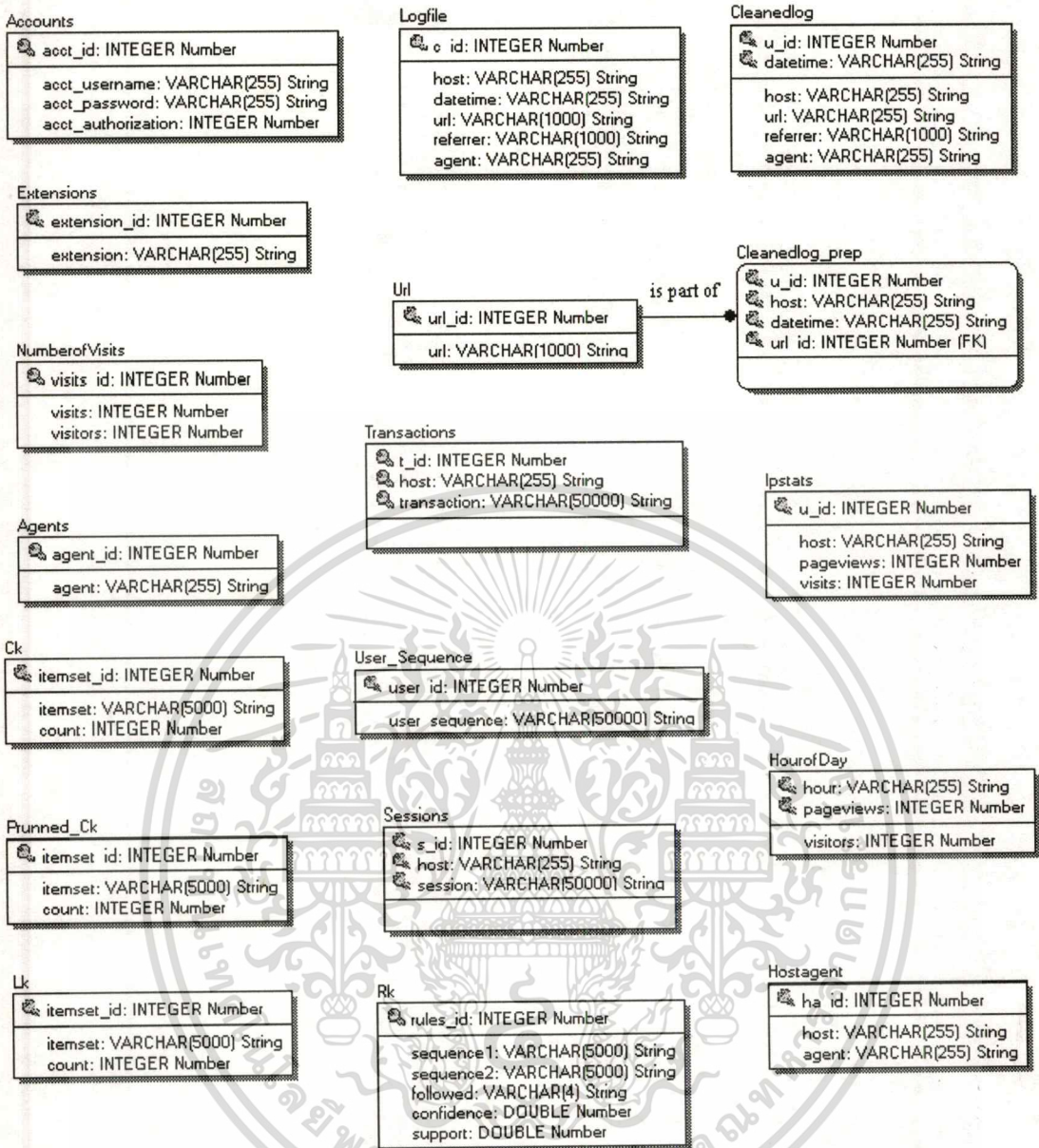
10. เอนทิตี Transactions คือรายการทรานแซกชันของผู้ใช้
11. เอนทิตี User_Sequence คือ ลำดับผู้ใช้ ซึ่งประกอบไปด้วยทรานแซกชันผู้ใช้ที่จัดเรียงตามลำดับเวลา
12. เอนทิตี Ck คือ เซตลำดับคู่แข่งที่มีสมาชิก k ตัว
13. เอนทิตี Prunned_Ck คือ เซตลำดับคู่แข่งที่มีสมาชิก k ตัว ที่ได้ผ่านการตัดเล่มแล้ว
14. เอนทิตี Lk คือ เซตของลำดับขนาดใหญ่ที่มีสมาชิก k ตัว
15. เอนทิตี Rk คือ กฎรูปแบบลำดับ
16. เอนทิตี HourofDay คือ สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง
17. เอนทิตี Ipstats คือ สถิติแยกตาม ไอพีแอดเดรสของผู้ใช้
18. เอนทิตี NumberofVisits คือ สถิติแยกตามจำนวนการเข้าชมเว็บไซต์

เอนทิตีเหล่านี้ไม่ได้มีความสัมพันธ์กันโดยตรง จึงไม่มีการเชื่อมโยงระหว่างเอนทิตีต่างๆ ยกเว้น เอนทิตี CleanedLog_prep และเอนทิตี Url ซึ่งจะมีการเชื่อมโยงกันด้วยความสัมพันธ์ที่ว่า ผู้ใช้คนหนึ่ง สามารถเยี่ยมชมไปได้หลาย Url ในส่วนเอนทิตีที่เหลือที่ไม่ได้มีความสัมพันธ์กันโดยตรงเพราะเอนทิตีเหล่านี้เป็นเอนทิตีที่เกิดขึ้นมา โดยการนำอีกเอนทิตีมาทำการประมวลผล และจัดรูปแบบให้เหมาะสมกับการประมวลผลในขั้นตอนต่างๆ การเกิดขึ้นของเอนทิตีต่างๆแสดงดังนี้

- เอนทิตี Logfile เป็นเอนทิตีที่ใช้สร้างเอนทิตี CleanedLog เอนทิตี Agent เอนทิตี HostAgent และเอนทิตี Url
- เอนทิตี HostAgent และ เอนทิตี Logfile เป็นเอนทิตีที่ใช้สร้างเอนทิตี CleanedLog_prep
- เอนทิตี CleanedLog_prep เป็นเอนทิตีที่ใช้สร้างเอนทิตี Sessions
- เอนทิตี Sessions ใช้สร้างเอนทิตี Transactions
- เอนทิตี Transactions ใช้สร้างเอนทิตี User_Sequence
- เอนทิตี User_Sequence ใช้สร้างเอนทิตี Lk เอนทิตี Ck และ Prunned_Ck
- เอนทิตี Lk ใช้สร้างเอนทิตี Rk
- เอนทิตี CleanedLog_prep ใช้สร้างเอนทิตี HourofDay
- เอนทิตี CleanedLog_prep และเอนทิตี Sessions ใช้สร้างเอนทิตี Ipstats
- เอนทิตี Sessions ใช้สร้างเอนทิตี NumberofVisits
- ส่วนเอนทิตี Accounts และเอนทิตี Extensions นั้นไม่ได้เกิดจากเอนทิตีใดๆ

ฐานข้อมูลของระบบประกอบด้วยตารางต่างๆแสดงดังรูปที่ 3.9 และพจนานุกรมข้อมูลแสดงดังตารางที่ 3.7 – 3.24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 ฐานข้อมูลของระบบ

ในขั้นตอนการเตรียมข้อมูลการใช้งาน จะใช้เอนทิตีต่างๆดังนี้ เอนทิตี Logfile, CleanedLog, CleanedLog_prep, Agent, Extensions, Url, HostAgent, Sessions, Transactions, User_Sequence ส่วนในขั้นตอนการค้นหารูปแบบ จะใช้เอนทิตีต่างๆดังนี้ เอนทิตี Lk, Ck, Prunned_Ck, Rk, HourofDay, NumberofVisits และ IpStats ส่วนในการวิเคราะห์รูปแบบ จะนำผลลัพธ์ที่ได้จากขั้นตอนการค้นหารูปแบบมาวิเคราะห์ แปลผล และนำไปประยุกต์ใช้

ตารางที่ 3.7 ตาราง Accounts

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
acct_id	รหัสผู้ใช้โปรแกรม	INTEGER(4)	PK	
acct_username	ชื่อผู้ใช้ระบบ	VARCHAR(255)		
acct_password	รหัสผ่าน	VARCHAR(255)		
acct_authorization	สิทธิ์การเข้าใช้งาน	INTEGER(4)		

ตารางที่ 3.8 ตาราง Extensions

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
extension_id	รหัสส่วนต่อท้าย	INTEGER(4)	PK	
extension	ส่วนต่อท้าย	VARCHAR(255)		

ตารางที่ 3.9 ตาราง Logfile

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
c_id	รหัสล็อกที่ผ่านการทำ ความสะอาดแล้ว	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)		
datetime	วันและเวลาที่เยี่ยมชม	VARCHAR(255)		
url	เส้นทางของไฟล์ที่ถูก กระทำ	VARCHAR(1000)		
agent	เอเจนต์	VARCHAR(255)		

ตารางที่ 3.10 ตาราง Cleanedlog

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
u_id	รหัสผู้เยี่ยมชมหลังทำ ความสะอาดแล้ว	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)		
datetime	วันและเวลาที่เยี่ยมชม	VARCHAR(255)	PK	
url	เส้นทางของไฟล์ที่ถูก กระทำ	VARCHAR(1000)		
agent	เอเจนต์	VARCHAR(255)		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.11 ตาราง Cleanedlog_prep

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
u_id	รหัสผู้เยี่ยมชมหลังทำ ความสะอาดแล้ว	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)	PK	
datetime	วันและเวลาที่เยี่ยมชม	VARCHAR(255)	PK	
url_id	รหัสเส้นทางของไฟล์ ที่ถูกกระทำ	INTEGER(4)	FK	Url

ตารางที่ 3.12 ตาราง Url

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
url_id	รหัสเส้นทางของไฟล์ ที่ถูกกระทำ	INTEGER(4)	PK	
url	เส้นทางของไฟล์ที่ถูก กระทำ	VARCHAR(1000)		

ตารางที่ 3.13 ตาราง Agents

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
agent_id	รหัสเอเจนต์	INTEGER(4)	PK	
agent	เอเจนต์	VARCHAR(255)		

ตารางที่ 3.14 ตาราง HostAgent

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
ha_id	รหัสผู้เยี่ยมชม	INTERGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)		
agent	เอเจนต์	VARCHAR(255)		

ตารางที่ 3.15 ตาราง UserSequence

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
user_id	รหัสลำดับผู้เยี่ยมชม	INTEGER(4)	PK	
user_sequence	ลำดับผู้เยี่ยมชม	VARCHAR(50000)		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.16 ตาราง Sessions

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
s_id	รหัสเซสชัน	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)	PK	
Session	เซสชัน	VARCHAR(50000)	PK	

ตารางที่ 3.17 ตาราง Transactions

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
t_id	รหัสทรานแซกชัน	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)	PK	
transaction	ทรานแซกชัน	VARCHAR(50000)	PK	

ตารางที่ 3.18 ตาราง Ck

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
itemset_id	รหัสเซตของรายการ	INTEGER(4)	PK	
itemset	เซตของรายการ	VARCHAR(5000)		
count	ค่านับสนุน	INTEGER(255)		

ตารางที่ 3.19 ตาราง Pruned_Ck

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
itemset_id	รหัสเซตของรายการ	INTEGER(4)	PK	
itemset	เซตของรายการ	VARCHAR(5000)		
count	ค่านับสนุน	INTEGER(255)		

ตารางที่ 3.20 ตาราง Lk

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
itemset_id	รหัสเซตของรายการ	INTEGER(4)	PK	
itemset	เซตของรายการ	VARCHAR(5000)		
count	ค่านับสนุน	INTEGER(255)		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.21 ตาราง Rk

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
rules_id	รหัสรูปแบบลำดับ	INTEGER(4)	PK	
sequence1	ลำดับที่1	VARCHAR(5000)		
followed	ตามด้วย	VARCHAR(4)		
sequence2	ลำดับที่2	VARCHAR(5000)		
confidence	ค่าความเชื่อมั่น	DOUBLE(5,2)		
support	ค่าสนับสนุน	DOUBLE(5,2)		

ตารางที่ 3.22 ตาราง HourOfDay

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
hour	ช่วงเวลาที่เยี่ยมชม	INTEGER(4)	PK	
pageviews	จำนวนครั้งของการเยี่ยมชม	INTEGER(255)		
visitors	จำนวนผู้ชม	INTEGER(255)		

ตารางที่ 3.23 ตาราง IpStats

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
u_id	รหัสผู้เยี่ยมชม	INTEGER(4)	PK	
host	หมายเลขเครื่อง	VARCHAR(255)		
pageviews	จำนวนครั้งของการเยี่ยมชม	INTEGER(255)		
visits	จำนวนการเยี่ยมชมจริง	INTEGER(255)		

ตารางที่ 3.24 ตาราง NumberOfVisits

ชื่อแอตทริบิวต์	คำอธิบาย	ชนิดข้อมูล	คีย์	ตารางอ้างอิง
visits_id	รหัสผู้เยี่ยมชม	INTEGER(4)	PK	
visits	ระดับครั้งการเยี่ยมชม	INTEGER(255)		
visitors	จำนวนผู้เยี่ยมชม	INTEGER(255)		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงวิธีการทดลองและผลการทดลอง โดยใช้โปรแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อก ที่ได้พัฒนาขึ้นจากแบบจำลองที่ได้นำเสนอไปแล้วในบทที่ 3 เพื่อใช้ในการทดลอง ส่วนวิธีการใช้งาน โปรแกรมแสดงในภาคผนวก ค

4.1 สภาพแวดล้อมในการพัฒนาระบบ

โปรแกรมนี้พัฒนาในลักษณะของวินโดวส์แอปพลิเคชัน สภาพแวดล้อมของการพัฒนาระบบประกอบด้วย 2 ส่วน คือ รายละเอียดทางด้านฮาร์ดแวร์และซอฟต์แวร์ ซึ่งสามารถสรุปได้ดังนี้

1. รายละเอียดทางด้านฮาร์ดแวร์ประกอบด้วย
 - เครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผลกลางเป็น Pentium® โดยทำงานที่ 2.00 GHz
 - หน่วยความจำขนาด 1 GB
2. รายละเอียดทางด้านซอฟต์แวร์ประกอบด้วย
 - ระบบปฏิบัติการ Windows XP
 - โปรแกรม NetBeans 5.5
 - Java Development Kit 1.5.0_09
 - โปรแกรม MySQL Server 5.0
 - โปรแกรม iReport 1.2.4

4.2 วิธีการทดลอง

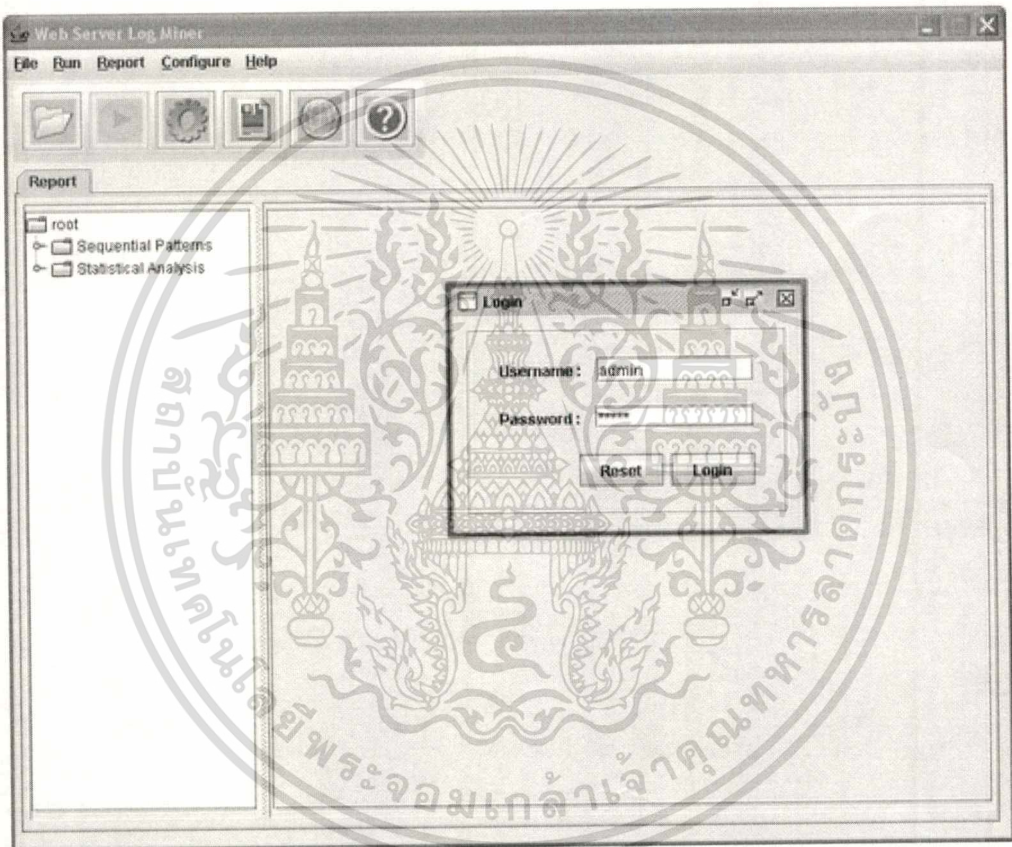
ในการทดลอง ผู้ทดลองจะนำเว็บเซิร์ฟเวอร์ล็อกที่บันทึกการเข้าใช้งานที่เว็บไซต์ <http://www.differentasia.com> เป็นข้อมูลนำเข้าของโปรแกรม จากนั้นมีการกำหนดค่าและจัดการข้อมูลต่างๆที่นำมาใช้ในการทำเหมืองข้อมูลก่อน เช่น ข้อมูลส่วนต่อท้ายที่ใช้ในการทำความสะอาดรายงานที่ต้องการ ชื่อของเว็บไซต์ ค่าสนับสนุนขั้นต่ำ ค่าความเชื่อมั่นขั้นต่ำ และค่า k-itemset ที่มากที่สุด แล้วจึงสั่งให้โปรแกรมทำการประมวลผล หลังจากทีโปรแกรมประมวลผลเสร็จก็จะได้รับรายงานตามที่ได้กำหนดไว้ วิธีการทดลองดังกล่าวมีขั้นตอนแสดงดังต่อไปนี้

1. ล็อกอินเข้าสู่โปรแกรมแสดงดังรูปที่ 4.1
2. นำเข้าเว็บเซิร์ฟเวอร์ล็อกของวันที่ 14 พฤษภาคม 2549 แสดงดังรูปที่ 4.2 และ 4.3

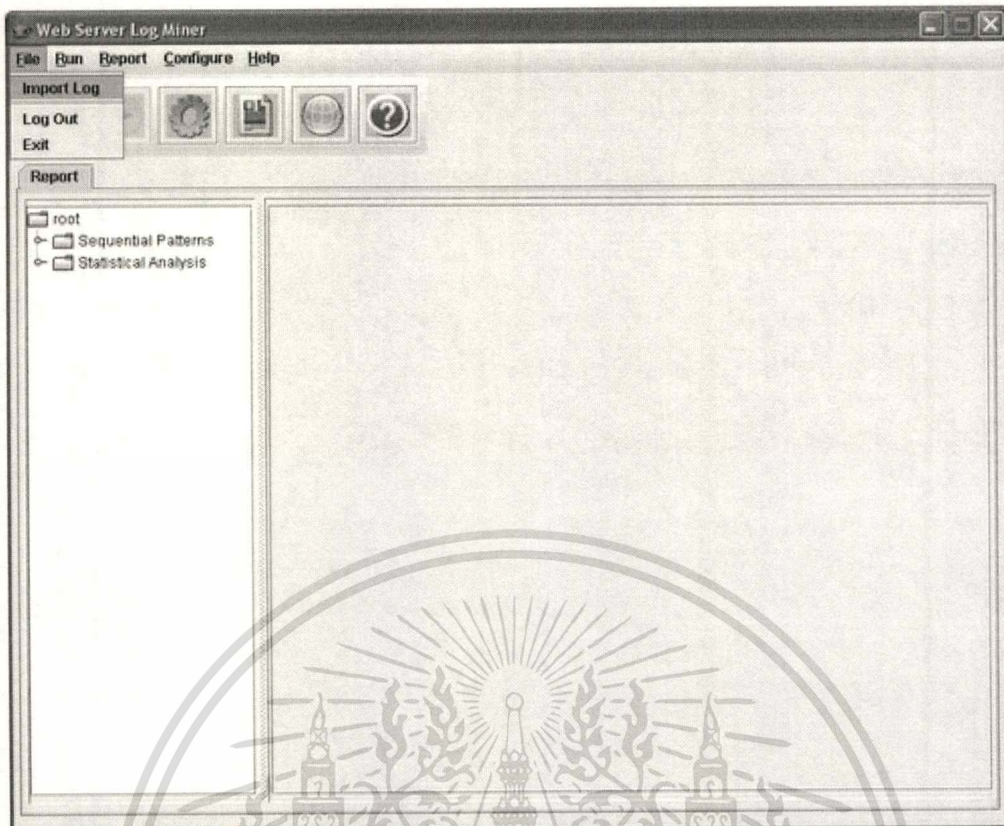
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

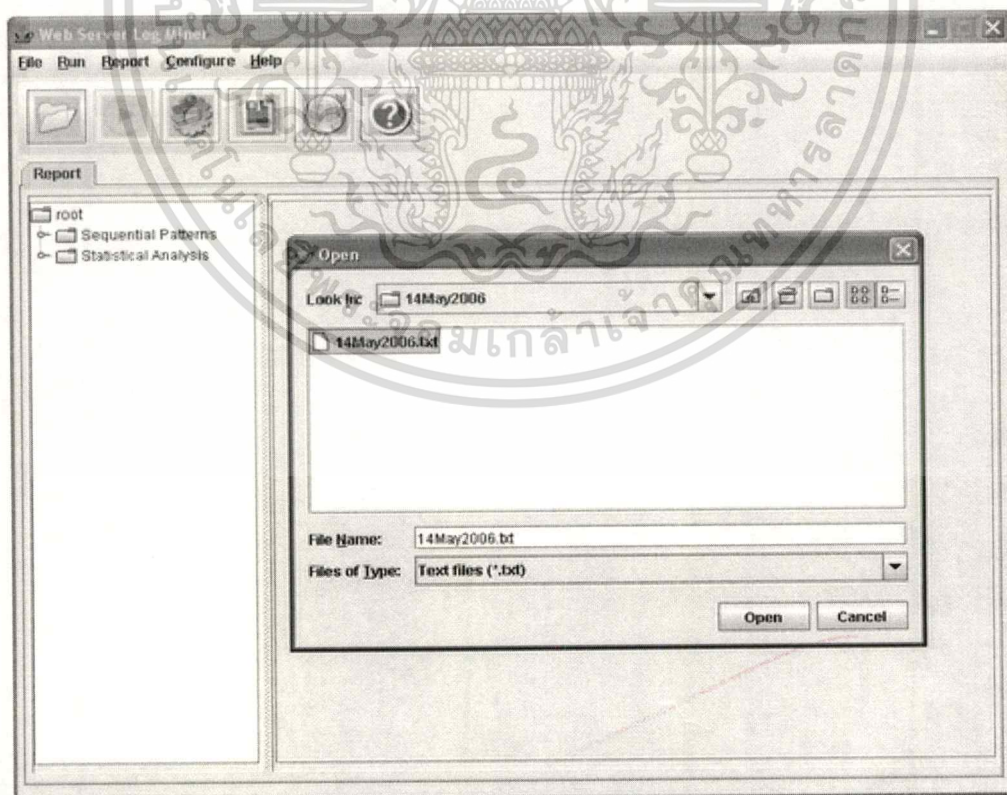
3. กำหนดค่าและจัดการข้อมูลที่ต้องใช้ในการทดลอง
 - กำหนดค่าเซสชันที่ 30 นาที แสดงดังรูปที่ 4.4
 - กำหนดค่า Minimum Support เท่ากับ 1.5 % ค่า Minimum Confidence เท่ากับ 80% และ ค่า Maximum k-Itemset เท่ากับ 4 แสดงดังรูปที่ 4.5
4. กำหนดชื่อเว็บไซต์และรายงานที่ต้องการให้โปรแกรมสร้างแสดงดังรูปที่ 4.6 ในการทดลองนี้ กำหนดให้โปรแกรมแสดงรายงานทุกชนิด
5. สั่งให้โปรแกรมทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อก แสดงดังรูปที่ 4.7



รูปที่ 4.1 หน้าจอล็อกอินเข้าสู่โปรแกรม

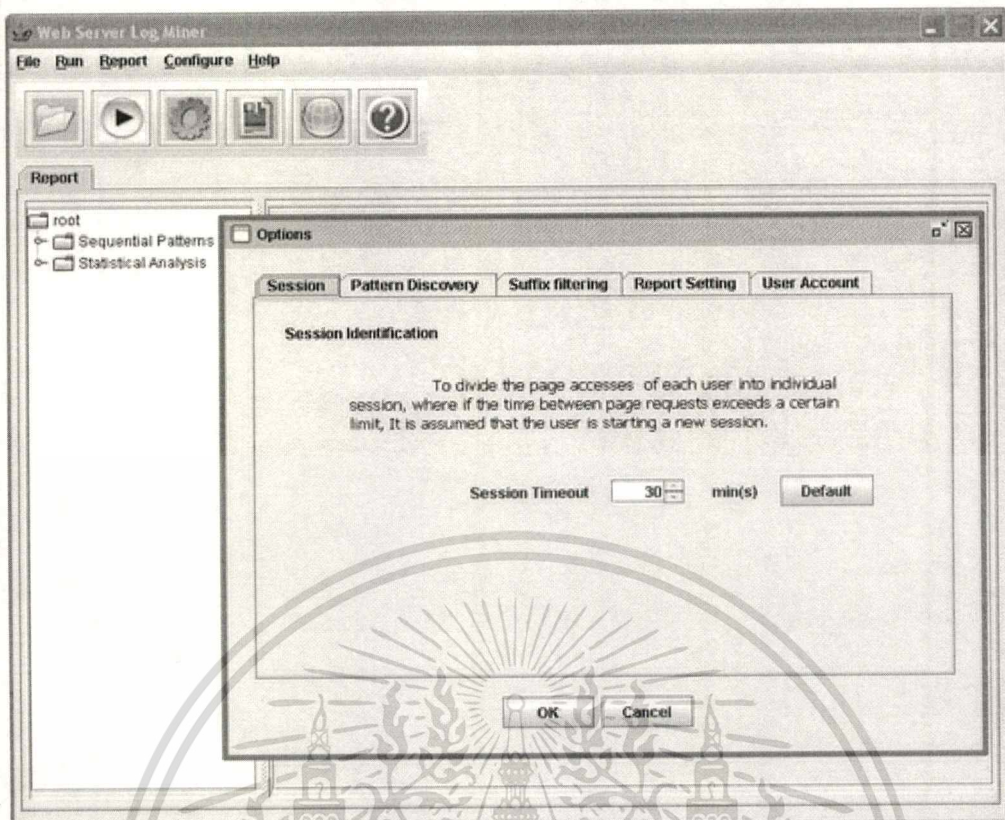


รูปที่ 4.2 หน้าจอการนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์

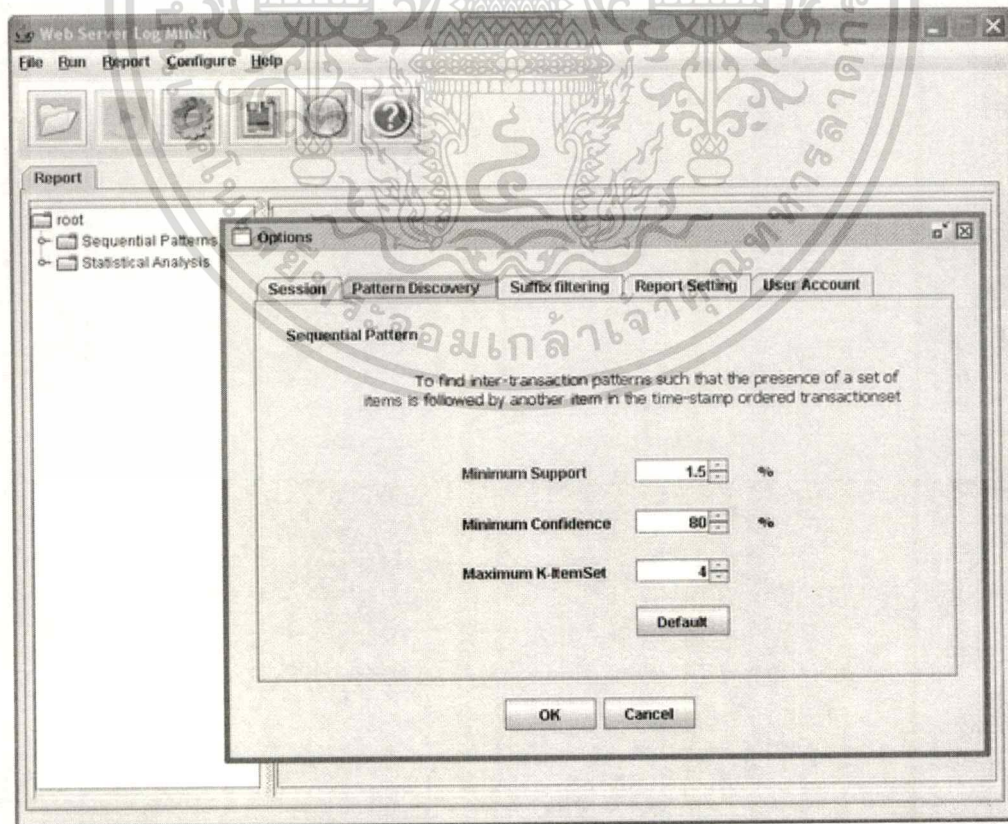


รูปที่ 4.3 หน้าจอการเลือกไฟล์ที่ต้องการนำเข้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

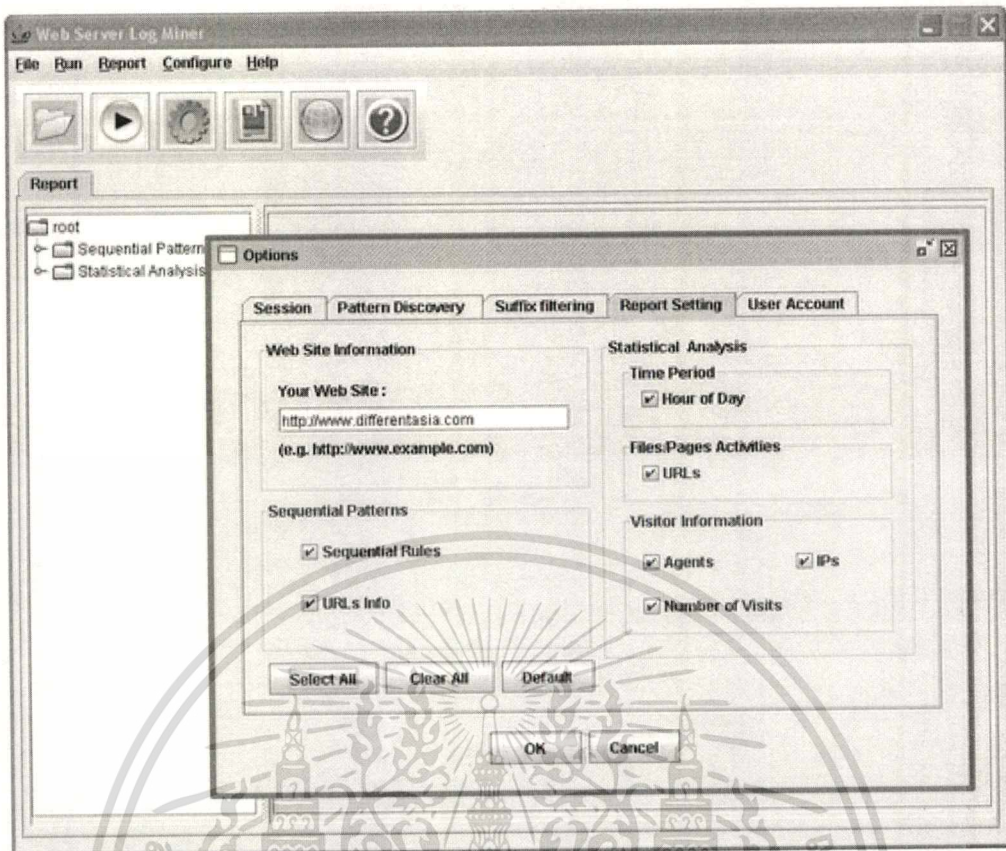


รูปที่ 4.4 หน้าจอการกำหนดค่าเวลาออก

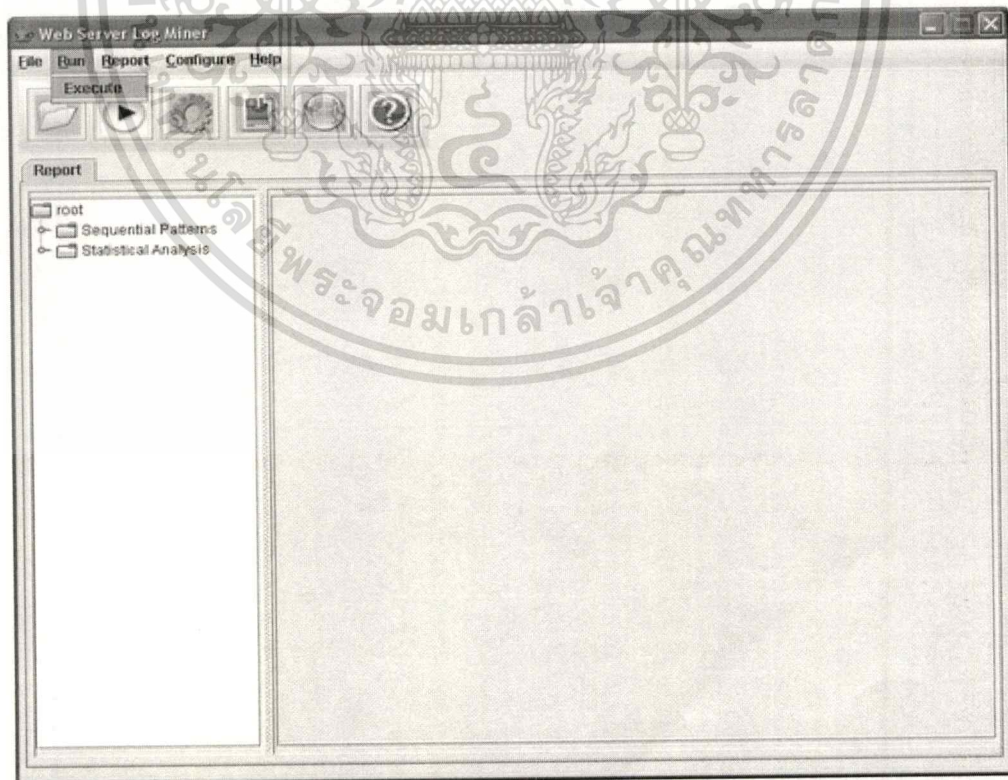


รูปที่ 4.5 หน้าจอการกำหนดค่าที่ใช้ในการค้นหารูปแบบลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 หน้าจอการกำหนดชื่อเว็บไซต์และรายงานที่ต้องการ



รูปที่ 4.7 หน้าจอการสั่งให้โปรแกรมประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลการทดลอง

หลังจากที่โปรแกรมประมวลผลเสร็จได้ผลการทดลอง แบ่งออกเป็นสองส่วนคือ รายงานรูปแบบลำดับและรายงานทางสถิติ ดังนี้

4.3.1 รายงานรูปแบบลำดับ

รูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 2 ตัว

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 80.0 %

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
1	{38}	==>	{4}	80.0	1.66
2	{38}	==>	{63}	80.0	1.66
3	{38}	==>	{84}	80.0	1.66
4	{38}	==>	{89}	80.0	1.66
5	{49}	==>	{4}	80.0	1.66
6	{49}	==>	{5}	80.0	1.66
7	{49}	==>	{15}	80.0	1.66
8	{49}	==>	{52}	80.0	1.66
9	{49}	==>	{54}	80.0	1.66
10	{50}	==>	{4}	100.0	1.66
11	{50}	==>	{5}	100.0	1.66
12	{50}	==>	{15}	100.0	1.66
13	{50}	==>	{52}	100.0	1.66
14	{50}	==>	{57}	100.0	1.66
15	{50}	==>	{62}	100.0	1.66
16	{50}	==>	{84}	100.0	1.66
17	{54}	==>	{4}	100.0	1.66
18	{56}	==>	{4}	100.0	1.66
19	{56}	==>	{38}	100.0	1.66
20	{56}	==>	{61}	100.0	1.66
21	{56}	==>	{62}	100.0	1.66
22	{56}	==>	{63}	100.0	1.66
23	{56}	==>	{84}	100.0	1.66
24	{57}	==>	{62}	100.0	1.66
25	{57}	==>	{84}	100.0	1.66
26	{59}	==>	{4}	80.0	1.66
27	{59}	==>	{38}	80.0	1.66
28	{59}	==>	{61}	80.0	1.66
29	{59}	==>	{62}	80.0	1.66
30	{59}	==>	{63}	80.0	1.66
31	{59}	==>	{83}	80.0	1.66
32	{59}	==>	{84}	80.0	1.66
33	{60}	==>	{4}	80.0	1.66

Page 1 of 2

25/02/07 16:03

รูปที่ 4.8 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 2 ตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 80.0 %

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
34	{60}	==>	{61}	80.0	1.66
35	{60}	==>	{62}	80.0	1.66
36	{60}	==>	{63}	80.0	1.66
37	{60}	==>	{84}	80.0	1.66
38	{63}	==>	{4}	80.0	1.66
39	{63}	==>	{84}	80.0	1.66



25/02/07 16:03

Page 2 of 2

รูปที่ 4.8 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 2 ตัว (ต่อ)

จากการทดลองพบว่ารายงานรูปแบบลำดับที่ประกอบด้วยสมาชิก 2 ตัว มีทั้งสิ้น 39 ลำดับ ซึ่งรูปแบบลำดับเหล่านี้ มีความเป็นจริงในการเกิดรูปแบบลำดับตั้งแต่ 80 - 100 เปอร์เซ็นต์ และมีค่าสนับสนุน 1.66 เปอร์เซ็นต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 3 ตัว

3-SEQUENCE

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 80.0 %

Rules id	Sequence1	followed	Sequence2	Confidence	Support
1	{38}	==>	{63,4}	80.0	1.66
2	{38,63}	==>	{4}	100.0	1.66
3	{38}	==>	{63,84}	80.0	1.66
4	{38,63}	==>	{84}	100.0	1.66
5	{49}	==>	{54,4}	80.0	1.66
6	{49,54}	==>	{4}	100.0	1.66
7	{50}	==>	{57,62}	100.0	1.66
8	{50,57}	==>	{62}	100.0	1.66
9	{50}	==>	{57,84}	100.0	1.66
10	{50,57}	==>	{84}	100.0	1.66
11	{50}	==>	{62,84}	100.0	1.66
12	{50,62}	==>	{84}	100.0	1.66
13	{55,57}	==>	{62}	100.0	1.66
14	{55,57}	==>	{84}	100.0	1.66
15	{55,62}	==>	{84}	100.0	1.66
16	{55,63}	==>	{4}	100.0	1.66
17	{55,63}	==>	{84}	100.0	1.66
18	{55,83}	==>	{84}	100.0	1.66
19	{56}	==>	{38,4}	100.0	1.66
20	{56,38}	==>	{4}	100.0	1.66
21	{56}	==>	{38,63}	100.0	1.66
22	{56,38}	==>	{63}	100.0	1.66
23	{56}	==>	{38,84}	100.0	1.66
24	{56,38}	==>	{84}	100.0	1.66
25	{56}	==>	{61,4}	100.0	1.66
26	{56,61}	==>	{4}	100.0	1.66
27	{56}	==>	{61,84}	100.0	1.66
28	{56,61}	==>	{84}	100.0	1.66
29	{56}	==>	{62,84}	100.0	1.66
30	{56,62}	==>	{84}	100.0	1.66
31	{56}	==>	{63,4}	100.0	1.66
32	{56,63}	==>	{4}	100.0	1.66
33	{56}	==>	{63,84}	100.0	1.66

25/02/07 16:03

Page 1 of 2

รูปที่ 4.9 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 3 ตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 80.0 %

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
34	{56,63}	==>	{84}	100.0	1.66
35	{57}	==>	{62,84}	100.0	1.66
36	{57,62}	==>	{84}	100.0	1.66
37	{59}	==>	{38,4}	80.0	1.66
38	{59,38}	==>	{4}	100.0	1.66
39	{59}	==>	{38,63}	80.0	1.66
40	{59,38}	==>	{63}	100.0	1.66
41	{59}	==>	{38,84}	80.0	1.66
42	{59,38}	==>	{84}	100.0	1.66
43	{59}	==>	{61,4}	80.0	1.66
44	{59,61}	==>	{4}	100.0	1.66
45	{59}	==>	{61,84}	80.0	1.66
46	{59,61}	==>	{84}	100.0	1.66
47	{59}	==>	{62,84}	80.0	1.66
48	{59,62}	==>	{84}	100.0	1.66
49	{59}	==>	{63,4}	80.0	1.66
50	{59,63}	==>	{4}	100.0	1.66
51	{59}	==>	{63,84}	80.0	1.66
52	{59,63}	==>	{84}	100.0	1.66
53	{59}	==>	{83,84}	80.0	1.66
54	{59,83}	==>	{84}	100.0	1.66
55	{60}	==>	{61,4}	80.0	1.66
56	{60,61}	==>	{4}	100.0	1.66
57	{60}	==>	{61,84}	80.0	1.66
58	{60,61}	==>	{84}	100.0	1.66
59	{60}	==>	{62,84}	80.0	1.66
60	{60,62}	==>	{84}	100.0	1.66
61	{60}	==>	{63,4}	80.0	1.66
62	{60,63}	==>	{4}	100.0	1.66
63	{60}	==>	{63,84}	80.0	1.66
64	{60,63}	==>	{84}	100.0	1.66

25/02/07 16:03

Page 2 of 2

รูปที่ 4.9 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 3 ตัว (ต่อ)

จากการทดลองพบว่ารายงานรูปแบบลำดับที่ประกอบด้วยสมาชิก 3 ตัว มีทั้งสิ้น 64 ลำดับ ซึ่งรูปแบบลำดับเหล่านี้ มีความเป็นจริงในการเกิดรูปแบบลำดับตั้งแต่ 80 - 100 เปอร์เซ็นต์ และมีค่าสนับสนุน 1.66 เปอร์เซ็นต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 4 ตัว

4-SEQUENCE

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 80.0 %

Rules id	Sequence1	followed	Sequence2	Confidence	Support
1	{50}	==>	{57,62,84}	100.0	1.66
2	{50,57}	==>	{62,84}	100.0	1.66
3	{50,57,62}	==>	{84}	100.0	1.66



25/02/07 16:04

Page 1 of 1

รูปที่ 4.10 รายงานรูปแบบลำดับที่ประกอบด้วยสมาชิกจำนวน 4 ตัว

จากการทดลองพบว่ารายงานรูปแบบลำดับที่ประกอบด้วยสมาชิก 4 ตัว มีทั้งสิ้น 3 ลำดับ ซึ่งรูปแบบลำดับเหล่านี้ มีความเป็นจริงในการเกิดรูปแบบลำดับ 100 เปอร์เซ็นต์ และมีค่าสนับสนุน 1.66 เปอร์เซ็นต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 รายงานทางสถิติ

สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง

Hour of Day		
14/May/2006		
Hour	PageViews	Visits
00:00 - 00:59	107	7
01:00 - 01:59	16	9
02:00 - 02:59	6	5
03:00 - 03:59	5	4
04:00 - 04:59	22	14
05:00 - 05:59	24	19
06:00 - 06:59	58	17
07:00 - 07:59	34	10
08:00 - 08:59	18	16
09:00 - 09:59	15	11
10:00 - 10:59	18	16
11:00 - 11:59	46	17
12:00 - 12:59	18	14
13:00 - 13:59	20	13
14:00 - 14:59	16	17
15:00 - 15:59	18	12
16:00 - 16:59	41	13
17:00 - 17:59	34	16
18:00 - 18:59	17	16
19:00 - 19:59	24	17
20:00 - 20:59	193	13
21:00 - 21:59	20	10
22:00 - 22:59	70	7
23:00 - 23:59	54	7

รูปที่ 4.11 รายงานทางสถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง

จากการทดลองพบว่าเพจวิวสูงสุดอยู่ที่ช่วง 20:00 – 20:59 น. โดยมีเพจวิวเท่ากับ 193 ครั้ง ส่วนการเยี่ยมชมสูงสุดอยู่ที่ช่วงเวลา 05:00 – 05:59 น. โดยมีการเยี่ยมชมเท่ากับ 19 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถิติแยกตามไอพีแอดเดรสของผู้ใช้

IP Stats

14/May/2006

U ID	Host	PageViews	Visits
1	66.154.102.65	37	15
2	68.142.250.193	2	2
3	81.236.187.22	3	1
4	72.30.110.26	1	1
5	68.142.251.40	1	1
6	68.142.251.72	1	1
7	64.124.85.79	4	1
8	68.142.250.50	2	2
9	66.249.66.171	278	16
10	207.46.98.45	7	5
11	72.30.111.202	1	1
12	68.142.249.176	2	2
13	72.30.102.71	1	1
14	72.30.132.222	1	1
15	207.46.98.44	10	8
16	72.30.134.37	1	1
17	72.30.132.214	1	1
18	68.142.251.108	1	1
19	68.142.250.101	2	2
20	68.142.249.118	1	1
21	72.30.216.15	1	1
22	68.142.249.117	1	1
23	68.142.250.152	1	1
24	72.30.132.151	1	1
25	72.30.104.202	1	1
26	72.30.101.135	1	1
27	68.142.249.196	1	1
28	72.30.111.71	1	1
29	72.30.102.84	1	1
30	72.30.56.204	1	1
31	72.30.98.219	1	1
32	217.229.1.80	1	1
33	203.144.160.242	40	1
34	72.30.129.106	3	2
35	68.142.250.87	1	1

25/02/07 16:05

Page 1 of 7

รูปที่ 4.12 รายงานทางสถิติแยกตามไอพีแอดเดรสของผู้ใช้

จากการทดลองพบว่าไอพีแอดเดรสที่มีเพจวิวสูงสุดคือ ไอพีหมายเลข 66.249.66.171 โดยมีจำนวนเพจวิวเท่ากับ 278 ครั้ง ส่วนการเยี่ยมชมที่มากที่สุดมาจากไอพีแอดเดรสเดียวกัน โดยมีจำนวนครั้งที่เยี่ยมชมเท่ากับ 16 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถิติแยกตามเอเจนต์

Agent Stats

14/May/2006

agent_id	agent	PageViews	Visits
1	Gigabot/2.0/gigablast.com/spider.html	39	2
2	Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)	193	181
3	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)	44	10
4	Mozilla/5.0 (compatible; BecomeBot/3.0; MSIE 6.0 compatible; +http://www.become.com/site_owners.html)	4	1
5	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)	278	1
6	msnbot/1.0 (+http://search.msn.com/msnbot.htm)	18	3
7	Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.7.10) Gecko/20050724 Firefox/1.0.6	1	1
8	Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt)	42	2
9	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; InfoPath.1)	1	1
10	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; yplus 5.1.02b)	1	1
11	Mozilla/4.0 (compatible; MSIE 6.0b; Windows 98)	1	1
12	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; sbcydsl 3.12; YPC 3.0.3; SV1)	1	1
13	Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/312.8 (KHTML, like Gecko) Safari/312.6	4	1
14	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)	16	7
15	Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)	2	1
16	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)	60	6
17	Mozilla/4.0 (compatible; MSIE 5.17; Mac_PowerPC)	7	1
18	Mozilla/5.0 (Windows; U; Windows NT 5.1; fi-FI; rv:1.7.5) Gecko/20041108 Firefox/1.0	2	1
19	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; iebar; MaxiFiles; .NET CLR 1.1.4322)	1	1
20	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511	1	1
21	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50215)	2	1
22	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; FunWebProducts; .NET CLR 1.1.4322)	2	1
23	Mozilla/5.0 (compatible; LinksManager.com_bot +http://linksmanager.com/linkchecker.html)	4	1
24	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Q312461; SV1; .NET CLR 1.0.3705; .NET CLR 1.1.4322)	2	2
25	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; (R1 1.3); .NET CLR 1.1.4322)	5	4
26	Java/1.5.0	154	1
27	Scooter/3.3	3	1
28	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; FunWebProducts)	1	1

25/02/07 16:05

Page 1 of 2

รูปที่ 4.13 รายงานทางสถิติแยกตามเอเจนต์

จากการทดลองพบว่าเอเจนต์ที่มีเพจวิวมากที่สุดคือ Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html) โดยมีเพจวิวเท่ากับ 278 ครั้ง ส่วนเอเจนต์ที่มีการเยี่ยมชมมากที่สุดคือ Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp) โดยมีการเยี่ยมชมเท่ากับ 181 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถิติแยกตามจำนวนการเข้าชมเว็บไซต์

Number of Visits

14/May/2006

Visits	Visitors
1 Visits	224
2 Visits	13
3 Visits	0
4 Visits	0
5 Visits	1
6 Visits	0
7 Visits	0
8 Visits	1
9 Visits	0
10 Visits	0
11 Visits	0
12 Visits	0
13 Visits	0
14 Visits	0
15+ Visits	2

25/02/07 16:05

Page 1 of 1

รูปที่ 4.14 รายงานทางสถิติแยกตามจำนวนการเข้าชมเว็บไซต์

จากการทดลองพบว่าจำนวนครั้งที่มีการเยี่ยมชมมากที่สุดคือการเยี่ยมชมเว็บไซต์ 1 ครั้ง โดยมีจำนวนผู้ชมเท่ากับ 224 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถิติแยกตามเว็บเพจที่เข้าชม

URL Stats

14/May/2006

url_id	url	PageViews	Visits
1	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Cabana_Gr and_View_Hotel_Spa.php	2	2
2	http://www.differentasia.com/Thailand/Phuket/Kata_Beach/Peach_Hill_Hotel_Reso rt.php	2	2
3	http://www.differentasia.com/Vietnam/Chau_Doc_Town/Victoria_Chau_Doc_Hotel. php	3	3
4	http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php	11	7
5	http://www.differentasia.com/Thailand/Chiang_Mai/Old_Town/Rachamankha_Chia ng_Mai.php	6	6
6	http://www.differentasia.com/home/index.php	1	1
7	http://www.differentasia.com/Cambodia/off_Monivong_Boulevard/Sangkat_Wat_P hnomRaffles_Hotel_Le_Royal.php	2	2
8	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Imperial_B oat_House.php	3	3
9	http://www.differentasia.com/Thailand/Trad/Koh_Chang_Surrounding_Islands/Aan a_Resort_Spa.php	3	3
10	http://www.differentasia.com/Thailand/Chonburi/Pattaya/Rabbit_Resort.php	1	1
11	http://www.differentasia.com/Thailand/Bangkok/Sathon/Metropolitan_Hotel_Bangk ok.php	4	4
12	http://www.differentasia.com/Thailand/Bangkok/Silom/Solitel_Silom.php	2	2
13	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Beachcom ber_Hotel.php	2	2
14	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Pinnacle_S amui_Resort_Spa.php	3	3
15	http://www.differentasia.com/Thailand/Krabi/Muang_Town/Tubkaak_Resort_Krabi. php	6	6
16	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Nara_Gard en_Beach_Resort.php	2	2
17	http://www.differentasia.com/Thailand/Chiang_Mai/Doi_Suthep/Belle_Villa_Resort Chiang_Mai.php	1	1
18	http://www.differentasia.com/Thailand/Bangkok/Sathon/Sukhothai_Bangkok.php	2	2
19	http://www.differentasia.com/Thailand/Bangkok/Huay_Kwang/Swissotel_Le_Conco rde_Bangkok.php	5	5
20	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Bandara_R esort_Spa.php	1	1
21	http://www.differentasia.com/Cambodia/Central_Park/Victoria_Angkor_Hotel.php	2	2
22	http://www.differentasia.com/Thailand/Trad/Koh_Chang_Surrpunding_Islands/Ama ri_Emerald_Cove_Resort.php	2	2
23	http://www.differentasia.com/Thailand/Rayong/Koh_Samed/Mooban_Talay_Resort .php	1	1
24	http://www.differentasia.com/Thailand/Surat_Thani_Samui/Koh_Samui/Nora_Bea ch_Resort_Spa.php	1	1

25/02/07 16:06

Page 1 of 18

รูปที่ 4.15 รายงานทางสถิติแยกตามเว็บเพจที่เข้าชม

จากการทดลองพบว่า URL ที่มีเพจวิวมากที่สุดคือ http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php โดยมีจำนวนเพจวิวเท่ากับ 11 ครั้ง และเป็น URL ที่มีการเยี่ยมชมมากที่สุดคือ 7 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 วิธีการนำรูปแบบลำดับไปใช้ปรับโครงสร้างลิงก์ของเว็บไซต์

ปัจจัยที่สำคัญที่เราจะพิจารณาคือรูปแบบลำดับนั้นประกอบไปด้วยเพจต่างๆ โดยที่เพจเหล่านั้นมีการเชื่อมต่อระหว่างกันโดยตรงหรือไม่ ถ้าเพจเหล่านั้นไม่มีการเชื่อมต่อระหว่างกันโดยตรงเราควรที่จะแทรกลิงก์ (Shortcut Link) ระหว่างเพจเหล่านั้น โดยลิงก์จากเพจต้นทาง (Source Page) ไปยังเพจปลายทาง (Target Page) เพื่อให้การเข้าถึงใช้ระยะเวลาสั้นยิ่งขึ้น ถ้าไม่แทรกลิงก์แล้ว จะทำให้ผู้ใช้ต้องไปยังเส้นทางอื่นๆ เพื่อเข้าถึงเพจปลายทาง อีกส่วนที่พิจารณาคือการเปลี่ยนสภาพที่ปรากฏของลิงก์ที่มีอยู่ ตัวอย่างเช่น ถ้าอินเด็กซ์เพจมีลิงก์ที่ประกอบไปด้วยเซตของเพจหลายๆเพจ ก็ควรเน้นลิงก์เหล่านี้ให้โดดเด่นขึ้น (Highlighted Link) เมื่อผู้เยี่ยมชมมาเยี่ยมชมครั้งแรกก็สามารถที่จะเข้าถึงในส่วนต่างๆของเว็บไซต์ได้อย่างง่าย Mikroyannidis (2004 : 144)

The screenshot shows the School of Informatics website. The navigation bar includes 'School', 'Undergraduate', 'Postgraduate Taught', 'Postgraduate Research', 'Research', and 'Intranet'. The 'Quick links' section is circled and contains three links: 'Additional resources for CT121', 'Additional resources for CT338', and 'Additional resources for CT438'. The 'Module Resources' section lists 'Level 1 modules', 'Level 2 modules', 'Level 3 modules', and 'Level 4 modules'. The 'Level 1 Modules' section contains a detailed description of the modules and a table of module codes and titles. The table is circled and includes the following data:

Module code	Dept.	Module code	Module title
1 CN-1003	CT103		Software Engineering
2 CN-1005	CT105		Databases 1
3 CN-1006	CT106		Computer Architecture
4 CN-1007	CT116		Artificial Intelligence
5 CN-1010	CT111		Personal and Professional Development
6 CN-1012	CT118		Java Programming I
7 CN-1014	CT120		Hypertext and the World Wide Web
8 CN-1015	CT121		Level 1 Programming A
9 CN-1016	CT122		Level 1 Programming B
10 MA-1006	CT108		Mathematical Techniques for Computing (1&1)
11 MA-1007	CT114		Discrete Mathematics
12 SM-1006	CT107		Fundamentals of Management
13 SM-1010	CT201		Accounting
14 SM-1017	CT202		Business Economics

รูปที่ 4.16 ตัวอย่างเว็บเพจที่มีการปรับปรุงโครงสร้าง

จากรูปที่ 4.18 เนื่องจากมีผู้ใช้เข้าถึง Additional Resources for CT121, Additional Resources for CT338 และ Additional Resources for CT438 บ่อยครั้ง เนื่องจากเป็นวิชาที่มีความนิยม แต่ไม่มีลิงก์เชื่อมต่อระหว่างลิงก์เหล่านี้กับเว็บเพจนี้ จึงมีการเพิ่มลิงก์เหล่านี้ไว้ที่มุมบนซ้ายได้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Quick Links สำหรับเว็บเพจที่มีความนิยมสูง และมีลิงก์ที่เว็บเพจนี้เรียบร้อยแล้ว ก็จะถูกนำเสนอให้กับผู้ใช้โดยทำลิงก์เหล่านี้ให้โดดเด่นขึ้น เช่น ลิงก์ที่นำไปสู่เว็บเพจ Personal and Professional Development

ในโครงการนี้จะเน้นที่การแทรกลิงก์ระหว่างเพจที่เข้าเยี่ยมชมด้วยกันบ่อยๆ แต่ไม่ได้มีการลิงก์ถึงกันโดยตรง และการเปลี่ยนแปลงสภาพที่ปรากฏของลิงก์ที่ได้รับความนิยมสูง

เมื่อทดสอบขั้นตอนวิธีการค้นหารูปแบบลำดับกับข้อมูลเว็บเซิร์ฟเวอร์ล็อกไฟล์จากเว็บไซค์ <http://www.differentasia.com> วันที่ 14 พฤษภาคม 2549 เสร็จเรียบร้อยแล้ว นำรูปแบบลำดับที่น่าสนใจนำมาแปลงให้เป็น URL ต่างๆ แสดงดังตารางที่ 4.1 ซึ่งแต่ละ URL จะแสดงเนื้อหาเกี่ยวกับเพจต่างๆ ในรูปแบบลำดับแสดงดังตารางที่ 4.2

ตารางที่ 4.1 ตัวอย่างรูปแบบลำดับที่ค้นพบ

รูปแบบลำดับ	ค่าความเชื่อมั่น (%)	ค่าสนับสนุน (%)
http://www.differentasia.com/destinations/Chiang_Mai.php --> http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php	80	1.66
http://www.differentasia.com/destinations/Koh_Chang_Surrounding_Islands.php --> http://www.differentasia.com/destinations/Pranburi.php	100	1.66
http://www.differentasia.com/destinations/Koh_Samui.php --> http://www.differentasia.com/destinations/Pranburi.php , http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php	80	1.66
http://www.differentasia.com/index.php , http://www.differentasia.com/destinations/Cha_Am.php --> http://www.differentasia.com/destinations/Krabi.php , http://www.differentasia.com/faq/faq.php	100.0	1.66

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 เนื้อหาเกี่ยวกับเพจต่างๆ ในตัวอย่างรูปแบบลำดับ

รูปแบบลำดับ	ค่าความเชื่อมั่น (%)	ค่าสนับสนุน (%)
เพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดเชียงใหม่ --> เพจแนะนำโรงแรมศรีลันตา ที่เกาะลันตาใหญ่ จังหวัดกระบี่	80	1.66
เพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะช้าง จังหวัดตราด --> เพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี	100	1.66
เพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะสมุย จังหวัดสุราษฎร์ธานี --> เพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี, เพจแนะนำโรงแรมศรีลันตา ที่เกาะลันตาใหญ่ จังหวัดกระบี่	80	1.66
เพจหลักของเว็บ ซึ่งจะแสดงลิงก์ไปยังสถานที่ท่องเที่ยวและโรงแรมต่างๆ, เพจแนะนำโรงแรมที่พักต่างๆ ที่ชะอำ จังหวัดเพชรบุรี --> เพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดกระบี่, เพจคำถามที่พบบ่อยของเว็บ	100.0	1.66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปแบบลำดับที่แสดงในตารางที่ 4.1 สามารถอธิบายได้ดังนี้

- รูปแบบลำดับที่ 1 สามารถอธิบายได้ว่า ผู้ชม 80 เปอร์เซ็นต์ของทรานแซกชันทั้งหมดที่เยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดเชียงใหม่ จะเยี่ยมชมเพจแนะนำโรงแรมศรีลันตาที่เกาะลันตาใหญ่ จังหวัดกระบี่ โดยที่จำนวนทรานแซกชันที่เยี่ยมชมทั้ง 2 เพจคิดเป็น 1.66 เปอร์เซ็นต์ของทรานแซกชันทั้งหมด
- รูปแบบลำดับที่ 2 สามารถอธิบายได้ว่า ผู้ชม 100 เปอร์เซ็นต์ของทรานแซกชันทั้งหมดที่เยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะช้าง จังหวัดตราด จะเยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี โดยที่จำนวนทรานแซกชันที่เยี่ยมชมทั้ง 2 เพจคิดเป็น 1.66 เปอร์เซ็นต์ของทรานแซกชันทั้งหมด
- รูปแบบลำดับที่ 3 สามารถอธิบายได้ว่า ผู้ชม 80 เปอร์เซ็นต์ของทรานแซกชันทั้งหมดที่เยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะสมุย จังหวัดสุราษฎร์ธานี จะเยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี และเพจแนะนำโรงแรมศรีลันตา ที่เกาะลันตาใหญ่ จังหวัดกระบี่ โดยที่จำนวนทรานแซกชันที่เยี่ยมชมทั้ง 3 เพจคิดเป็น 1.66 เปอร์เซ็นต์ของทรานแซกชันทั้งหมด
- รูปแบบลำดับที่ 4 สามารถอธิบายได้ว่า ผู้ชม 100 เปอร์เซ็นต์ของทรานแซกชันทั้งหมดที่เยี่ยมชมเพจหลักของเว็บ และเพจแนะนำโรงแรมที่พักต่างๆ ที่ชะอำ จังหวัดเพชรบุรี จะเยี่ยมชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดกระบี่ และเพจคำถามที่พบบ่อยของเว็บ โดยที่จำนวนทรานแซกชันที่เยี่ยมชมทั้ง 4 เพจคิดเป็น 1.66 เปอร์เซ็นต์ของทรานแซกชันทั้งหมด

จากรูปแบบลำดับที่ได้ เราจะพิจารณารูปแบบลำดับนั้น ว่าประกอบไปด้วยเพจต่างๆ โดยที่เพจเหล่านั้นมีการเชื่อมต่อระหว่างกันโดยตรงหรือไม่ ถ้าเพจเหล่านั้นไม่มีการเชื่อมต่อระหว่างกันโดยตรงเราควรที่จะแทรกลิงก์ (Shortcut Link) ระหว่างเพจเหล่านั้น โดยลิงก์จากเพจต้นทาง (Source Page) ไปยังเพจปลายทาง (Target Page) เพื่อให้การเข้าถึงใช้ระยะเวลาสั้นยิ่งขึ้น อีกส่วนที่พิจารณาก็คือการเปลี่ยนสภาพที่ปรากฏของลิงก์ที่มีอยู่ เพื่อให้ ลิงก์เหล่านี้ให้โดดเด่นขึ้น (Highlighted Link) เมื่อผู้เยี่ยมชมมาเยี่ยมชมครั้งแรกก็สามารถที่จะเข้าถึงในส่วนต่างๆ ของเว็บไซต์ได้อย่างง่าย

- เมื่อพิจารณารูปแบบลำดับที่ 1 พบว่า เพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดเชียงใหม่ เป็นตัวชักนำให้ผู้ชมมาชมเพจแนะนำโรงแรมศรีลันตา ที่เกาะลันตาใหญ่ จังหวัดกระบี่ เมื่อพิจารณาที่โครงสร้างของเว็บไซต์พบว่าไม่มีลิงก์ของเพจแนะนำโรงแรมศรีลันตา ในหน้าของเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดเชียงใหม่ ดังนั้นจึงควรเพิ่มลิงก์ที่ไปยังเพจแนะนำโรงแรมศรีลันตา ลงในเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดเชียงใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เมื่อพิจารณารูปแบบลำดับที่ 2 พบว่า เพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะช้าง จังหวัดตราดเป็นตัวชักนำให้ผู้ชมมาชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี เมื่อพิจารณาที่โครงสร้างของเว็บไซต์พบว่ามีลิงก์ของเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี ในหน้าของเพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะช้าง จังหวัดตราดเรียบร้อยแล้ว แสดงดังรูปที่ 4.22 แต่สภาพที่ปรากฏของลิงก์ที่มีอยู่ยังไม่โดดเด่น จึงควรเน้นลิงก์นี้ให้โดดเด่นและวางในตำแหน่งที่เหมาะสมเพื่อให้มองเห็นได้ง่ายขึ้น

- เมื่อพิจารณารูปแบบลำดับที่ 3 พบว่า เพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะสมุย จังหวัดสุราษฎร์ธานี เป็นตัวชักนำให้ผู้ชมมาชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี และเพจแนะนำโรงแรมศรีลันตา ที่เกาะลันตาใหญ่ จังหวัดกระบี่ เมื่อพิจารณาที่โครงสร้างของเว็บไซต์พบว่ามีลิงก์ของเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี ในหน้าของเพจแนะนำโรงแรมที่พักต่างๆ ที่เกาะสมุย จังหวัดสุราษฎร์ธานี แสดงดังรูปที่ 4.23 แต่สภาพลิงก์ของหน้าปราณบุรีที่อยู่ในหน้าสมุยยังไม่โดดเด่น จึงควรเน้นลิงก์นี้ให้โดดเด่นและวางในตำแหน่งที่เหมาะสมเพื่อให้มองเห็นได้ง่ายขึ้น นอกจากนี้ยังพบว่าไม่มีลิงก์ของเพจแนะนำโรงแรมศรีลันตา ในหน้าของเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี ดังนั้นจึงควรเพิ่มลิงก์ที่ไปยังเพจแนะนำโรงแรมศรีลันตาลงในเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดปราณบุรี

- เมื่อพิจารณารูปแบบลำดับที่ 4 พบว่า เพจหลักของเว็บไซต์ และเพจแนะนำโรงแรมที่พักต่างๆ ที่ชะอำ จังหวัดเพชรบุรี เป็นตัวชักนำให้ผู้ชมมาชมเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดกระบี่ และเพจคำถามที่พบบ่อยของเว็บไซต์ เมื่อพิจารณาที่โครงสร้างของเว็บไซต์พบว่ามีลิงก์ของเพจแนะนำโรงแรมที่พักต่างๆ ในจังหวัดกระบี่ และเพจคำถามที่พบบ่อยของเว็บไซต์ ในหน้าของเพจหลักของเว็บไซต์ และเพจแนะนำโรงแรมที่พักต่างๆ ที่ชะอำ จังหวัดเพชรบุรีแล้ว แสดงดังรูปที่ 4.27, 4.28 และ 4.30 จัดได้ว่ามีการวางโครงสร้างเว็บไซต์ที่ดีแล้ว แต่สภาพลิงก์ของหน้ากระบี่ที่ปรากฏอยู่ในหน้าชะอำยังไม่โดดเด่น จึงควรเน้นลิงก์นี้ให้โดดเด่นและวางในตำแหน่งที่เหมาะสม

The screenshot displays the 'differentasia.com' website interface. At the top, there is a search bar with 'Thailand / North / Chiang Mai' selected. Below this, a 'Destination : Chiang Mai' section provides a 'Brief overview' of the city, describing it as a modern, cosmopolitan city with a rich cultural heritage. A 'Chiang Mai hotel Spotlight' section features two featured hotels: 'Central Duangtawan Hotel' and 'D2hotel Chiang Mai'. The right sidebar contains a 'Register for promotions' section, a list of 'All Chiang Mai Hotels' with price filters (<\$50, \$50-\$99, \$100+), and several promotional banners for 'AlphaVista Babel Fish' and 'International Herald Tribune'.

รูปที่ 4.17 เว็บไซต์ http://www.differentasia.com/destinations/Chiang_Mai.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Luxury Hotels				
	➤ Chedi Chiang Mai	from: \$267	Chiang Mai Riverside	Book Now
	➤ D2hotel Chiang Mai	from: \$153	Old Town	Book Now
	➤ Four Seasons Resort Chiangmai	from: \$419	Mae Rim	Book Now
	➤ Mandarin Oriental Dhara Dhevi	from: \$387	Chiang Mai Riverside	Book Now
	➤ Rachamanka Chiang Mai	from: \$146	Old Town	Book Now
	➤ Sukantara Cascade Resort	from: \$129	Mae Rim	Book Now
	➤ Amari Rincome	from: \$78	Huay Kaew	Book Now
	➤ Belle Villa Resort Chiang Mai	from: \$57	Doi Suthep	Book Now
	➤ Krisdadoi Chiangmai Aprime Resort Hotel	from: \$78	Hang Dong	Book Now
	➤ Maekok River Village Hotel	from: \$68	Chiang Mai Riverside	Book Now
	➤ Royal Princess Chiangmai	from: \$64	Muang	Book Now
	➤ Sheraton Chiang Mai	from: \$70	Chiang Mai Riverside	Book Now
	➤ Tamarind Village	from: \$66	Old Town	Book Now
	➤ Amora Tapae Hotel ,Chiang Mai	from: \$42	Muang	Book Now
	➤ Baantai Hotel	from: \$22	Muang	Book Now
	➤ Botanic Resort Chiangmai	from: \$47	Mae Rim	Book Now
	➤ BP Chiangmai City Hotel	from: \$22	Muang	Book Now
	➤ Central Duangtawan Hotel	from: \$40	Muang	Book Now
	➤ Chiang Dao Hill Resort Hotel	from: \$29	Chiang Dao	Book Now
	➤ Chiang Mai Gate Hotel	from: \$26	Muang	Book Now
	➤ Chiang Mai Orchid Hotel	from: \$49	Muang	Book Now
	➤ Chiangmai Hill 2000 Hotel	from: \$31	Huay Kaew	Book Now
	➤ Chiangmai Plaza Hotel	from: \$42	Chiang Mai Riverside	Book Now
	➤ Downtown Inn Hotel	from: \$26	Chiang Mai Riverside	Book Now
	➤ Empress Chiang Mai Hotel	from: \$36	Chiang Mai Riverside	Book Now
	➤ Eurasia Hotel	from: \$49	Muang	Book Now
	➤ Hmong Hilltribe Lodge Hotel	from: \$49	Mae Rim	Book Now
	➤ Holiday Garden Hotel & Resort	from: \$22	Muang	Book Now
	➤ Imperial Chiangmai Resort Hotel	from: \$45	Mae Rim	Book Now
	➤ Imperial Mae Ping Hotel	from: \$31	Chiang Mai Riverside	Book Now
	➤ Lanna Palace Hotel	from: \$31	Chiang Mai Riverside	Book Now
	➤ Lotus Pang Suankaew Hotel	from: \$29	Huay Kaew	Book Now
	➤ Northern Heritage Resort Hotel	from: \$42	Chiang Mai Riverside	Book Now
	➤ Novotel Chiang Mai Hotel	from: \$35	Old Town	Book Now
	➤ Park Hotel	from: \$28	Chiang Mai Riverside	Book Now
	➤ Raming lodge	from: \$33	Muang	Book Now
	➤ Red Hibiscus Guesthouse Hotel	from: \$21	Old Town	Book Now
	➤ Royal Peninsula Hotel and Services Apartment	from: On request	Muang	Book Now
	➤ Star Hotel	from: \$24	Chiang Mai Riverside	Book Now
	➤ Suan Bua Hotel	from: \$40	Doi Suthep	Book Now
	➤ Surivongse Hotel	from: \$29	Chiang Mai Riverside	Book Now
	➤ Thaton Chalet Hotel	from: \$35	Muang	Book Now
	➤ Tough Star Resort	from: \$29	Hang Dong	Book Now

- We Feature:**
- A wide selection of destinations
 - Extensive information & descriptions
 - Attractive aesthetic with clear photos
 - Reputable hotels of all budgetary levels
 - Highly competitive rates with no extra charges
 - Real time stock & availability
 - Instant confirmation for selected hotels
 - Flexible cancellation & amendment policy
 - Many satisfied customers
 - We are a registered member of the TAT



License No. 11/3849

home | about us | terms & conditions | faq | travel resources | site map | careers

รูปที่ 4.17 เว็บเพจ http://www.differentasia.com/destinations/Chiang_Mai.php (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

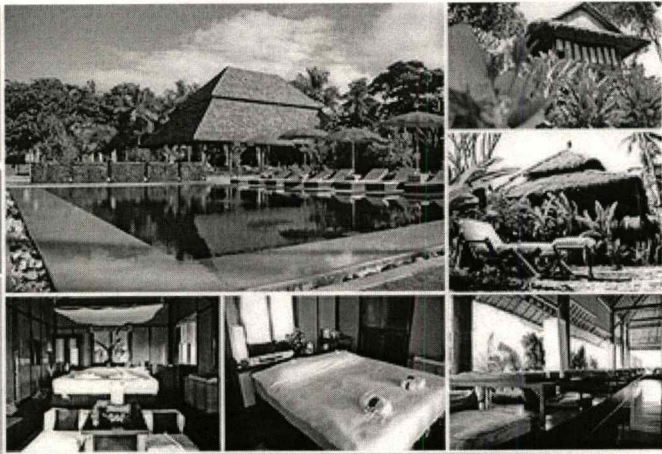


Hotel Quick Search for instant confirmation at over 800 Indochina Hotels

Country and location search fields with dropdown menus and a 'Go' button.

- Destinations: Home, Koh Lanta Review, Overview, Geography & Orientation, When to go?, What to do?, Culture, Diving, Food, National Parks, Photo Gallery, Booking, Thailand Information, Laos Information, Vietnam Information, Cambodia Information, Myanmar Information.

- All Bangkok Hotels, All Cha Am Hotels, All Chiang Mai Hotels, All Hua Hin Hotels, All Krabi Hotels, All Koh Chang Hotels, All Koh Samui Hotels, All Pattaya Hotels, All Phuket Hotels, All Pranburi Hotels.



differentasia.com unique destinations, different experiences

Thailand / South / Koh Lanta Sri Lanta

111 Moo 6, Klongnin Beach, Koh Lanta Yai, Lanta Island Krabi Style : Balinese Boutique Hotel Setting : Koh Lanta

- Overview, Accommodation, Restaurants, Activities, Highlights, Facilities, Map/Photo Gallery

Sri Lanta Overview

Covering a wide area of landscaped hillside and beachfront, Sri Lanta Resort blends effortlessly with its surroundings. The designers of the resort, acting on the motto "the resplendence of life" have created a retreat that sits perfectly in tune with its unspoiled natural environment.

The use of natural materials, water walls and ponds, and the careful placement of the buildings all encourage a sense of peace and serenity. Sri Lanta's 48 elevated thatched villas, terraced on the hills with winding paths, offer ocean and / or mountain view. Minimalist in style with thoughtfully designed natural furnishings and equipped with essential amenities, the Sri Lanta villas present guests with a rustic yet luxurious form of traditional living.

Sri Lanta's approach is one of subtle simplicity, offering those things are essential. There is the back tiled pool by the beach, named Anodana after another of the Hindu Gods. On the beach itself, guests find shaded sun beds, with full service from the Chedi Beach bar and Surya Chandra Restaurant. Nearby is the Sri Lanta Spa where professional therapists offer an array of revitalizing treatments developed from traditional healing methods, using local, natural ingredients.

Sri Lanta's island location, the surroundings, the resort's design and philosophy and the friendly "feel free" atmosphere management and staff create for each and every guest, make Sri Lanta the ideal place for nature-lovers and those seeking a complete escape from the distractions and pressures of life.

Click here for our full list of hotel members

Check Rates & Availability form with fields for Check in, Check out, Room(s), Adult(s), Child(s), and Bed Type #.



The Indochina Collection

Register for promotions, mailing lists and more privileges.

- Bookmark this page, Print Whole Report, E-mail a friend, Feedback

Alternative Koh Lanta Hotels

- Price filters: <\$50, \$50-\$99, \$100+; Hotel list: Layana Resort and Spa, Ravi Warin Resort & Spa, Sri Lanta, Lanta Casuarina Beach Resort, Pimalai Resort & Spa, Twin Lotus Resort & Spa

AltaVista Babel Fish translation tool with flags for various languages.



our Vision differentasia.com

Why differentasia.com?

- We Feature: A wide selection of destinations, Extensive information & descriptions, Attractive aesthetic with clear photos, Reputable hotels of all budgetary levels, Highly competitive rates with no extra charges, Real time stock & availability, Instant confirmation for selected hotels, Flexible cancellation & amendment policy, Many satisfied customers, We are a registered member of the TAT



รูปที่ 4.18 เว็บไซต์ http://www.differentasia.com/Thailand/Krabi/Koh_Lanta/SriLanta.php เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Hotel Quick Search
for instant confirmation
for over 800 Indochina
Hotels

--- country ---
--- location ---
Go

Destinations

Home

About Koh Chang & Surrounding Islands

- Overview
- Geography & Orientation
- When to go ?
- What to do ?
- Culture
- Diving
- Trekking

Places to Stay

Specific Travel

- Beach Hotels
- Boutique Hotels
- City Hotels
- Cultural Hotels
- Diving Hotels
- Eco Hotels
- Health Resorts
- Romantic Hotels
- Family Hotels
- Designer Hotels
- Luxury Hotels

Thailand Information

Laos Information

Vietnam Information

Cambodia Information

Myanmar Information

- All Bangkok Hotels
- All Cha Am Hotels
- All Chiang Mai Hotels
- All Hua Hin Hotels
- All Krabi Hotel
- All Koh Chang
- All Koh Samui
- All Pattaya Ho
- All Phuket Hotels
- All Pranburi Hotels



differentasia.com
unique destinations. different perspective.

.. / Thailand / East / Koh Chang

➤ **Destination : Koh Chang & Surrounding Islands**

Brief overview

Koh Chang is huge, the second largest island in the Kingdom but relatively speaking very undeveloped. Overall, the island is made up of lush green forested hills (70%) and mountains with the Khao Jom Prasat Mountain peaking at 2,428 feet. At the coastline, forests back down to beautiful sandy coral rimmed beaches. ...more

Koh Chang & Surrounding Islands hotel spotlight

Aana Resort Spa	Panviman Resort Koh Chang	Amari Emerald Cove Resort
		
Klong Prao Beach Koh Chang East Thailand	Klong Prao Beach Koh Chang East Thailand	Klong Prao Beach Koh Chang East Thailand

price < \$50 \$51 > price > \$99 price > \$100

➤ Aiyapura Resort & Spa	from: \$102	White Sand Beach	Book Now
➤ Amari Emerald Cove Resort	from: \$135	Klong Prao Beach	Book Now
➤ Aana Resort Spa	from: \$77	Klong Prao Beach	Book Now
➤ Barali Beach Resort	from: \$75	Klong Prao Beach	Book Now
➤ Bhumiyama Beach Resort	from: \$57	Kai Bae Beach	Book Now
➤ Koh Chang Cliff Beach Resort Hotel	from: \$69	Kai Bae Beach	Book Now
➤ Koh Chang Resort & Spa Hotel	from: \$56	Klong Prao Beach	Book Now
➤ Koh Chang Tropicana Resort & Spa	from: \$88	Klong Prao Beach	Book Now
➤ Panviman Resort Koh Chang	from: \$95	Klong Prao Beach	Book Now
➤ Shantaa Koh Kood Hotel	from: \$73	Koh Kood	Book Now
➤ KC Grande Resort	from: \$23	Klong Prao Beach	Book Now
➤ Klong Prao Resort Hotel	from: \$29	White Sand Beach	Book Now
➤ Koh Chang Kacha Resort & Spa Hotel	from: \$38	White Sand Beach	Book Now
➤ Koh Mak Resort Hotel	from: \$22	Koh Mak	Book Now
➤ Ramayana Koh Chang Resort	from: \$45	Klong Prao Beach	Book Now

Register for promotions, mailing lists and more privileges.

All Koh Chang & Surrounding Islands Hotels

<\$50 \$50-\$99 \$100+

- Aiyapura Resort & Spa
- Amari Emerald Cove Resort
- Aana Resort Spa
- Barali Beach Resort
- Bhumiyama Beach Resort
- Koh Chang Cliff Beach Resort Hotel

Google Translate for full list of hotel members

ALL HOTELS IN KOH CHANG & SURROUNDING ISLANDS

INTERNATIONAL Herald Tribune

our Vision differentasia.com

Why differentasia.com?

We Feature:


- A wide selection of destinations
- Extensive information & descriptions
- Attractive aesthetic with clear photos
- Reputable hotels of all budgetary levels
- Highly competitive rates with no extra charges
- Real time stock & availability
- Instant confirmation for selected hotels
- Flexible cancellation & amendment policy
- Many satisfied customers
- We are a registered member of the TAT



License No. 11/3849

รูปที่ 4.19 เว็บเพจ http://www.differentasia.com/destinations/Koh_Chang_Surrounding_Islands.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Hotel Quick Search
for instant confirmation at over 800 Indochina Hotels

Go

Destinations

--- country ---
--- location ---

Go

D Home
D About Koh Samui

- Overview
- Geography & Orientation
- When to go ?
- What to do ?
- Culture
- Nightlife
- Getting there & around


D Places to Stay

D Specific Travel

- Beach Hotels
- Boutique Hotels
- City Hotels
- Cultural Hotels
- Diving Hotels
- Eco Hotels
- Health Resorts
- Romantic Hotels
- Family Hotels
- Designer Hotels
- Luxury Hotels

D Thailand Information
D Laos Information
D Vietnam Information
D Cambodia Information
D Myanmar Information

- All Bangkok Hotels
- All Cha Am Hotels
- All Chiang Mai Hotels
- All Hua Hin Hotels
- All Krabi Hotels
- All Koh Chang
- All Koh Samui
- All Pattaya Hot
- All Phuket Hotels
- All Pranburi Hotels



differentasia.com
unique destinations different perspectives


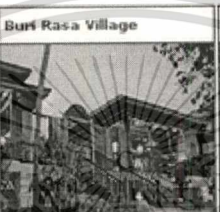

.. / Thailand / South / Koh Samui

➤ **Destination : Koh Samui**

Brief overview

Koh Samui experiences the best all year round weather of all the islands, boasts an array of sensational high end and hip resorts, beautiful beaches, gourmet restaurants, luxury spas and a plethora of activities. Once the budget travelers first stop, Koh Samui, like much of modern Thailand has become an exotic retreat for people to whom budget is not the main concern. Access to the island is easy due to its local airport.more

Koh Samui hotel Spotlight

 Anantara Resort Koh Samui	 Buri Rasa Village	 Santiburi Resort Koh Samui
Bophut Koh Samui South Thailand	Chaweng Koh Samui South Thailand	Maenam Koh Samui South Thailand

[Click here for our full list of hotel members](#)

ALL HOTELS IN KOH SAMUI

price < \$50		\$51 > price > \$99	price > \$100	
➤ Amari Palm Reef Resort	from: \$151	Chaweng	Book Now	
➤ Anantara Resort Koh Samui	from: \$178	Bophut	Book Now	
➤ Baan Haad Ngam Boutique Resort	from: \$114	Chaweng	Book Now	
➤ Bandara Resort Spa	from: \$139	Bophut	Book Now	
➤ Buri Rasa Village	from: \$136	Chaweng	Book Now	
➤ Central Samui Beach Resort	from: \$169	Chaweng	Book Now	
➤ Chaba Cabana Samui Resort	from: \$126	Bophut	Book Now	
➤ Chaweng Regent Beach Resort	from: \$168	Chaweng	Book Now	
➤ Imperial Boat House	from: \$109	Chaweng Mon	Book Now	
➤ Muang Samui Spa Resort	from: \$146	Chaweng	Book Now	
➤ Napasai	from: \$217	Maenam	Book Now	
➤ Nora Beach Resort & Spa	from: \$109	Chaweng	Book Now	
➤ Paradise Beach Resort	from: \$115	Maenam	Book Now	
➤ Poppies Samui	from: \$221	Chaweng	Book Now	
➤ Renaissance Koh Samui Resort & Spa	from: \$203	Lamai	Book Now	
➤ Samui Palm Beach Resort	from: \$100	Maret	Book Now	
➤ Samui Peninsula Spa Resort	from: \$199	Bophut	Book Now	
➤ Sila Evason Hideaway	from: \$101	Bophut	Book Now	
➤ The Tongyai Bay	from: \$186	Choeng Mon	Book Now	
	from: \$392	Bophut	Book Now	
	from: \$230	Tongsai Beach	Book Now	

differentasia.com
unique destinations different perspectives

Register for promotions, mailing lists and more privileges.

All Koh Samui Hotels

- <\$50
- \$50-\$99
- \$100+

- Amari Palm Reef Resort
- Anantara Resort Koh Samui
- Baan Haad Ngam Boutique Resort
- Bandara Resort Spa
- Buri Rasa Village
- Central Samui Beach Resort

differentasia.com
unique destinations different perspectives


Register for promotions, mailing lists and more privileges.

All Koh Samui Hotels

- <\$50
- \$50-\$99
- \$100+

- Amari Palm Reef Resort
- Anantara Resort Koh Samui
- Baan Haad Ngam Boutique Resort
- Bandara Resort Spa
- Buri Rasa Village
- Central Samui Beach Resort

AltaVista
Babel Fish
To translate this page, click a flag!



INTERNATIONAL Herald Tribune

Our Vision
differentasia.com

Why?
differentasia.com?

We Feature:

- A wide selection of destinations
- Extensive information & descriptions
- Attractive aesthetic with clear photos
- Reputable hotels of all budgetary levels
- Highly competitive rates with no extra charges
- Real time stock & availability
- Instant confirmation for selected hotels
- Flexible cancellation & amendment policy
- Many satisfied customers
- We are a registered member of the IAT

มีลิงก์ไปที่หน้าปรามบุรี

รูปที่ 4.20 เว็บเพจ http://www.differentasia.com/destinations/Koh_Samui.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



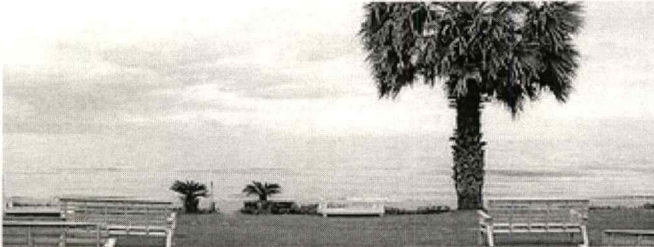
License No. 11/3849

➤ Aloha Resort Samui	from: \$59	Maret	Book Now
➤ Amarin Victoria Resort	from: \$52	Maenam	Book Now
➤ Arayaburi Boutique Resort	from: \$52	Bophut	Book Now
➤ Baan Samui Resort	from: \$94	Chaveng	Book Now
➤ Banana Fan Sea Samui	from: \$62	Bophut	Book Now
➤ Banburee Resort & Spa	from: \$77	Laemset	Book Now
➤ Beachcomber Hotel	from: \$82	Chaveng	Book Now
➤ Blue Lagoon	from: \$94	Chaveng	Book Now
➤ Cabana Grand View Hotel & Spa	from: \$57	Bophut	Book Now
➤ Central Samui Village	from: \$75	Chaveng	Book Now
➤ Chaba Samui Resort	from: \$92	Bophut	Book Now
➤ Chaveng Buri Resort	from: \$85	Chaveng	Book Now
➤ Coral Bay Resort & Spa	from: \$53	Chaveng	Book Now
➤ First Bungalow Beach Resort	from: \$57	Chaveng	Book Now
➤ Implana Resort Samui	from: \$78	Chaveng	Book Now
➤ Iyara Beach Hotel Plaza	from: \$83	Chaveng	Book Now
➤ Pavilion Samui Boutique Resort	from: \$95	Santi Beach	Book Now
➤ Peace Resort	from: \$66	Bophut	Book Now
➤ Punn Pree Da Hip Resort	from: \$88	Bophut	Book Now
➤ Santiburi Resort Koh Samui	from: \$68	Maenam	Book Now
➤ Seafan Beach Resort	from: \$66	Maenam	Book Now
➤ Thai Ayodhya Villas & Spa	from: \$85	Maret	Book Now
➤ Thai House Beach Resort	from: \$71	Maret	Book Now
➤ The Fair House Beach Resort & Hotel	from: \$54	Bophut	Book Now
➤ Zazen Boutique Resort & Spa	from: \$90	Maenam	Book Now
➤ Al's Hut Resort	from: \$33	Bophut	Book Now
➤ Baan Hin Sai Resort	from: On request	Lamai	Book Now
➤ Best Western Samui Bayview Resort	from: \$45	Bophut	Book Now
➤ Bo Phut Resort & Spa	from: \$43	Bophut	Book Now
➤ Chaveng Resort Hotel	from: \$49	Chaveng	Book Now
➤ Coral Cove Chalet Hotel	from: \$41	Lamai	Book Now
➤ Muang Kulaypan Hotel	from: On request	Chaveng	Book Now
➤ Nova Samui Resort	from: \$47	Chaveng	Book Now
➤ Pinnacle Samui Resort & Spa	from: \$40	Maenam	Book Now
➤ Rajapruek Samui Resort	from: On request	Lipanoi	Book Now
➤ Samui Amanda Resort	from: On request	Nathon Village	Book Now
➤ Samui First House	from: \$31	Bophut	Book Now
➤ Samui Laguna Resort Hotel	from: \$38	Lamai	Book Now
➤ Samui Orchid Resort & Aquarium	from: \$31	Maret	Book Now
➤ Samui Paradise Chaveng Beach Resort	from: \$43	Maenam	Book Now
➤ Samui Yacht Club	from: \$35	Tongsai Beach	Book Now
➤ Sand Sea Resort & Spa	from: \$26	Maret	Book Now
➤ Seascape Beach Resort	from: \$46	Chaveng	Book Now
➤ The Imperial Samui Hotel	from: On request	Bophut	Book Now
➤ The Island Resort & Spa	from: \$35	Bophut	Book Now
➤ White House Hotel	from: \$20	Choeng Mon	Book Now

[home](#) | [about us](#) | [terms & conditions](#) | [faq](#) | [travel resources](#) | [site map](#) | [careers](#)

รูปที่ 4.20 เว็บเพจ http://www.differentasia.com/destinations/Koh_Samui.php (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Hotel Quick Search
for instant confirmation
at over 800 Indochina
Hotels

Go

Destinations

--- country ---

--- location ---

Go

- D Home
- D About Pranburi
 - Overview
 - Geography & Orientation
 - When to go ?
 - What to do ?
 - National Parks
- D Places to Stay
- D Specific Travel
 - Beach Hotels
 - Boutique Hotels
 - City Hotels
 - Cultural Hotels
 - Diving Hotels
 - Eco Hotels
 - Health Resorts
 - Romantic Hotels
 - Family Hotels
 - Designer Hotels
 - Luxury Hotels

- D Thailand Information
- D Laos Information
- D Vietnam Information
- D Cambodia Information
- D Myanmar Information
 - All Bangkok Hotels
 - All Cha Am Hotels
 - All Chiang Mai Hotels
 - All Hua Hin Hotels
 - All Krabi Hotels
 - All Koh Chang Hotels
 - All Koh Samui Hotels
 - All Pattaya Hotels
 - All Phuket Hotels
 - All Pranburi Hotels

[differentasia.com](http://www.differentasia.com)
Local destinations. Different perspectives.

.. / Thailand / Center / Pranburi

Destination : Pranburi

Brief overview

Around 30 KM's south of Hua Hin, for the time being Pranburi, the Kingdom's new coastal hip hotspot encapsulates serenity and tranquility. Set well away from anything that could be described as mega-developed, the beaches are lovely and quiet and the water is blue. This is why pranburi is the delightful destination for some of our favorite boutique resorts: Prasaban, Evason Hideaway and Aleenta Resort & Spa. The town of Pranburi is small and charming with a scattering of houses, a local school and the Na Huam temple making up the majority of the local infrastructure. If you get restless, visit the Sam Roi national park, explore the Tham Praya Nakhon cave or dime the Khao Kaleke hill. And if that's not enough, slightly more upbeat Hua Hin isn't too far away to go for the day or at nighttime.

...more

Pranburi hotel Spotlight



Pranburi Center Thailand Pranburi Center Thailand

[Click here for our full list of hotel members](#)

ALL HOTELS IN PRANBURI			
	price < \$50	\$51 > price > \$99	price > \$100
>> Evason Hideaway Hua Hin	from: \$372	Pranburi	Book Now
>> Praseban Resort	from: \$105	Pranburi	Book Now
>> Tanaosri Resort	from: \$57	Pranburi	Book Now

Register for promotions, mailing lists and more privileges.

All Pranburi Hotels

<\$50 \$50-\$99 \$100+

- Evason Hideaway Hua Hin
- Praseban Resort
- Tanaosri Resort

Alta Vista
Babel Fish
To translate this page, click a flag!

INTERNATIONAL
Herald Tribune

Our Vision
differentasia.com

Why
differentasia.com?

- We Feature:**
- A wide selection of destinations
 - Extensive information & descriptions
 - Attractive aesthetic with clear photos
 - Reputable hotels of all budgetary levels
 - Highly competitive rates with no extra charges
 - Real time stock & availability
 - Instant confirmation for selected hotels
 - Flexible cancellation & amendment policy
 - Many satisfied customers
 - We are a registered member of the TAT

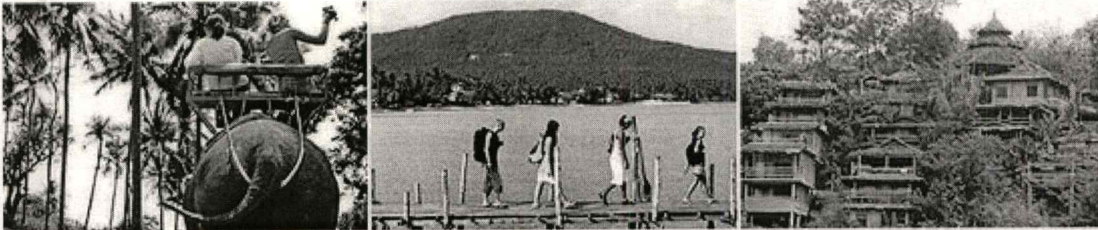


License No. 11/3849

[home](#) | [about us](#) | [terms & conditions](#) | [faq](#) | [travel resources](#) | [site map](#) | [careers](#)

รูปที่ 4.21 เว็บเพจ <http://www.differentasia.com/destinations/Pranburi.php>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



differentasia.com
unique destinations. different perspective.

Destinations

--- country ---
--- location ---
GO

Home
Specific Travel
Beach Hotels
Boutique Hotels
City Hotels
Cultural Hotels
Diving Hotels
Eco Hotels
Health Resorts
Romantic Hotels
Family Hotels
Designer Hotels
Luxury Hotels
Bangkok For A Day
Bangkok For A Night
Thailand Festivals & Events
South East Asia Festivals & Events

Differentasia.com features a unique handpicked collection of exclusive hotels, four and five star resorts, designer retreats, boutique escapes and unusual holiday homes all perfect for the ultimate Indochina & Thailand Vacation. Unlike regular portal sites which blanket the entire Asia region, we at differentasia.com take pride in selecting only the most special and unique hotels & resorts which we know will offer our guests the most memorable holiday experiences.

Hotel Quick Search

Select a Destination: -- select a destination --
Check in: yyyy-mm-dd
Check out: yyyy-mm-dd
Room(s): 1
Adult(s): 1
Child(s) not over 12: 0
Bed Type # 01: Double (sleeps 2)
Hotels Search

Simply input your dates, destination and number of rooms required and we'll promptly show you all of our available stock for guaranteed room reservation/ instant confirmation, followed by those available on request, in which case you'll receive a confirmation within 24 hours.

Hotel Members : Looking for the best hotels in Thailand & Indochina?
Click & go to here for our full list of hotel members

Destination Spotlight

Sun, sea and sand - Koh Samui - Voted in Forbs Magazine's 13 most "sexy islands in the world" survey.

Peace and solitude - Khao Lak - one of the Kingdom's last remaining undeveloped eco-hotspots.

Divers paradise - Koh Racha - an undeveloped wonderland of breathtaking natural beauty.

Desert island escape - Koh Phi Phi - two truly majestic islands just off the Andaman Coast.

Siem Reap & the UNESCO listed Angkor Wat - one of the greatest religious monuments in the world.

This Months Retreats

Baan Krating Jungle Beach - the delights of tropical jungle living combined with blissful seaside frolicking.

The Tubkaak - commanding one of the most beautiful stretches of sand in the Krabi province.

The Rachamankha - this magnificent documentation of Northern Thai architecture is located within the walls of Chiang Mai's traditional old town.

D2 Hotel Chiang Mai - chic new boutique hotel located just off the famous Chang Klan Road.

On one of the lesser developed islands, Koh Lanta, Pimalai Resort & Spa offers 5-star luxury whilst maintaining natural beauty, tradition and charm.

Register for promotions, mailing lists and more privileges.

Recommended Hotels

Le, Vinnan
Koh Samet

The Cliff
Krabi

SriLanta
Koh Lanta, Krabi

Click & go to Special Package

INTERNATIONAL
Herald Tribune

our Vision
differentasia.com

Why
differentasia.com?

We Feature:

- A wide selection of destinations
- Extensive information & descriptions
- Attractive aesthetic with clear photos
- Reputable hotels of all budgetary levels
- Highly competitive rates with no extra charges
- Real time stock & availability

รูปที่ 4.22 เว็บไซต์ <http://www.differentasia.com/index.php>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



➤ All Thailand Hotels :

- All Bangkok Hotels
- All Cha Am Hotels
- All Chiang Mai Hotels
- All Hua Hin Hotels
- All Krabi Hotels
- All Pranburi Hotels

มีลิงก์ไปที่หน้าชะอำ



➤ Boutique Hotels :

- Boutique Thailand Hotels
- Bangkok Boutique Hotels
- Koh Samui Boutique Hotels
- Phuket Boutique Hotels
- Chiang Mai Boutique Hotels

มีลิงก์ไปที่หน้ากระบี่



➤ Luxury Hotels :

- Luxury Thailand Hotels
- Bangkok Luxury Hotels
- Koh Samui Luxury Hotels
- Phuket Luxury Hotels
- Chiang Mai Luxury Hotels
- Hua Hin Luxury Hotels
- Pranburi Luxury Hotels
- Luang Prabang Luxury Hotels



➤ Thai Islands :

- Koh Chang Island
- Koh Lanta Island
- Koh Maiton Island
- Koh Phi Phi Island
- Koh Phuket Island
- Koh Samui Island
- Koh Yao Island
- Koh Samet Island
- Koh Racha Island

South East Asia Events

The well-worn backpackers' trail of South East Asia is still full of surprises, especially if you time your visit to coincide with one of the many Festivals or Events that can be found throughout the year in this vast continent. With its multitude of climates, cultures, languages and natural wonders you can be guaranteed an exciting event is taking place, somewhere at any given time. So, haul yourself off that seat and out of that square box you call an office and join in the fun.

Thailand / South East Asia Festivals and Events.

Bangkok by Day

The city of Bangkok is a fascinating place and a great city for those who enjoy hunting out bargains, visiting sacred temples, relaxing at a spa, or simply traveling around by unusual modes of transport and seeing the city. Read more ...

Bangkok by Night

Bangkok is a truly fascinating city and a great diverse night life scene. Although over the dramatic efforts to curb the whole Kingdom's reputation Thailand's 'seedy' image, Bangkok's night life's vast array of high end clubs, rustic bars and more ...

มีลิงก์ไปที่หน้าคำถามที่พบบ่อย

home | about us | terms & conditions | faq | travel resources | site map | careers

- Instant confirmation for selected hotels
- Flexible cancellation & amendment policy
- Many satisfied customers
- We are a registered member of the TAT



License No. 11/3849

รูปที่ 4.22 เว็บไซต์ <http://www.differentasia.com/index.php> (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Hotel Quick Search
for instant confirmation at over 800 Indochina Hotels

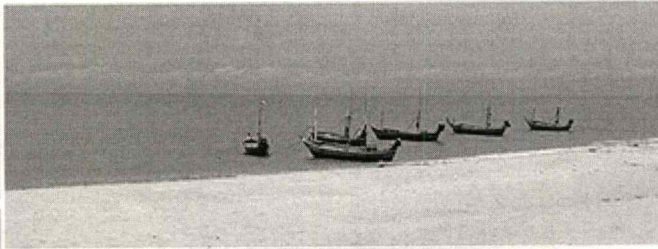
... / Thailand / Center / Cha-Am

Destinations

... country ...
... location ...

Home
About Cha-Am
Overview
Geography & Orientation
When to go ?
What to do ?
Food
National Parks
Places to Stay
Specific Travel
Beach Hotels
Boutique Hotels
City Hotels
Cultural Hotels
Diving Hotels
Eco Hotels
Health Resorts
Romantic Hotels
Family Hotels
Designer Hotels
Luxury Hotels
Thailand Information
Laos Information
Vietnam Information
Cambodia Information
Myanmar Information

All Bangkok
All Cha Am H
All Chiang Ma
All Hua Hin Hotels
All Krabi Hotels
All Koh Chang Hotels
All Koh Samui Hotels
All Pattaya Hotels
All Phuket Hotels
All Pranburi Hotels



differentasia.com
unique destinations, different perspectives

.. / Thailand / Center / Cha-Am

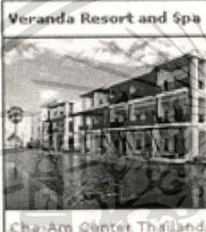
➤ **Destination : Cha-Am**

Brief overview

As with it's somewhat more well-known neighbor Hua Hin, Cha Am is also a popular holiday destination for holidaying Thai's due to its proximity to Bangkok. The beaches are nice, the water is clean and if you come here on a weekday, you'll probably have the entire beach to yourself. With a quiet almost non-existent nightlife, this seaside town is great for families with young children those seeking peaceful evenings and a more relaxing alternative to Hua Hin and Pattaya. And, if resorting and beaching it gets a bit monotonous, nearby are some very interesting sites to venture out to see. ...more


Cha-Am hotel Spotlight

Veranda Resort and Spa



Cha-Am Center Thailand

Springfield Beach Resort



Cha-Am Center Thailand

price < \$50 \$51 > price > \$99 price > \$100

Hotel Name	Price	Cha-Am	Book Now
➤ Springfield Village	from: \$119	Cha-Am	Book Now
➤ Springfield Beach Resort	from: \$67	Cha-Am	Book Now
➤ Veranda Resort and Spa	from: \$95	Cha-Am	Book Now
➤ Long Beach Cha Am Hotel	from: \$44	Cha-Am	Book Now

Register for promotions, mailing lists and more privileges.

All Cha-Am Hotels

<\$50 \$50-\$99 \$100+

- Springfield Village
- Springfield Beach Resort
- Veranda Resort and Spa
- Long Beach Cha Am Hotel

AltaVista Babel Fish
To translate this page, click a flag!

INTERNATIONAL
Herald Tribune

Our Vision
differentasia.com

Why differentasia.com?

We Feature:

- A wide selection of destinations
- Extensive information & descriptions
- Attractive aesthetic with clear photos
- Reputable hotels of all budgetary levels
- Highly competitive rates with no extra charges
- Real time stock & availability
- Instant confirmation for selected hotels
- Flexible cancellation & amendment policy
- Many satisfied customers
- We are a registered member of the TAT

License No. 11/3849

รูปที่ 4.23 เว็บเพจ http://www.differentasia.com/destinations/Cha_Am.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Hotel Quick Search
for instant confirmation
at over 800 Indochina
Hotels

Go

Destinations

--- country ---
--- location ---

Go

Home

About Krabi

- Overview
- Geography & Orientation
- When to go ?
- What to do ?
- Culture
- Diving
- Food
- Festivals
- National Parks

Places to Stay

Specific Travel

- Beach Hotels
- Boutique Hotels
- City Hotels
- Cultural Hotels
- Diving Hotels
- Eco Hotels
- Health Resorts
- Romantic Hotels
- Family Hotels
- Designer Hotels
- Luxury Hotels



differentasia.com
unique destinations different people like

.. / Thailand / South / Krabi

➤ **Destination : Krabi**

Brief overview

Krabi is without a doubt one of southern Thailand's most picturesque provinces, with a very diverse look and one of breathtaking natural beauty, especially represented in its five national parks. Inland, Krabi features dense rainforest rich in fauna and flora. However, Krabi is most famously noted for the beauty of its islands and beaches. There are over 150 islands with only a handful inhabited. The islands generally feature mangroves on the eastern coast with dense inland forests backing on to pristine white sand beaches on the western coast. ...more

Krabi hotel Spotlight

 SriLanta Koh Lanta South Thailand	 Central Krabi Bay Resort Ao Nang & Rai Lei Krabi South Thailand	 Alis Hotel Ao Nang & Rai Lei Krabi South Thailand
---	---	---

[Click here for our full list of hotel members](#)

ALL HOTELS IN KRABI

price < \$50	\$51 > price > \$99	price > \$100
➤ Central Krabi Bay Resort	from: \$197	Ao Nang
➤ Nakamanda Resort & Spa	from: \$210	Muang (Town)
➤ Pakasai Resort Hotel	from: \$100	Muang (Town)

The Indochina Collection

Register for promotions, mailing lists and more privileges.

All Krabi Hotels

<\$50 \$50-\$99 \$100+

- Central Krabi Bay Resort
- Nakamanda Resort & Spa
- Pakasai Resort Hotel
- Rayavadee
- Sheraton Krabi Beach Resort
- The Tribkaak Krabi

AltaVista **Babel Fish**
To translate this page, click a flag!

INTERNATIONAL **Herald Tribune**

our **Vision**
differentasia.com

Why
differentasia.com?

รูปที่ 4.24 เว็บไซต์ <http://www.differentasia.com/destinations/Krabi.php>


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thailand Information	Rayavadee	from: \$571	Ao Nang	Book Now
Laos Information	Sheraton Krabi Beach Resort	from: \$190	Muang (Town)	Book Now
Vietnam Information	The Tubkaak Krabi	from: \$163	Muang (Town)	Book Now
Cambodia Information	Alis Hotel	from: \$63	Ao Nang	Book Now
Myanmar Information	Ao Nang Villa Resort	from: \$52	Muang (Town)	Book Now
All Bangkok Hotels	Golden Beach Resort Hotel	from: \$63	Muang (Town)	Book Now
All Cha Am Hotels	Krabi National Park Success Hotel	from: \$85	Ao Nang	Book Now
All Chiang Mai Hotels	Maritime Park & Spa Resort Hotel	from: \$56	Muang (Town)	Book Now
All Hua Hin Hotels	Pavilion Queen's Bay	from: \$90	Ao Nang	Book Now
All Krabi Hotels	Peace Laguna Resort Hotel	from: \$68	Muang (Town)	Book Now
All Koh Chang Hotels	Phranang Inn Resort Hotel	from: \$52	Muang (Town)	Book Now
All Koh Samui Hotels	The Cliff Ao Nang Resort	from: \$88	Ao Nang	Book Now
All Pattaya Hotels	Vogue Pranang Bay Resort And Spa Hotel	from: \$56	Muang (Town)	Book Now
All Phuket Hotels	Ao Nang Buri Hotel	from: \$38	Ao Nang	Book Now
All Pranburi Hotels	Ao Nang Princeville Resort Hotel	from: \$35	Muang (Town)	Book Now
	Aonang Paradise Resort	from: \$32	Muang (Town)	Book Now
	Aonang Sunset Hotel	from: \$26	Ao Nang	Book Now
	Beach Terrace Hotel	from: \$42	Ao Nang	Book Now
	Krabi La Playa Resort	from: \$49	Ao Nang	Book Now
	Krabi Resort Hotel	from: \$45	Ao Nang	Book Now
	Sand Sea Resort Hotel	from: \$30	Ao Nang	Book Now
	Semkiet Buri Resort Hotel	from: \$47	Muang	
	Srisuksant Resort	from: \$29		
	Thai Village Resort Hotel	from: On request		
	Tipa Resort Hotel	from: \$38		

มีลิงก์ไปที่หน้าคำถามที่พบบ่อย

We Feature:

- A wide selection of destinations
- Extensive information & descriptions
- Attractive aesthetic with clear photos
- Reputable hotels of all budgetary levels
- Highly competitive rates with no extra charges
- Real time stock & availability
- Instant confirmation for selected hotels
- Flexible cancellation & amendment policy
- Many satisfied customers
- We are a registered member of the TAT



License No. 11/3849

รูปที่ 4.24 เว็บไซต์ <http://www.differentasia.com/destinations/Krabi.php> (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

บทสรุป

5.1 สรุปผลโครงการ

โครงการพัฒนาโปรแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติจัดทำขึ้นเพื่อค้นหารูปแบบการเยี่ยมชมเว็บไซต์ รวมทั้งสถิติการเยี่ยมชมเว็บไซต์ ซึ่งองค์ความรู้ที่ได้เหล่านี้ สามารถนำไปประยุกต์ใช้ในการปรับปรุงโครงสร้างเว็บไซต์ ทำให้ผู้ใช้สามารถเข้าถึงเว็บไซต์ได้อย่างทั่วถึงและรวดเร็ว

ผู้จัดทำดำเนินโครงการโดยใช้ขั้นตอนการทำเหมืองข้อมูลการใช้งานเว็บไซต์ ซึ่งมี 3 ขั้นตอน คือ การเตรียมข้อมูลการใช้งาน การค้นหารูปแบบ และการวิเคราะห์รูปแบบ

ในการเตรียมข้อมูล เริ่มจากการทำความสะอาดเว็บเซิร์ฟเวอร์ล็อกไฟล์ โดยการกรองข้อมูลที่ไม่เกี่ยวข้องออก แล้วนำข้อมูลที่ได้จากขั้นตอนนี้มาระบุตัวผู้ใช้โดยใช้โฮสต์และเอเจนต์เป็นตัวระบุ จากนั้นจึงนำข้อมูลที่ผ่านขั้นตอนนี้ มาระบุเซสชันผู้ใช้โดยใช้เวลานานอก ซึ่งโดยปกติแล้วจะใช้ค่า 30 นาที เมื่อได้เซสชันผู้ใช้แล้ว จะนำมาระบุทรานแซกชันโดยใช้อัลกอริทึมการอ้างอิงไปข้างหน้าไกลที่สุด จากนั้นนำข้อมูลที่ได้จากขั้นตอนการระบุทรานแซกชันมาสร้างเป็นลำดับผู้ใช้เพื่อใช้ในขั้นตอนการค้นหาแบบ

ในการค้นหาแบบจะใช้อัลกอริทึม AprioriAll ซึ่งจะมีขั้นตอนการทำงานคล้ายกับอัลกอริทึม Apriori แต่จะมีความแตกต่างตรงที่วิธีการสร้างลำดับคู่แข่ง โดยจะถือว่าลำดับก่อนหลังมีความแตกต่างกัน หลังจากที่ได้ลำดับขนาดใหญ่ทั้งหมดที่เป็นไปได้แล้ว จะนำลำดับขนาดใหญ่เหล่านี้ไปสร้างเป็นรูปแบบลำดับโดยใช้อัลกอริทึม genpattern

ในการวิเคราะห์รูปแบบจะนำรูปแบบลำดับที่ได้สร้างไว้ มาพิจารณาเพ่งต่างๆที่อยู่ในรูปแบบลำดับว่าควรที่จะแทรกถึงระหว่างเพ่งเหล่านั้นหรือไม่ ถ้าเพ่งเหล่านั้นไม่มีการเชื่อมต่อกันโดยตรง เราควรที่จะแทรกถึงระหว่างเพ่งเหล่านั้น อีกส่วนที่พิจารณา ก็คือการเปลี่ยนสภาพที่ปรากฏของลิงก์ที่มีอยู่ ถ้าเป็นเพ่งที่มีลิงก์ซึ่งเป็นลิงก์ที่ประกอบไปด้วยเซตของเพ่งหลายๆเพ่ง ก็ควรเน้นลิงก์เหล่านี้ให้โดดเด่นขึ้น เมื่อผู้เยี่ยมชมมาเยี่ยมชมครั้งแรกก็สามารถที่จะเข้าถึงในส่วนต่างๆของเว็บไซต์ได้อย่างง่าย

ในการวิเคราะห์ทางสถิติประกอบไปด้วย สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง สถิติแยกตามไอพีแอดเดรสของผู้ใช้ สถิติแยกตามเอเจนต์ สถิติแยกตามจำนวนการเข้าชมเว็บไซต์ และสถิติแยกตามเว็บเพจที่เข้าชม

ผลจากการพัฒนาระบบทำให้ได้โปรแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ โดยโปรแกรมนี้อาจให้องค์ความรู้ซึ่งช่วยให้ผู้พัฒนาเว็บไซต์นำมาใช้ประกอบการตัดสินใจปรับโครงสร้างลิงก์ของเว็บไซต์

5.2 ข้อเสนอแนะ

ผู้จัดทำมีข้อเสนอแนะในการพัฒนาขั้นตอนการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อกโดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ เพื่อให้มีประสิทธิภาพในการทำงานมากขึ้นดังนี้

- ปรับเปลี่ยนโครงสร้างที่ใช้เก็บข้อมูลเพื่อให้การทำงานของโปรแกรมทำงานเร็วขึ้น โดยการประยุกต์ใช้ทรี
- การระบุตัวผู้ใช้ด้วยไอพีแอดเดรสและเอเจนต์ไม่รับประกันความเป็นหนึ่ง จะล้มเหลวเมื่อมีการสับเปลี่ยนหมุนเวียนไอพี ดังนั้น จึงต้องหาวิธีระบุตัวผู้ใช้ที่แน่นอน โดยอาจจะใช้การลงทะเบียน
- เนื่องจากอัลกอริทึม AprioriAll ใช้เวลาในการประมวลผลค่อนข้างนาน ไม่เหมาะกับข้อมูลปริมาณมาก ดังนั้น จึงควรเปลี่ยนแปลงอัลกอริทึมที่ใช้ในการค้นหารูปแบบลำดับ

บรรณานุกรม

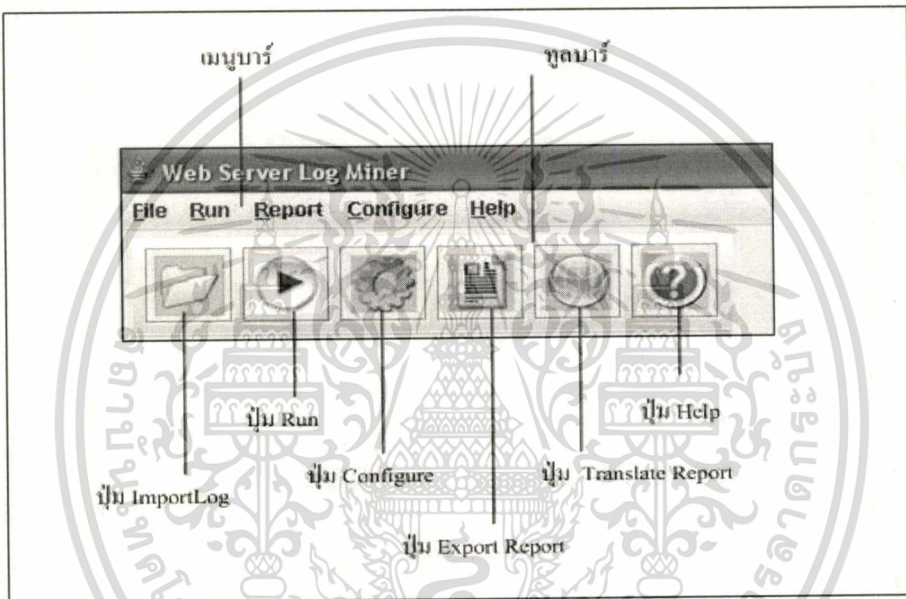
- Agrawal, R. et al. 1995. **Mining Sequential Patterns**. [Online]. Available: https://www.almaden.ibm.com/software/projects/iis/hdb/Publications/papers/icde95_rj.pdf.
- Berendt, B. et al. 2002. **Web Usage Mining for E-Business Applications**. [Online]. Available: <http://www.cs.helsinki.fi/events/ecmlpkdd/pdf/berendt-2.pdf>.
- Cooley, R. et al. 1999. **Data Preparation for Mining World Wide Web Browsing Patterns**. [Online]. Available: <http://maya.cs.depaul.edu/~classes/ect584/papers/cms-kais.pdf>.
- Cooley, R. et al. 1997. "Web Mining : Information and Pattern Discovery on the World Wide Web." **Tools with Artificial Intelligence, Proceedings of the Ninth IEEE International Conference**. Pp. 558-567.
- Lalani, A. 2003. "Data Mining of Web Access Logs." M.Sc.Thesis, Royal Melbourne Institute of Technology.
- Mikroyannidis, A. 2004. **Web Usage Driven Adaptation of the Semantic Web**. [Online]. Available : http://kmi.open.ac.uk/events/usersweb/papers/12_mikroyannidis_final.pdf.
- Ming-Syan, C. et al. 1998. "Efficient Data Mining for Path Traversal Patterns." **IEEE Transactions on Knowledge and Data Engineering**, 10(2) : 209-221.
- Mobasher, B. 2004. **Web Usage Mining and Personalization**. [Online]. Available: <http://maya.cs.depaul.edu/~mobasher/papers/IC-Handbook-04.pdf>. 2004.
- Mortazavi-Asl B. 2001. "Discovering and Mining User Web-Page Traversal Pattern." M.Sc.Thesis of Simon Fraser University.
- Srivastava, J. et al. 2000. "Web Usage Mining : Discovery and Applications of Usage Patterns from Web Data." **ACM SIGKDD**. 1(2) : 12-23.
- Zhu, T. 2001. **Web Usage Mining for Internet Recommendation**. [Online]. Available: <http://www.cs.ualberta.ca/~tszhu/paper/proposal.ps>. 2001.

ภาคผนวก ก.

คู่มือการใช้งานโปรแกรม

โปรแกรมการทำเหมืองข้อมูลเว็บเซิร์ฟเวอร์ล็อก โดยใช้เทคนิครูปแบบลำดับและการวิเคราะห์ทางสถิติ จะใช้เมนูบาร์ ซึ่งมีลักษณะเป็นพุดคาว์นเมนู หรือทูลบาร์ ที่ประกอบไปด้วยปุ่มต่างๆ เพื่อใช้เป็นคำสั่งหลักในการใช้งาน โปรแกรม

เมนูบาร์และทูลบาร์



รูปที่ ก.1 เมนูบาร์และทูลบาร์

รายละเอียดของเมนูบาร์แต่ละตัวมีดังนี้

- เมนู File เป็นเมนูที่ใช้สำหรับนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์เข้ามาประมวลผล
- เมนู Run เป็นเมนูที่ใช้สั่งให้โปรแกรมประมวลผล
- เมนู Report เป็นเมนูที่ใช้ออกรายงาน
- เมนู Configure เป็นเมนูที่ใช้กำหนดค่าและจัดการข้อมูลที่ระบบต้องใช้
- เมนู Help เป็นเมนูที่ใช้ให้ความช่วยเหลือ

รายละเอียดของทูลบาร์แต่ละตัวมีดังนี้

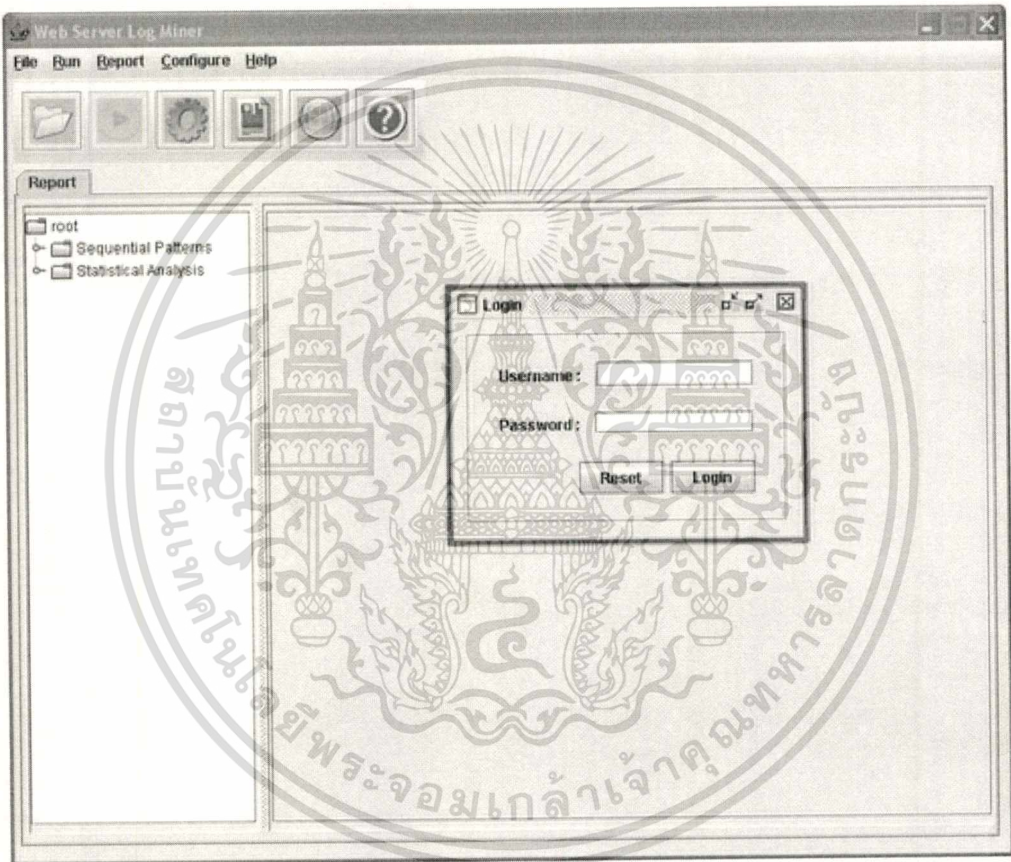
- ปุ่ม ImportLog เป็นปุ่มที่ใช้สำหรับนำเข้าเว็บเซิร์ฟเวอร์ล็อกไฟล์เข้ามาประมวลผล
- ปุ่ม Run เป็นปุ่มที่ใช้สั่งให้โปรแกรมประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปุ่ม Configure	เป็นปุ่มที่ใช้กำหนดค่าและจัดการข้อมูลที่ระบบต้องใช้
ปุ่ม Export Report	เป็นปุ่มที่ใช้ออกรายงาน โดยการ Export
ปุ่ม Translate Report	เป็นปุ่มที่ใช้ออกรายงาน โดยการ Translate
ปุ่ม Help	เป็นปุ่มที่ใช้ให้ความช่วยเหลือ

การเข้าสู่ระบบ

กรอกชื่อและรหัสผ่านเพื่อเข้าสู่โปรแกรมดังรูปที่ ก.2

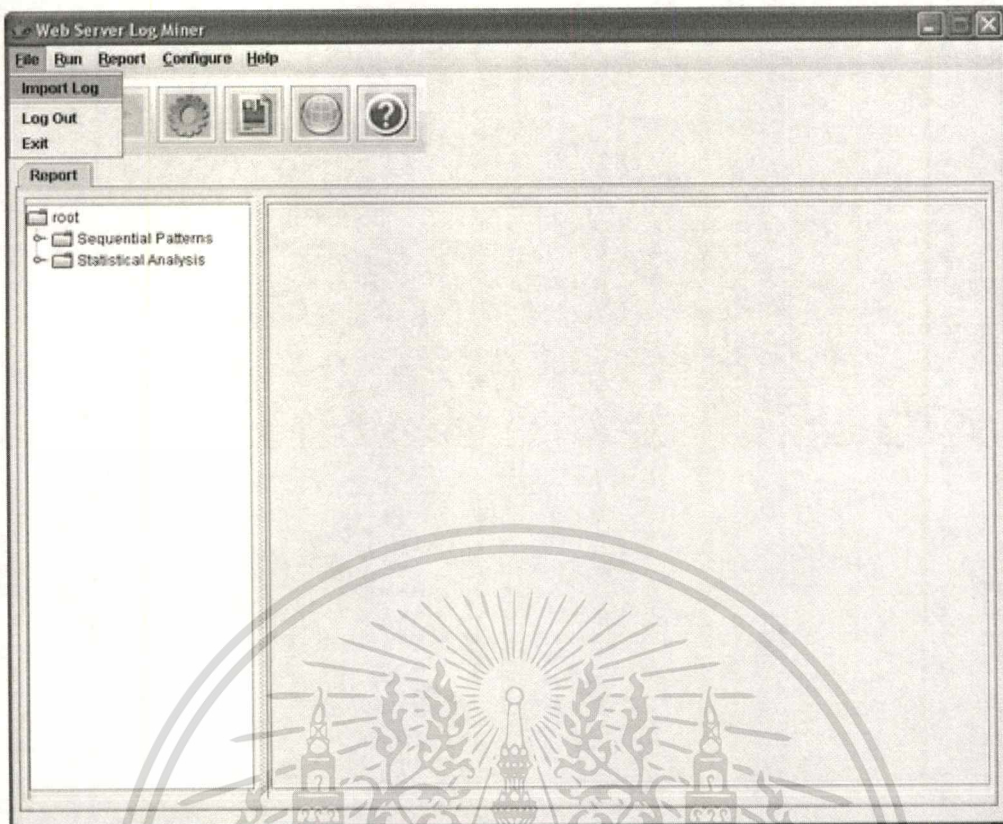


รูปที่ ก.2 หน้าจอการล็อกอิน

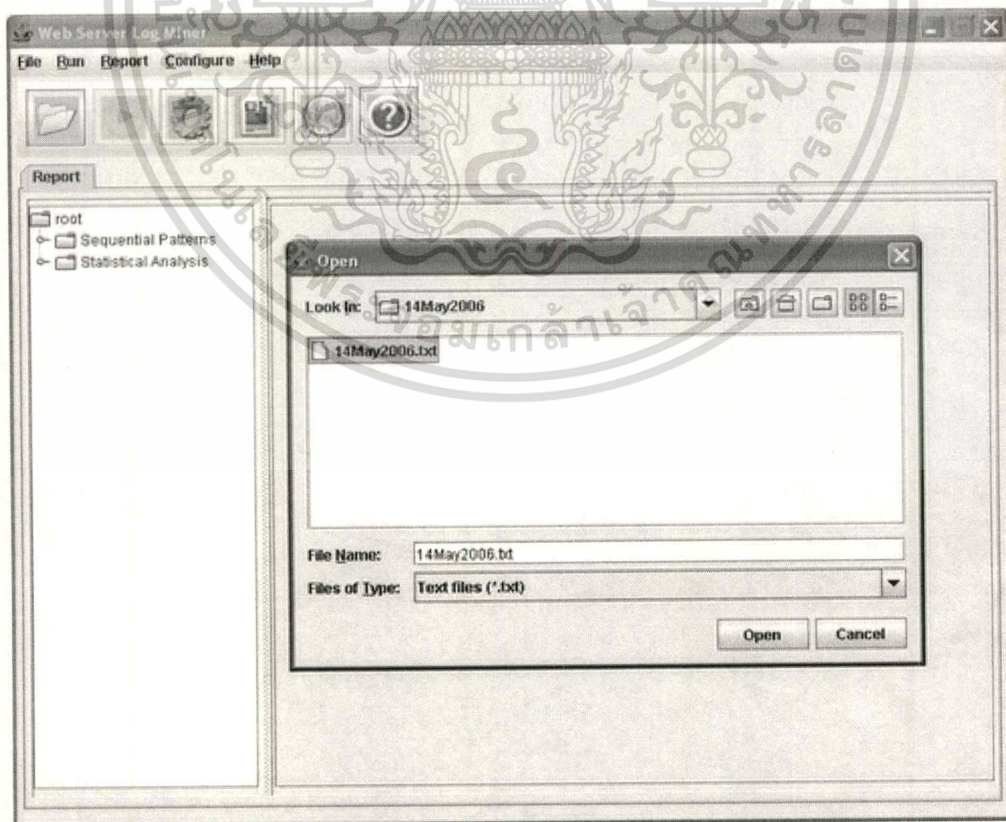
การนำเข้าเว็บเซิร์ฟเวอร์ล็อก

1. เลือกที่เมนู File -> ImportLog หรือปุ่ม ImportLog ดังรูปที่ ก.3
2. เลือกไฟล์ที่ต้องการนำเข้าจากนั้นคลิกที่ปุ่ม Open ดังรูปที่ ก.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.3 หน้าจอการนำเข้าเว็บเซิร์ฟเวอร์ล็อก

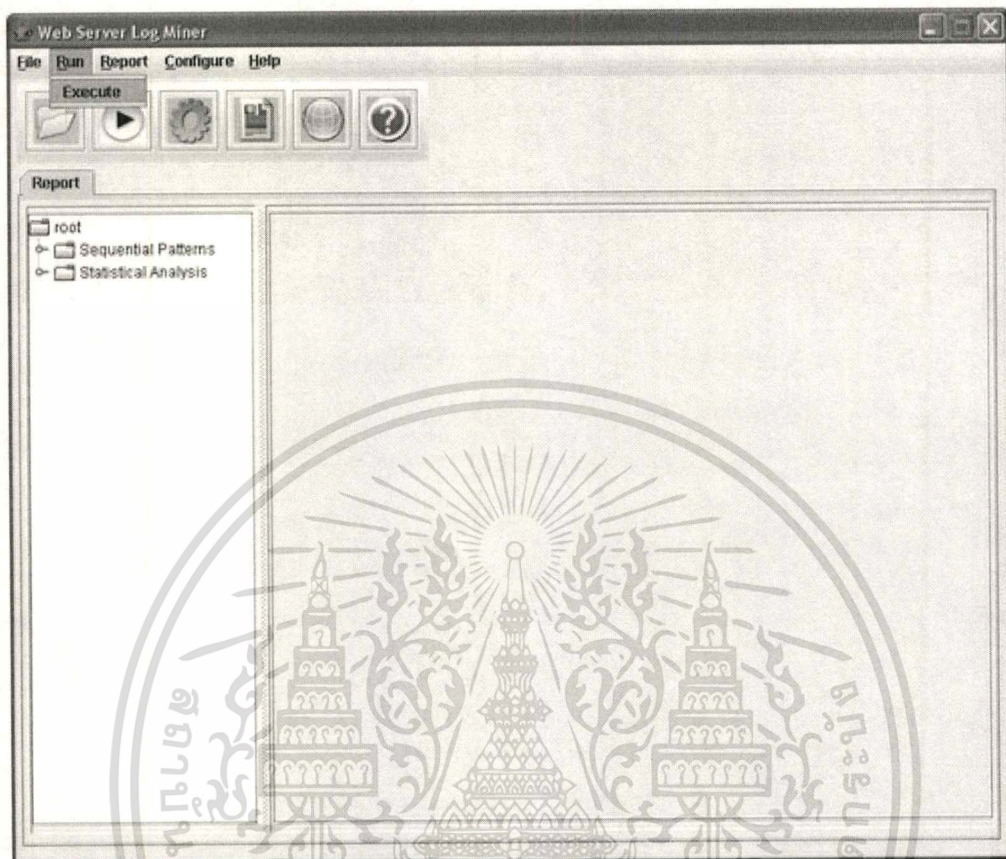


รูปที่ ก.4 หน้าจอการเลือกไฟล์ที่ต้องการนำเข้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสั่งให้โปรแกรมประมวลผล

เลือกที่เมนู Run -> Execute หรือปุ่ม Run ดังรูปที่ ก.5



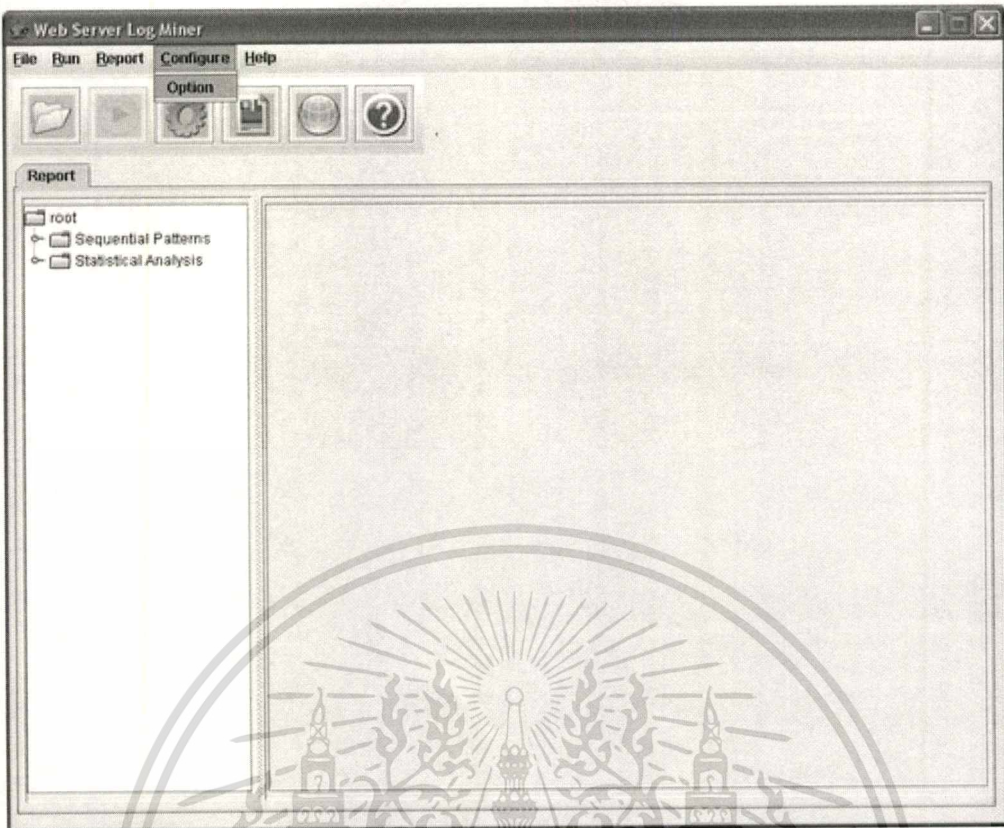
รูปที่ ก.5 หน้าจอการสั่งให้โปรแกรมประมวลผล

การกำหนดค่าและจัดการข้อมูลที่ใช้ในระบบ

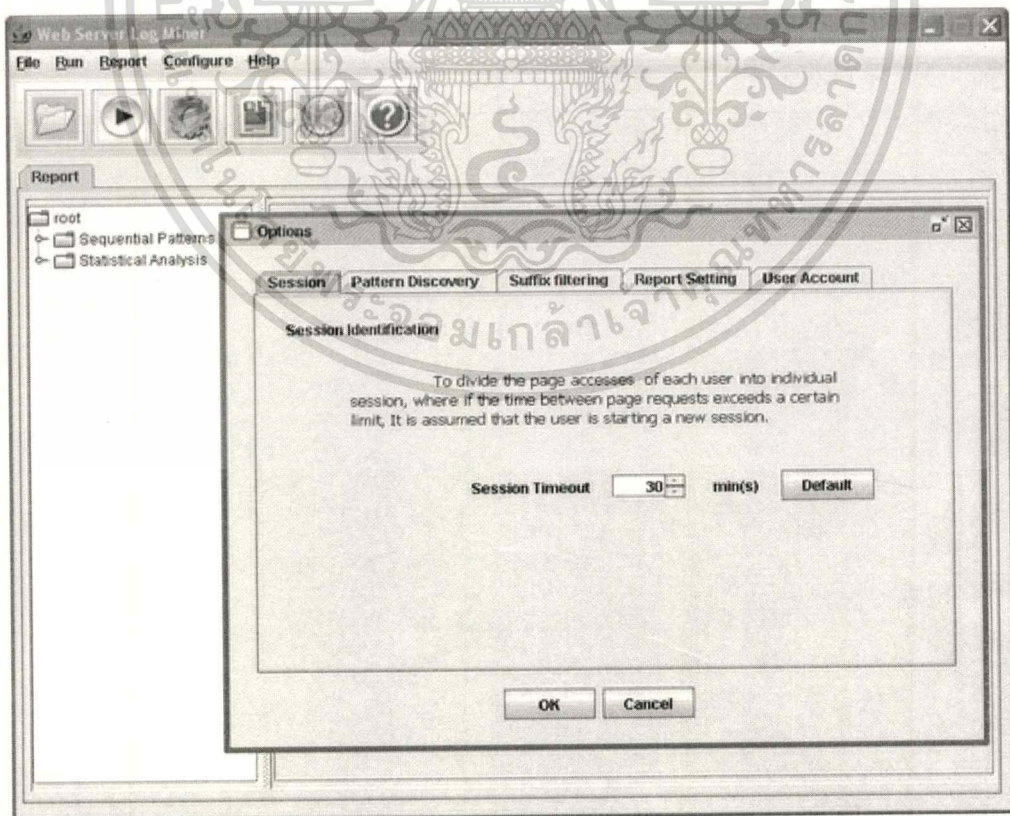
1. เลือกที่เมนู Configure -> Option หรือปุ่ม Configure ดังรูปที่ ก.6
2. เลือกที่แท็บ Session แล้วกำหนดค่าเซสชันที่ต้องการ ดังรูปที่ ก.7
3. เลือกที่แท็บ Pattern Discovery แล้วกำหนดค่า Minimum Support, Minimum Confidence และค่า Maximum k-Itemset ที่ต้องการ ดังรูปที่ ก.8
4. เลือกที่แท็บ Suffix Filtering แล้วเพิ่ม แก๊ไข หรือลบส่วนต่อท้ายที่ใช้กรอง ดังรูปที่ ก.9
5. เลือกที่แท็บ Report Setting แล้วกำหนดชื่อเว็บไซต์และรายงานที่ต้องการ ดังรูปที่ ก.10
6. เลือกที่แท็บ User Account แล้วเพิ่ม แก๊ไขหรือลบรายชื่อผู้ใช้งานระบบ ดังรูปที่ ก.11

หมายเหตุ ตาราง Accounts จะมีการกำหนดสิทธิ์การเข้าใช้งาน ในกรณีที่ป็นผู้ใช้ปกติเข้ามาใช้งาน โปรแกรม หน้าจอการกำหนดค่าและจัดการข้อมูลจะไม่แสดงในส่วนของ Suffix Filtering และ Report Setting แสดงดังรูปที่ ก.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

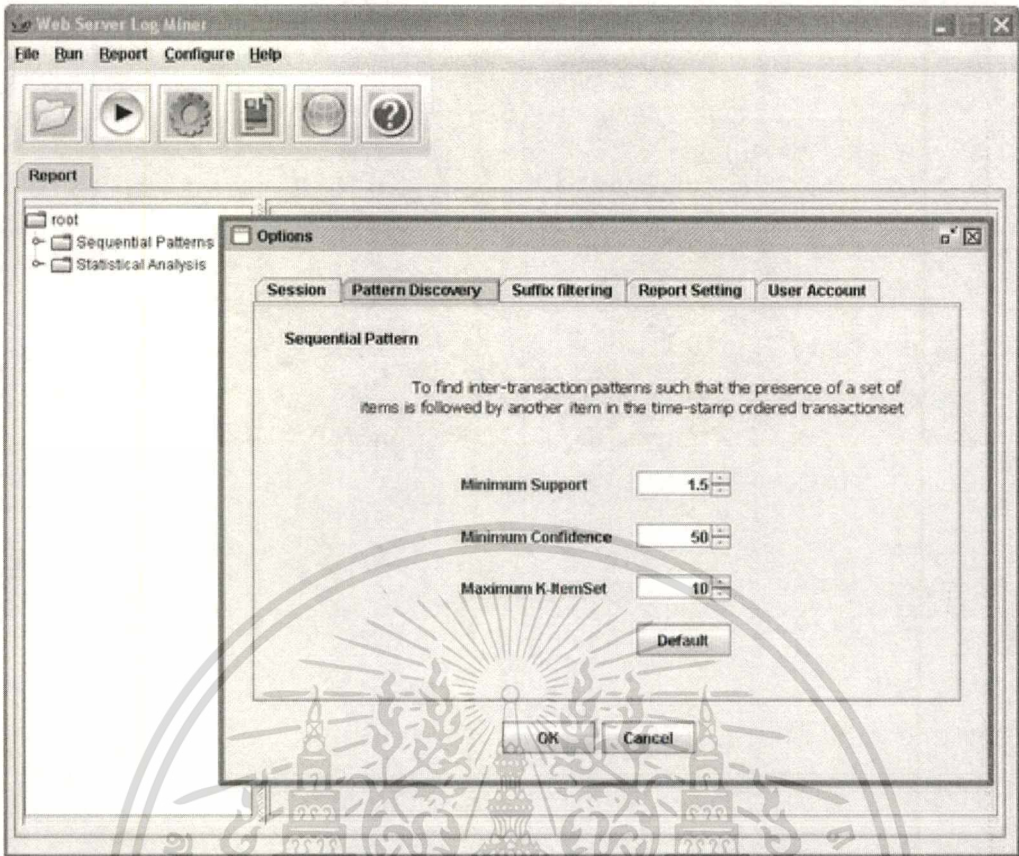


รูปที่ ก.6 หน้าจอการกำหนดค่าและจัดการข้อมูลที่ใช้ในระบบ

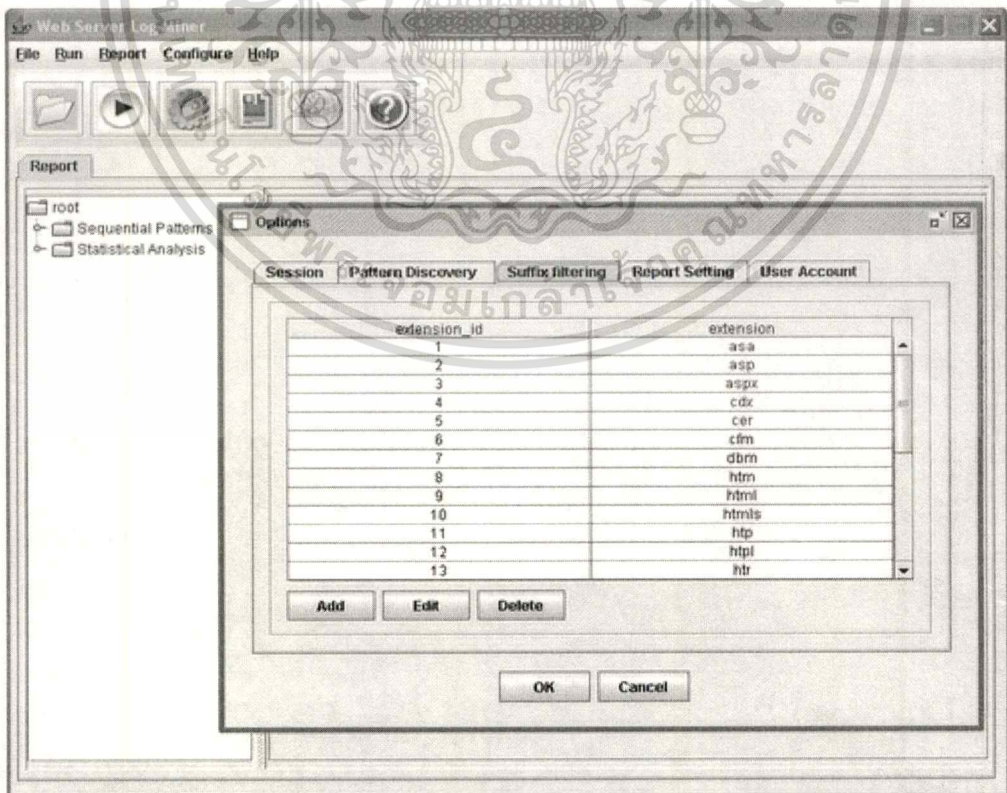


รูปที่ ก.7 หน้าจอการกำหนดเซสชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

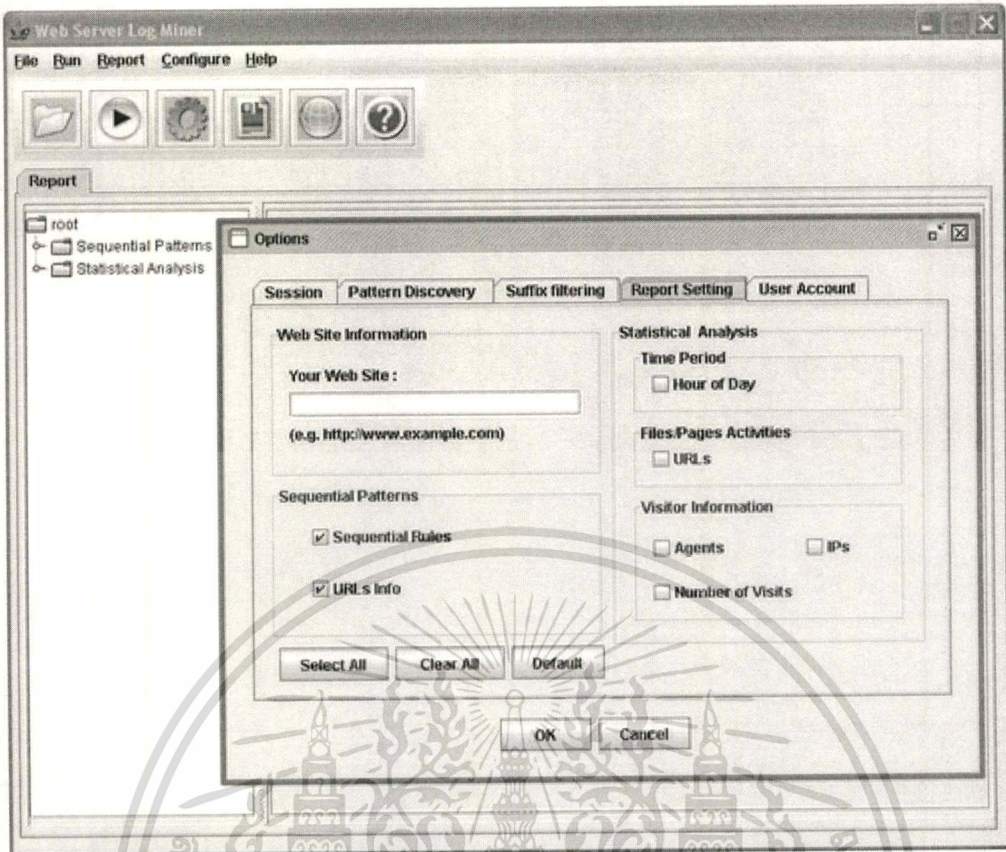


รูปที่ ก.8 หน้าจอการกำหนดค่าที่ใช้ในการค้นหารูปแบบลำดับ

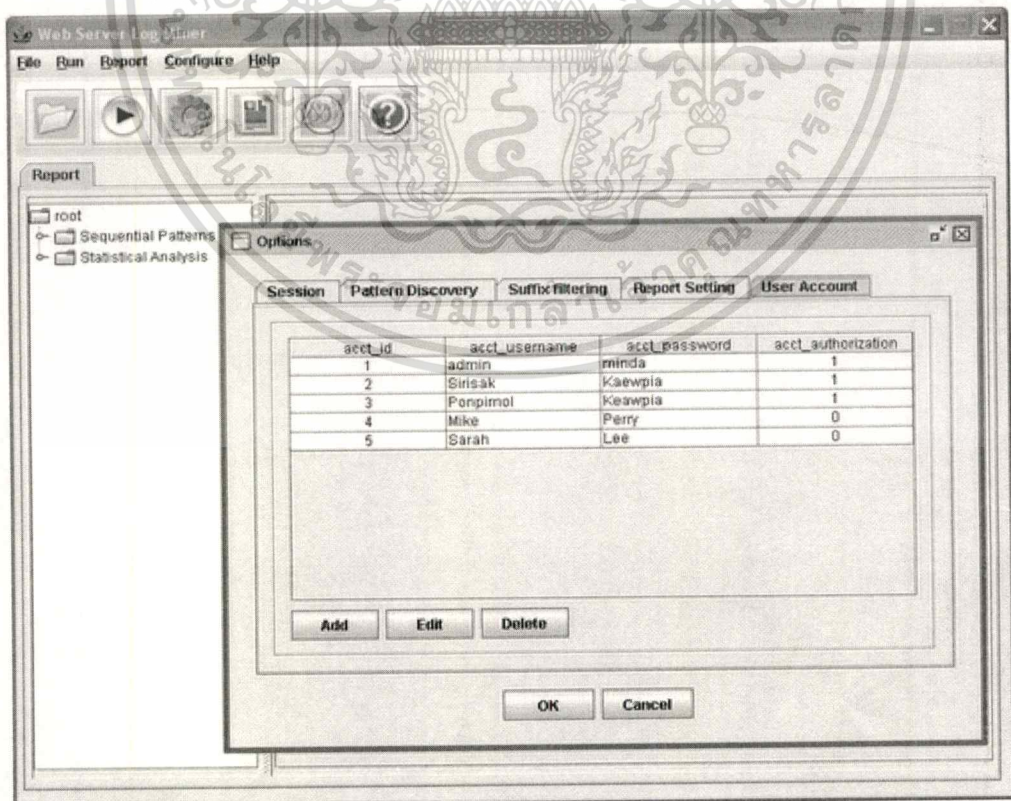


รูปที่ ก.9 หน้าจอการเพิ่ม แก้ไข หรือลบส่วนต่อท้ายที่ใช้กรอง

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือสงวนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีก้นำไปใช้

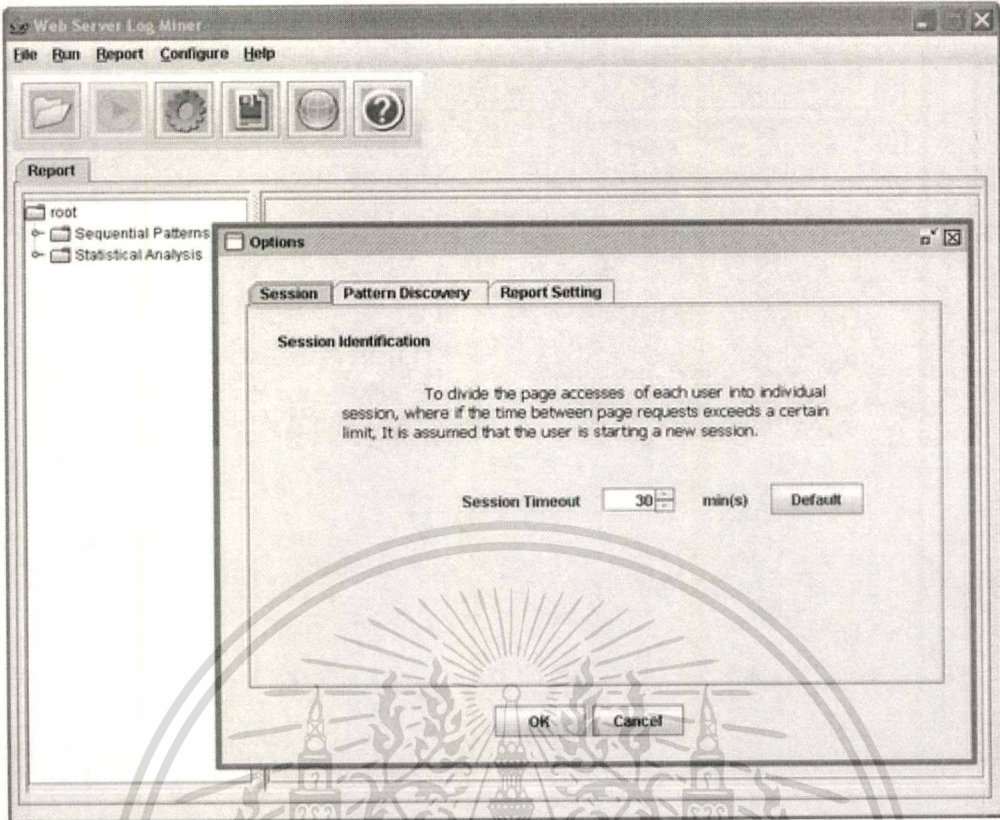


รูปที่ ก.10 หน้าจอการกำหนดชื่อเว็บไซต์และรายงานที่ต้องการ



รูปที่ ก.11 หน้าจอการเพิ่ม แก้ไขหรือลบรายชื่อผู้ใช้ในระบบ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.12 หน้าจอการกำหนดค่าของผู้ใช้ปกติ

การแสดงผลรายงาน

รายงานที่ระบบสร้างมี 2 ประเภทคือ รายงานรูปแบบลำดับและรายงานทางสถิติ

นักวิเคราะห์สามารถเลือกรายงานต่างๆ ได้โดยการเลือกที่ชื่อรายงานที่ต้องการในด้านซ้ายมือของหน้าจอรายงาน จากนั้น โปรแกรมจะแสดงรายงานทางด้านขวามือของหน้าจอรายงาน

รายงานรูปแบบลำดับ

1. รูปแบบลำดับขนาดต่างๆ โดยโปรแกรมจะมี ตั้งแต่ 2 – 10 ลำดับ ดังรูปที่ ก.13
2. URL ที่อ้างอิงกับตัวเลขที่ใช้แสดงในรูปแบบลำดับ ดังรูปที่ ก.14

รายงานทางสถิติ

1. สถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง ดังรูปที่ ก.15
2. สถิติแยกตามไอพีแอดเดรสของผู้ใช้ ดังรูปที่ ก.16
3. สถิติแยกตามเอเจนต์ ดังรูปที่ ก.17
4. สถิติแยกตามจำนวนการเข้าชมเว็บไซต์ ดังรูปที่ ก.18
5. สถิติแยกตามเว็บเพจที่เข้าชม ดังรูปที่ ก.19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมีให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

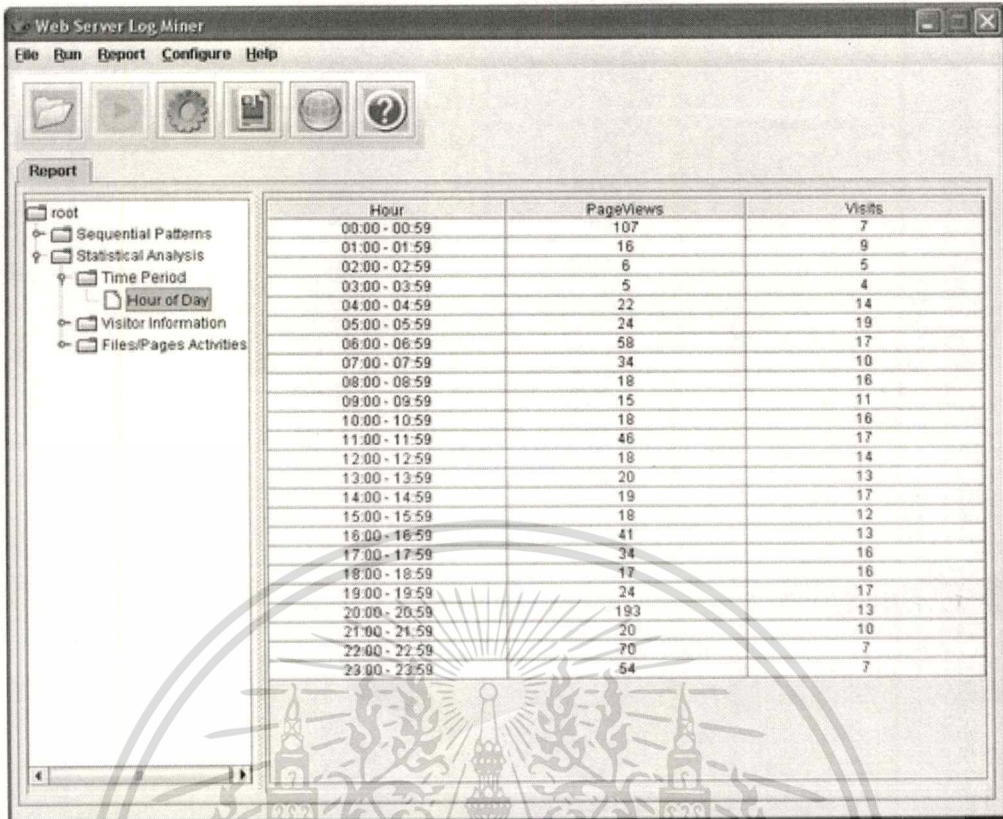
Rules_id	Sequence1	followed	Sequence2	Confidence	Support
1	(38)	==>	(4)	80	1.66
2	(38)	==>	(63)	80	1.66
3	(38)	==>	(84)	80	1.66
4	(38)	==>	(89)	80	1.66
5	(49)	==>	(4)	80	1.66
6	(49)	==>	(5)	80	1.66
7	(49)	==>	(15)	80	1.66
8	(49)	==>	(52)	80	1.66
9	(49)	==>	(54)	80	1.66
10	(50)	==>	(4)	100	1.66
11	(50)	==>	(5)	100	1.66
12	(50)	==>	(15)	100	1.66
13	(50)	==>	(52)	100	1.66
14	(50)	==>	(57)	100	1.66
15	(50)	==>	(62)	100	1.66
16	(50)	==>	(84)	100	1.66
17	(54)	==>	(4)	100	1.66
18	(56)	==>	(4)	100	1.66
19	(56)	==>	(38)	100	1.66
20	(56)	==>	(61)	100	1.66
21	(56)	==>	(62)	100	1.66
22	(56)	==>	(63)	100	1.66
23	(56)	==>	(84)	100	1.66
24	(57)	==>	(62)	100	1.66
25	(57)	==>	(84)	100	1.66
26	(59)	==>	(4)	80	1.66
27	(59)	==>	(38)	80	1.66
28	(59)	==>	(61)	80	1.66
29	(59)	==>	(62)	80	1.66

รูปที่ ก.13 หน้าจอแสดงรายงานรูปแบบลำดับ

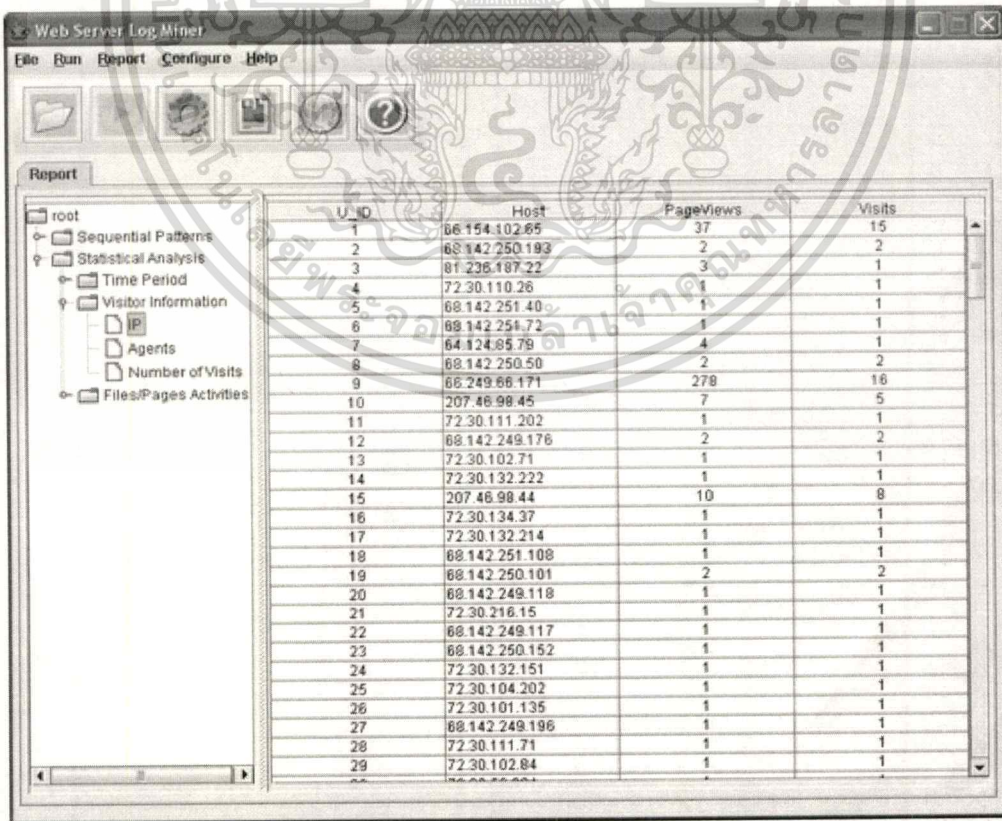
url_id	URL
1	/Thailand/Surat_Thani_Samui/Koh_Samui/Cabana_Grand_View_Hotel_Spa.php
2	/Thailand/Phuket/Gata_Beach/Peach_Hill_Hotel_Resort.php
3	/Vietnam/Chau_Doc_Town/Victoria_Chau_Doc_Hotel.php
4	/Thailand/Krabi/Koh_Lanta/SriLanta.php
5	/Thailand/Chiang_Mai/Old_Town/Rachamankha_Chiang_Mai.php
6	/home/index.php
7	/Cambodia/ot_Monivong_Boulevard/Sangkat_Wat_Phnorn/Raffles_Hotel_Le_Royal.php
8	/Thailand/Surat_Thani_Samui/Koh_Samui/Imperial_Boat_House.php
9	/Thailand/Trad/Koh_Chang_Surrounding_Islands/Aana_Resort_Spa.php
10	/Thailand/Chonburi/Pattaya/Rabbit_Resort.php
11	/Thailand/Bangkok/Sathon/Metropolitan_Hotel_Bangkok.php
12	/Thailand/Bangkok/Silom/Softtel_Silom.php
13	/Thailand/Surat_Thani_Samui/Koh_Samui/Beachcomber_Hotel.php
14	/Thailand/Surat_Thani_Samui/Koh_Samui/Pinnacle_Samui_Resort_Spa.php
15	/Thailand/Krabi/Muang_Town/Tubkaak_Resort_Krabi.php
16	/Thailand/Surat_Thani_Samui/Koh_Samui/Nara_Garden_Beach_Resort.php
17	/Thailand/Chiang_Mai/Doi_Suthep/Belle_Villa_Resort_Chiang_Mai.php
18	/Thailand/Bangkok/Sathon/Sukhothai_Bangkok.php
19	/Thailand/Bangkok/Huay_Kwang/Swissotel_Le_Concorde_Bangkok.php
20	/Thailand/Surat_Thani_Samui/Koh_Samui/Bandara_Resort_Spa.php
21	/Cambodia/Central_Park/Victoria_Angkor_Hotel.php
22	/Thailand/Trad/Koh_Chang_Surrounding_Islands/Amani_Emerald_Cove_Resort.php
23	/Thailand/Rayong/Koh_Samed/Mooban_Talay_Resort.php
24	/Thailand/Surat_Thani_Samui/Koh_Samui/Nora_Beach_Resort_Spa.php
25	/Thailand/Phuket/Cherng_Talay/Wilamanda_Laguna_Phuket.php
26	/Thailand/Trad/Koh_Chang_Surrounding_Islands/Klong_Prao_Resort_Hotel.php
27	/Thailand/Bangkok/Ploenchit/Plaza_Athenee.php
28	/Thailand/Trad/Koh_Chang_Surrounding_Islands/Ayapura_Resort_Spa.php
29	/packages/21.php

รูปที่ ก.14 หน้าจอแสดง URL ที่อ้างอิงกับตัวเลขที่ใช้แสดงในรูปแบบลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.15 หน้าจอแสดงสถิติแยกตามการเข้าใช้งานในแต่ละชั่วโมง



รูปที่ ก.16 หน้าจอแสดงสถิติแยกตามไอพีแอดเดรสของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Web Server Log Miner

File Run Report Configure Help

Report

- root
 - Sequential Patterns
 - Statistical Analysis
 - Time Period
 - Visitor Information
 - IP
 - Agents
 - Number of Visits
 - Files/Pages Activities

agent_id	agent
1	Gigabot/2.0/gigablast.com/spider.html
2	Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)
3	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)
4	Mozilla/5.0 (compatible; BecomeBot3.0; MSIE 6.0 compatible; +http://www.become.com/s
5	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
6	msnbot/1.0 (+http://search.msn.com/msnbot.htm)
7	Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.7.10) Gecko/20050724 Firefox/1.0.6
8	Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt)
9	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; InfoPath 1)
10	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; yplus 5.1.02
11	Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)
12	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; sbcydsi 3.12; YPC 3.0.3; SV1)
13	Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/312.8 (KHTML, like Gecko) Saf
14	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
15	Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)
16	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
17	Mozilla/4.0 (compatible; MSIE 5.17; Mac_PowerPC)
18	Mozilla/5.0 (Windows; U; Windows NT 5.1; it-IT; rv:1.7.5) Gecko/20041108 Firefox/1.0
19	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; lebar; MaxFiles; .NET CLR 1.1.4322)
20	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511
21	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2
22	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; FunWebProducts; .NET CLR 1.1
23	Mozilla/5.0 (compatible; LinksManager.com_bot +http://linksmanager.com/linkchecker.ht
24	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Q312461; SV1; .NET CLR 1.0.3705; ...
25	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; (R1 1.3); .NET CLR 1.1.4322)
26	Java/1.5.0
27	Secotier/3.3
28	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; FunWebProducts)
29	Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322; .NET CLR 2

รูปที่ ก.17 หน้าจอแสดงสถิติแยกตามเอเจนท์

Web Server Log Miner

File Run Report Configure Help

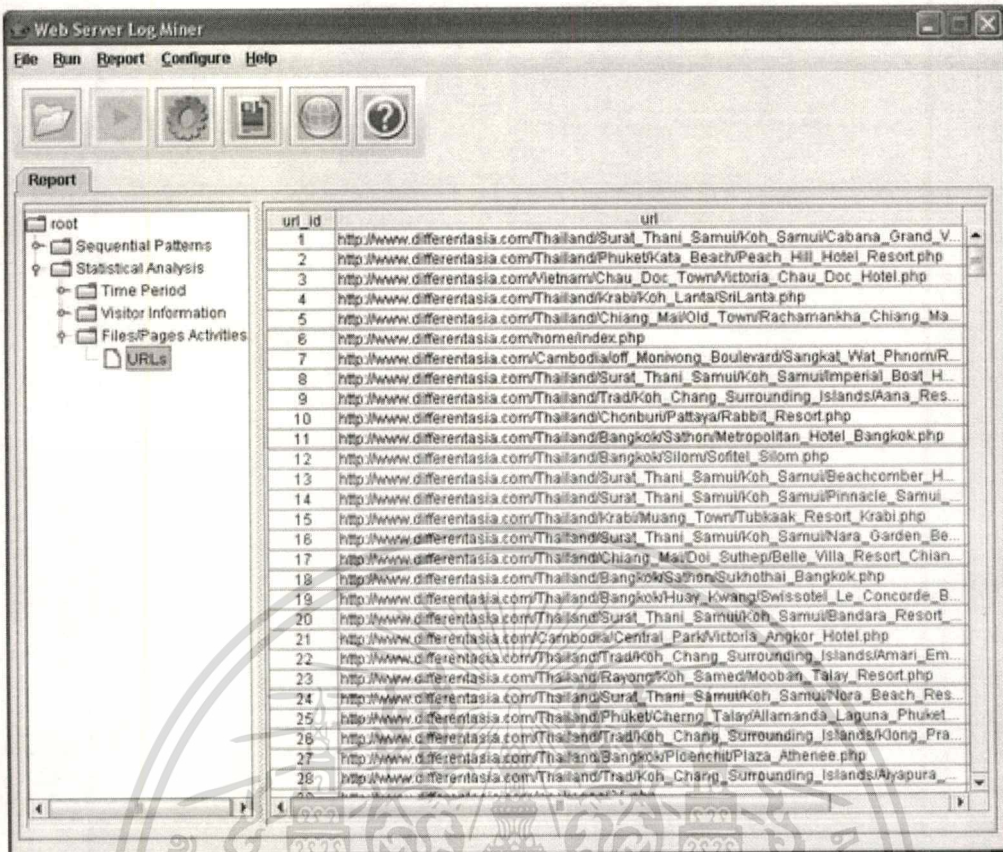
Report

- root
 - Sequential Patterns
 - Statistical Analysis
 - Time Period
 - Visitor Information
 - IP
 - Agents
 - Number of Visits
 - Files/Pages Activities

Visits	Visitors
1 Visits	224
2 Visits	13
3 Visits	0
4 Visits	0
5 Visits	1
6 Visits	0
7 Visits	0
8 Visits	1
9 Visits	0
10 Visits	0
11 Visits	0
12 Visits	0
13 Visits	0
14 Visits	0
15+ Visits	2

รูปที่ ก.18 หน้าจอแสดงสถิติแยกตามจำนวนการเข้าชมเว็บไซต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.19 หน้าจอแสดงสถิติแยกตามเว็บเพจที่เข้าชม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเรียงข้อมูลผลลัพธ์

เลือกที่ชื่อแอตทริบิวต์ใดๆในตารางแสดงผลที่ต้องการเรียง โดยจะมีเครื่องหมายแสดงการเรียงอยู่หลังชื่อแอตทริบิวต์ ตัวอย่างการเรียงข้อมูลจากการเรียงข้อมูลจากน้อยไปมากและการเรียงข้อมูลจากมากไปน้อยแสดงดังรูปที่ ค.20 และ ก.21

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
1	{38}	==>	{4}	80	1.66
2	{38}	==>	{63}	80	1.66
3	{38}	==>	{84}	80	1.66
4	{38}	==>	{89}	80	1.66
5	{49}	==>	{4}	80	1.66
6	{49}	==>	{5}	80	1.66
7	{49}	==>	{15}	80	1.66
8	{49}	==>	{52}	80	1.66
9	{49}	==>	{54}	80	1.66
26	{59}	==>	{4}	80	1.66
27	{59}	==>	{38}	80	1.66
28	{59}	==>	{61}	80	1.66
29	{59}	==>	{62}	80	1.66
30	{59}	==>	{63}	80	1.66
31	{59}	==>	{83}	80	1.66
32	{59}	==>	{84}	80	1.66
33	{60}	==>	{4}	80	1.66
34	{60}	==>	{61}	80	1.66
35	{60}	==>	{62}	80	1.66
36	{60}	==>	{63}	80	1.66
37	{60}	==>	{84}	80	1.66
38	{63}	==>	{4}	80	1.66
39	{63}	==>	{84}	80	1.66
10	{50}	==>	{4}	100	1.66
11	{50}	==>	{5}	100	1.66
12	{50}	==>	{15}	100	1.66
13	{50}	==>	{52}	100	1.66
14	{50}	==>	{57}	100	1.66
15	{50}	==>	{62}	100	1.66

รูปที่ ก.20 หน้าจอการเรียงข้อมูลจากน้อยไปมากที่ Confidence

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
10	(50)	==>	(4)	100	1.66
11	(50)	==>	(5)	100	1.66
12	(50)	==>	(15)	100	1.66
13	(50)	==>	(52)	100	1.66
14	(50)	==>	(57)	100	1.66
15	(50)	==>	(62)	100	1.66
16	(50)	==>	(84)	100	1.66
17	(54)	==>	(4)	100	1.66
18	(56)	==>	(4)	100	1.66
19	(56)	==>	(38)	100	1.66
20	(56)	==>	(61)	100	1.66
21	(56)	==>	(62)	100	1.66
22	(56)	==>	(63)	100	1.66
23	(56)	==>	(84)	100	1.66
24	(57)	==>	(62)	100	1.66
25	(57)	==>	(84)	100	1.66
1	(38)	==>	(4)	80	1.66
2	(38)	==>	(63)	80	1.66
3	(38)	==>	(84)	80	1.66
4	(38)	==>	(89)	80	1.66
5	(49)	==>	(4)	80	1.66
6	(49)	==>	(5)	80	1.66
7	(49)	==>	(15)	80	1.66
8	(49)	==>	(52)	80	1.66
9	(49)	==>	(54)	80	1.66
26	(59)	==>	(4)	80	1.66
27	(59)	==>	(38)	80	1.66
28	(59)	==>	(61)	80	1.66
29	(59)	==>	(62)	80	1.66

รูปที่ ก.21 หน้าจอการเรียงข้อมูลจากมากไปน้อยที่ Confidence

การออกรายงาน

การออกรายงานสามารถทำได้ โดยการเลือกที่ชื่อรายงานที่ต้องการจากนั้น เลือกวิธีการออกรายงาน ซึ่งมี 2 วิธี คือ Export และ Translate

- การออกรายงานด้วยวิธี Export

การออกรายงานด้วยวิธีนี้จะได้รายงานที่มีรูปแบบเหมือนกับรายงานที่แสดงในตัวโปรแกรม ตัวอย่างรายงานที่ได้จากการ Export แสดงดังรูปที่ ขึ้นตอนการออกรายงานเริ่มจาก

1. เลือกที่ชื่อรายงานที่ต้องการออก ในด้านซ้ายมือของหน้าจอรายงาน
2. เลือกที่เมนู Report -> Export หรือปุ่ม Export

- การออกรายงานด้วยวิธี Translate

การออกรายงานด้วยวิธีนี้จะได้รายงานรูปแบบลำดับที่แสดงรูปแบบลำดับด้วย URL

1. เลือกที่ชื่อรายงานที่ต้องการออกในด้านซ้ายมือของหน้าจอรายงาน
2. เลือกที่เมนู Report -> Translate หรือปุ่ม Translate

หมายเหตุ รายงานที่จะออกด้วยวิธี Translate ได้นั้นจะต้องเป็นรายงานรูปแบบลำดับเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2-SEQUENCE

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 50.0 %

Rules_id	Sequence1	followed	Sequence2	Confidence	Support
1	{38}	==>	{4}	80.0	1.66
2	{38}	==>	{63}	80.0	1.66
3	{38}	==>	{84}	80.0	1.66
4	{38}	==>	{89}	80.0	1.66
5	{49}	==>	{4}	80.0	1.66
6	{49}	==>	{5}	80.0	1.66
7	{49}	==>	{15}	80.0	1.66
8	{49}	==>	{52}	80.0	1.66
9	{49}	==>	{54}	80.0	1.66
10	{50}	==>	{4}	100.0	1.66
11	{50}	==>	{5}	100.0	1.66
12	{50}	==>	{15}	100.0	1.66
13	{50}	==>	{52}	100.0	1.66
14	{50}	==>	{57}	100.0	1.66
15	{50}	==>	{62}	100.0	1.66
16	{50}	==>	{84}	100.0	1.66
17	{54}	==>	{4}	100.0	1.66
18	{55}	==>	{4}	66.67	1.66
19	{55}	==>	{57}	66.67	1.66
20	{55}	==>	{62}	66.67	1.66
21	{55}	==>	{63}	66.67	1.66
22	{55}	==>	{83}	66.67	1.66
23	{55}	==>	{84}	66.67	1.66
24	{56}	==>	{4}	100.0	1.66
25	{56}	==>	{38}	100.0	1.66
26	{56}	==>	{61}	100.0	1.66
27	{56}	==>	{62}	100.0	1.66
28	{56}	==>	{63}	100.0	1.66
29	{56}	==>	{84}	100.0	1.66
30	{57}	==>	{62}	100.0	1.66
31	{57}	==>	{84}	100.0	1.66
32	{59}	==>	{4}	80.0	1.66
33	{59}	==>	{38}	80.0	1.66

23/02/07 12:15

Page 1 of 2

รูปที่ ก.22 ตัวอย่างรายงานรูปแบบลำดับที่ได้จากวิธี Export

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4-SEQUENCE

14/May/2006

Minimum Support : 1.5 % Minimum Confidence : 50.0 %

Rules_id : 1
 IF
<http://www.differentasia.com/index.php>
 THEN
http://www.differentasia.com/destinations/Cha_Am.php
<http://www.differentasia.com/destinations/Krabi.php>
<http://www.differentasia.com/faq/faq.php>

Confidence : 100.0

Support : 1.66

Rules_id : 2
 IF
<http://www.differentasia.com/index.php>
http://www.differentasia.com/destinations/Cha_Am.php
 THEN
<http://www.differentasia.com/destinations/Krabi.php>
<http://www.differentasia.com/faq/faq.php>

Confidence : 100.0

Support : 1.66

Rules_id : 3
 IF
<http://www.differentasia.com/index.php>
http://www.differentasia.com/destinations/Cha_Am.php
<http://www.differentasia.com/destinations/Krabi.php>
 THEN
<http://www.differentasia.com/faq/faq.php>

Confidence : 100.0

Support : 1.66

23/02/07 12:28

Page 1 of 1

รูปที่ ก.23 ตัวอย่างรายงานรูปแบบลำดับที่ได้จากวิธี Translate

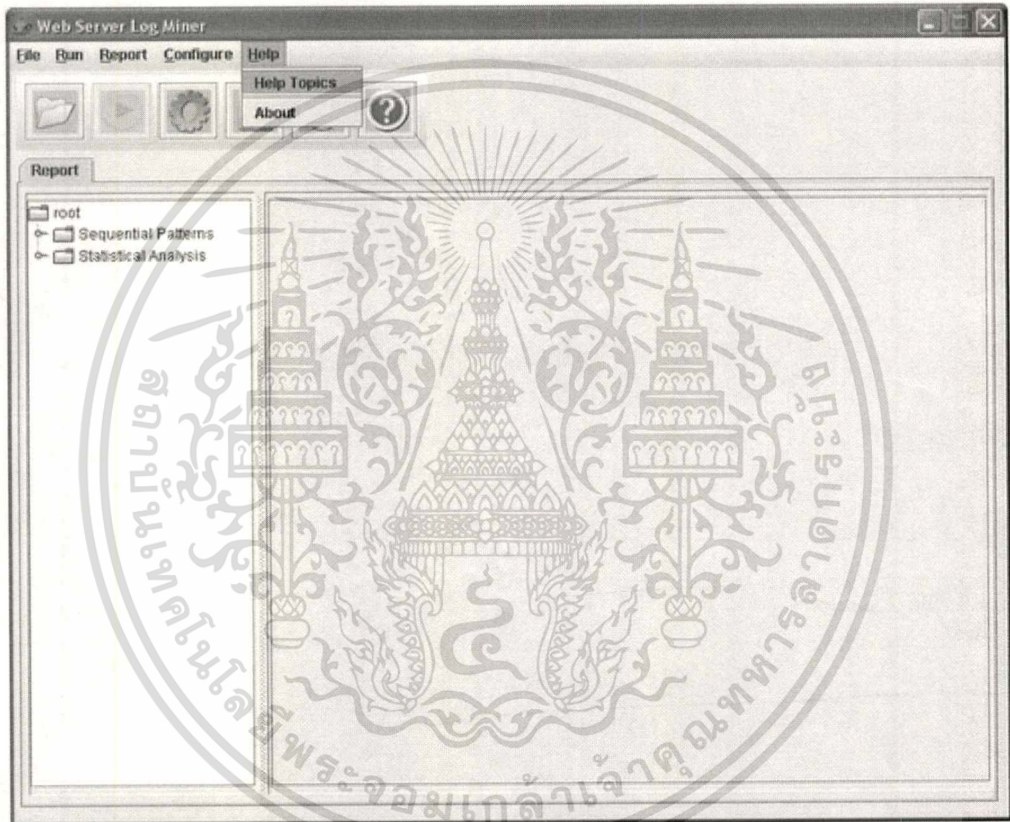
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การขอความช่วยเหลือ

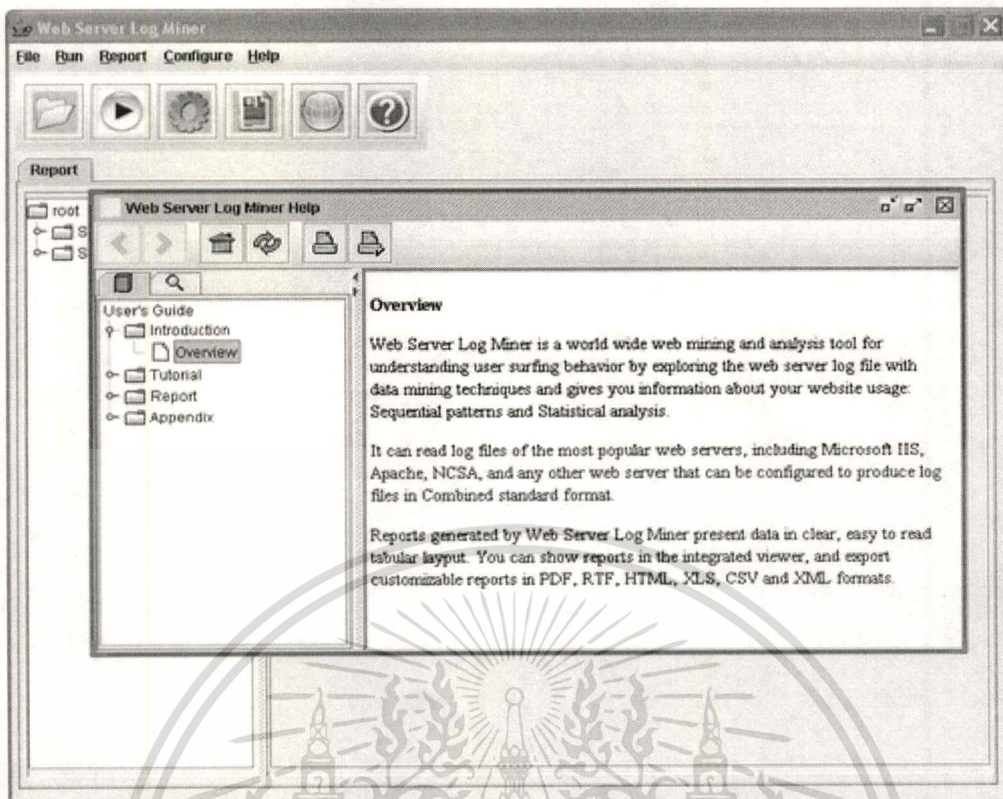
1. เลือกที่เมนู Help Topics หรือที่ปุ่ม Help เพื่อแสดงหน้าจอให้ความช่วยเหลือ ดังรูปที่ ก.24
2. เลือกหัวข้อที่ต้องการหรือค้นหาด้วยคำสำคัญที่ต้องการ ดังรูปที่ ก.25

การแสดงรายละเอียดเกี่ยวกับโปรแกรม

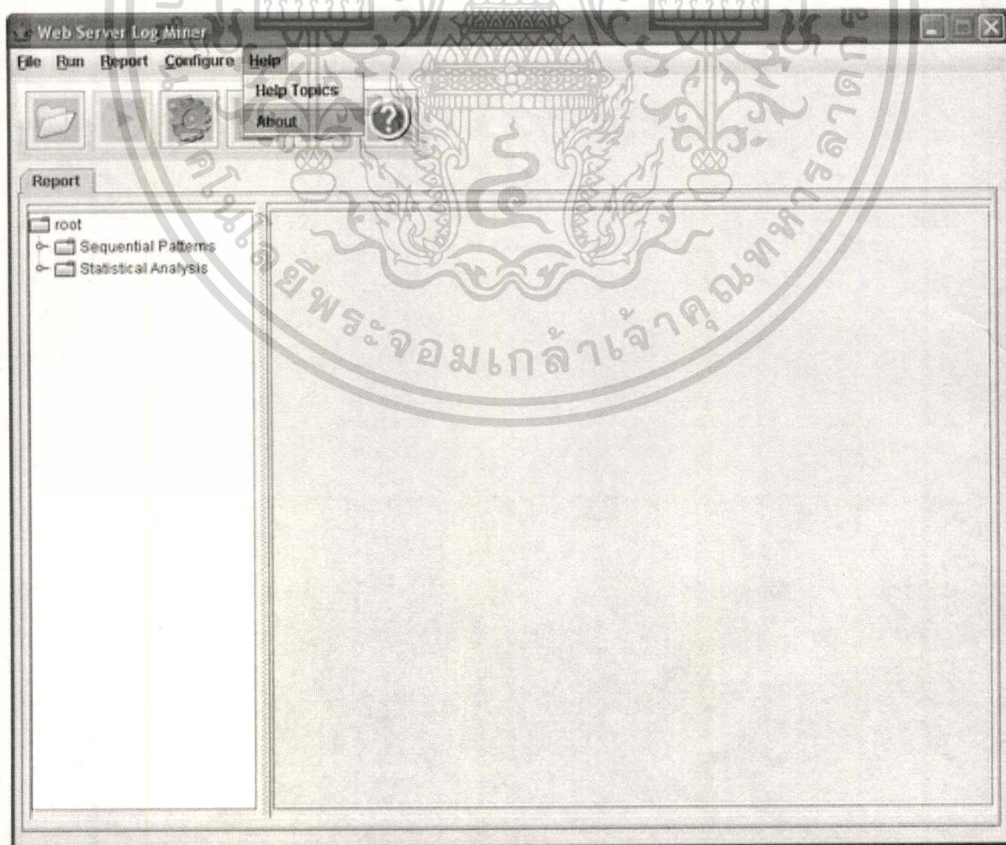
1. เลือกเมนู Help -> About เพื่อดูรายละเอียดเกี่ยวกับ โปรแกรม ดังรูปที่ ก.26
2. โปรแกรมจะแสดงรายละเอียดดังรูปที่ ก.27



รูปที่ ก.24 หน้าจอการขอความช่วยเหลือ



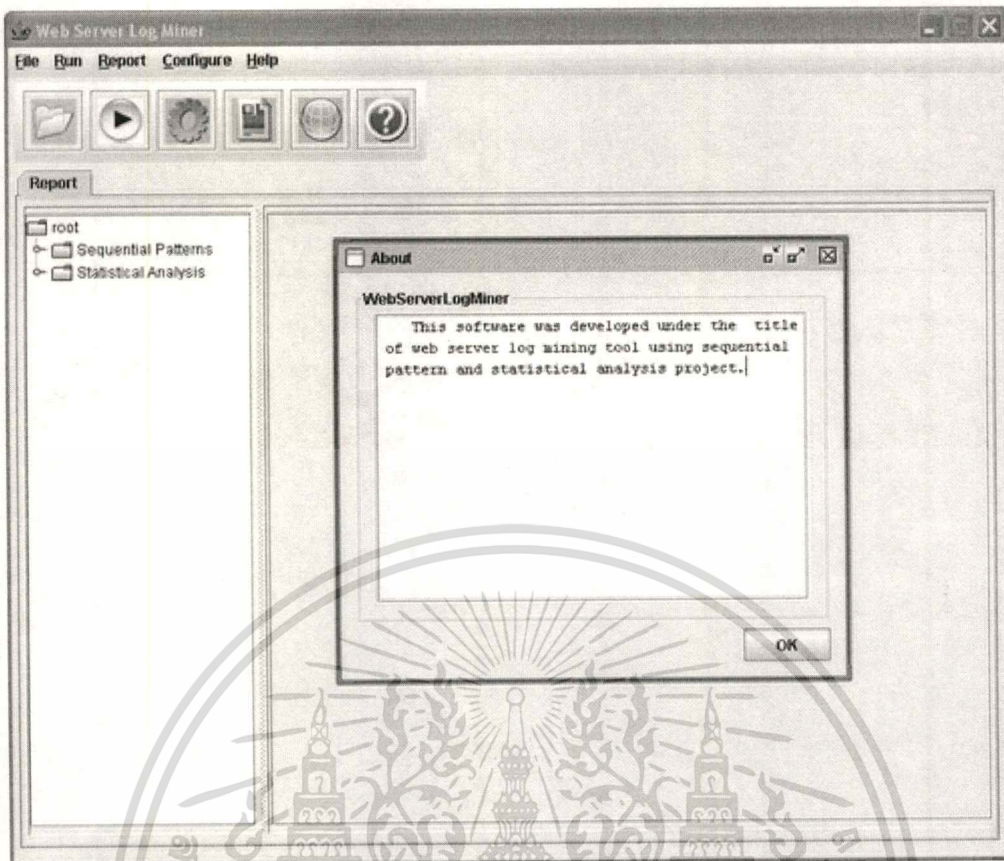
รูปที่ ก.25 หน้าจอการให้ความช่วยเหลือ



รูปที่ ก.26 หน้าจอการขอรายละเอียดเกี่ยวกับโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.27 หน้าจอแสดงรายละเอียดเกี่ยวกับโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นายศิริศักดิ์ แก้วเป็ย
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	วิทยาศาสตรบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีที่สำเร็จการศึกษา	2547



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้