

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การประยุกต์ใช้การค้นหากฎความสัมพันธ์กับระบบห้องสมุด

APPLICATION OF ASSOCIATION RULES DISCOVERY WITH
LIBRARY SYSTEM



อพ.
๗๖๑ก
๒๕๕๐

เลขหมู่.....
เลขทะเบียน..... 04458
วัน,เดือน,ปี..... 12 ส.ย. 2551

Barcode
H004458

b. 11923830
i.

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 1 ปีการศึกษา 2550

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**APPLICATION OF ASSOCIATION RULES DISCOVERY WITH
LIBRARY SYSTEM**



SUCHART KAEWYOTE

**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**

1/ 2007

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2007

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การประยุกต์ใช้การค้นหากฎความสัมพันธ์กับระบบห้องสมุด
นักศึกษา	นายสุชาติ แก้วยศ
รหัสนักศึกษา	47066633
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2550
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสูระเดช

บทคัดย่อ

โครงการนี้มีวัตถุประสงค์เพื่อเป็นการประยุกต์ใช้การค้นหากฎความสัมพันธ์ (Association Rules Discovery) ร่วมกับระบบห้องสมุดโดยใช้ความสามารถในส่วนคาด้าไมน์นิงของระบบจัดการฐานข้อมูล SQL Server 2005 โดยการนำข้อมูลรายการยืมของสมาชิกมาใช้ในการสร้างกฎความสัมพันธ์เพื่อช่วยในการแนะนำรายการทรัพยากรห้องสมุดที่มีความสัมพันธ์กันให้กับผู้ใช้บริการสืบค้นข้อมูลของระบบห้องสมุด โดยในการออกแบบและพัฒนาระบบครั้งนี้ได้ทำการพัฒนาโปรแกรมประยุกต์เพื่อใช้ในการสร้างและประมวลผลแบบจำลองคาด้าไมน์นิงบนระบบจัดการฐานข้อมูล SQL Server 2005, การพัฒนาเว็บเซอร์วิสที่ให้บริการในการทำนายกฎความสัมพันธ์ และการพัฒนาโปรแกรมประยุกต์แบบเว็บในการให้บริการสืบค้นข้อมูลแก่ผู้ใช้บริการห้องสมุดซึ่งจะเรียกใช้บริการเว็บเซอร์วิสที่สร้างขึ้น การพัฒนาระบบนี้จะช่วยอำนวยความสะดวกให้กับผู้ใช้บริการห้องสมุดและช่วยให้ผู้ใช้บริการสามารถใช้บริการห้องสมุดได้อย่างมีประสิทธิภาพมากยิ่งขึ้น นอกจากนี้แล้วยังสามารถประยุกต์ใช้ระบบนี้กับงานอื่นๆ ที่เกี่ยวข้องกับการค้นหากฎความสัมพันธ์ได้อีกด้วย

Title	Application of Association Rules Discovery with Library System
Student	Mr. Suchart Kaewyote
Student ID.	47066633
Degree	Master of Science
Programme	Information Science
Academic Year	2007
Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej

ABSTRACT

This report applies association rules discovery with library system. The study applies data mining feature of SQL Server 2005 to develop the recommendation engine for library system by using library member's circulation records to generate mining models and association rules. The study includes design and development of application to create and process mining models on SQL Server 2005, development of web services to predict association rules from the mining model and development of library system web application to consume the web service. The system will assist library system users to search for their interesting library resources effectively. In addition, it may be applied to other tasks that associate with association rules discovery.

กิตติกรรมประกาศ

โครงการพัฒนาระบบงานฉบับนี้สำเร็จได้อย่างดี ด้วยคำแนะนำ และคำปรึกษาจาก รศ.ดร. วรพจน์ กริสุระเดช ซึ่งเป็นอาจารย์ที่ปรึกษาโครงการ นอกจากนี้ ยังได้รับความกรุณาจากอาจารย์ กรรมการอีกสองท่านที่เสียสละเวลามาเป็นกรรมการสอบ คอยให้คำแนะนำและตรวจสอบ จนกระทั่งเสร็จสมบูรณ์ ข้าพเจ้ารู้สึกทราบบ้างในความอนุเคราะห์จากท่านอาจารย์ทุกท่าน และขอขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณคณาจารย์คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมา

ขอขอบคุณบัณฑิตศึกษาและบัณฑิตวิทยาลัยคณะเทคโนโลยีสารสนเทศ ที่ให้ความช่วยเหลือ ในเรื่องต่างๆ

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำโครงการพัฒนาระบบงานฉบับนี้สำเร็จลุล่วงด้วยดี

คุณค่าและประโยชน์อันพึงมาจากโครงการพัฒนาระบบงานฉบับนี้ ข้าพเจ้าขอบพระคุณแต่ผู้มีพระคุณทุกท่าน

สุชาติ แก้วยศ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	1
1.3 ขอบเขตการดำเนินการ.....	1
1.4 ขั้นตอนการดำเนินการ.....	2
1.5 เครื่องมือที่ใช้ในการดำเนินการ.....	2
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 คาด้าไมน์นิง.....	3
2.1 ความหมายของคาด้าไมน์นิง.....	3
2.2 งานของคาด้าไมน์นิง	4
2.2.1 การจัดหมวดหมู่.....	4
2.2.2 การรวมกลุ่ม.....	4
2.2.3 การค้นหาภูควมสัมพันธ์	5
2.2.4 การประเมินค่า.....	5
2.2.5 การทำนาย	5
2.2.6 การหารายละเอียด.....	5
2.3 กระบวนการทำคาด้าไมน์นิง.....	5
2.3.1 การกำหนดเป้าหมายทางธุรกิจ.....	5
2.3.2 การรวบรวมข้อมูล.....	5
2.3.3 การเตรียมข้อมูล	5
2.3.4 การสร้างแบบจำลอง	6

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.3.5 การประเมินผลแบบจำลอง	6
2.4 การค้นหากฎความสัมพันธ์	6
2.5 อัลกอริทึม Apriori	8
2.5.1 การค้นหา Frequent Itemset.....	8
2.5.2 การสร้างกฎความสัมพันธ์.....	12
2.6 SQL Server 2005.....	13
2.6.1 SQL Server Business Intelligence.....	13
2.6.2 ค่าค่าไมน์นิงของ SQL Server	14
2.7 OLE DB สำหรับค่าค่าไมน์นิง.....	14
2.8 อัลกอริทึม Microsoft Association Rules.....	16
บทที่ 3 การวิเคราะห์และออกแบบระบบงาน.....	18
3.1 ลักษณะการทำงานของระบบงานปัจจุบัน.....	18
3.2 ปัญหาที่พบในระบบงานปัจจุบัน.....	18
3.3 การศึกษาความต้องการของระบบงานใหม่.....	19
3.4 ยูสเคสของระบบงานใหม่.....	19
3.5 ซีเควนซ์ไดอะแกรมของระบบงานใหม่.....	27
3.6 คลาสไดอะแกรมของระบบงานใหม่.....	30
3.7 การออกแบบฐานข้อมูลของระบบงานใหม่.....	32
บทที่ 4 การพัฒนาระบบงาน.....	36
4.1 ความต้องการของการพัฒนาระบบ.....	36
4.2 การทำค่าค่าไมน์นิง.....	36
4.2.1 กำหนดวัตถุประสงค์.....	36
4.2.2 การรวบรวมข้อมูล.....	37
4.2.3 การเตรียมข้อมูล.....	37
4.2.4 การทำค่าค่าไมน์นิง.....	37
4.3 การพัฒนาโปรแกรม.....	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญเตเห็นเบาะเบาะระเียนดานการคำ

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.4 โปรแกรมสร้างและประมวลผลแบบจำลอง ไม่นิ่ง.....	37
4.5 เว็บเซอร์วิสสำหรับทำนายกฏความสัมพันธ์.....	40
4.6 ระบบสืบค้นข้อมูลห้องสมุดผ่านเว็บ.....	41
บทที่ 5 สรุปผลการดำเนินงาน.....	47
5.1 สรุปผลการดำเนินงาน.....	47
5.2 ข้อเสนอแนะ.....	47
บรรณานุกรม	48
ภาคผนวก ก คู่มือการติดตั้งระบบ.....	49
ภาคผนวก ข คู่มือการใช้งานระบบ.....	55
ประวัติผู้เขียน	63

สารบัญตาราง

ตารางที่	หน้า
3.1 รายละเอียดประกอบชุดทดสอบ Search Catalog.....	21
3.2 รายละเอียดประกอบชุดทดสอบ View Biblio Detail.....	22
3.3 รายละเอียดประกอบชุดทดสอบ Predict Association.....	23
3.4 รายละเอียดประกอบชุดทดสอบ Review Biblio.....	23
3.5 รายละเอียดประกอบชุดทดสอบ Login.....	24
3.6 รายละเอียดประกอบชุดทดสอบ Logout.....	24
3.7 รายละเอียดประกอบชุดทดสอบ Create New Mining Project.....	25
3.8 รายละเอียดประกอบชุดทดสอบ Reprocess Mining Database.....	26
3.9 ตาราง biblio.....	33
3.10 ตาราง publishers.....	33
3.11 ตาราง users.....	34
3.12 ตาราง user_reviews.....	34
3.13 ตาราง LibCir.....	35

สารบัญรูป

รูปที่	หน้า
2.1 คาด้าไมน์นิ่งเป็นขั้นตอนหนึ่งในกระบวนการค้นหาความรู้	4
2.2 ขั้นตอนการค้นหาทฤษฎีความสัมพันธ์	8
2.3 ฐานข้อมูล D.....	9
2.4 วิธีการหา Frequent Itemset	10
2.5 อัลกอริทึม Apriori	12
2.6 สถาปัตยกรรมคาด้าไมน์นิ่งของ SQL Server 2005	15
2.7 สถาปัตยกรรมพื้นฐานของ OLE DB for Data Mining	15
2.8 ตารางเคสแบบ Nested	17
3.1 ยูสเคสไคอะแกรมของระบบงานใหม่.....	20
3.2 ซีเควนซ์ไคอะแกรมของยูสเคส Search Catalog.....	27
3.3 ซีเควนซ์ไคอะแกรมของยูสเคส Login และ Logout.....	28
3.4 ซีเควนซ์ไคอะแกรมของยูสเคส View Biblio Detail.....	29
3.5 คลาสไคอะแกรมของเว็บระบบห้องสมุด.....	30
3.6 คลาสไคอะแกรมของการจัดการคาด้าไมน์นิ่ง.....	31
3.7 อีอาร์ไคอะแกรมของระบบเว็บห้องสมุด.....	32
3.8 อีอาร์ไคอะแกรมของข้อมูลที่ใช้ในการค้นหาทฤษฎีความสัมพันธ์.....	35
4.1 เมนูหลักของโปรแกรมจัดการคาด้าไมน์นิ่ง.....	38
4.2 การสร้างและประมวลผลฐานข้อมูลใหม่.....	39
4.3 การประมวลผลฐานข้อมูลที่ได้สร้างไว้แล้ว.....	40
4.4 WSDL ของเว็บเซอร์วิสที่ใช้ทำนายทฤษฎีความสัมพันธ์.....	41
4.5 หน้าโฮมเพจของเว็บระบบห้องสมุด.....	42
4.6 ส่วนของการสืบค้นข้อมูล.....	43
4.7 หน้าแสดงผลลัพธ์จากการสืบค้นข้อมูล.....	43
4.8 หน้าแสดงรายละเอียดของทรัพยากรห้องสมุด.....	44
4.9 หน้าแสดงส่วนของการล็อกอินเข้าสู่ระบบ.....	45
4.10 ส่วนของการเขียนแสดงความคิดเห็นของผู้ใช้งาน.....	46

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ห้องสมุดนับได้ว่าเป็นแหล่งความรู้ที่สำคัญในการศึกษาทุกระดับชั้น สถาบันการศึกษา และองค์กรต่างๆ มีห้องสมุดซึ่งมีหนังสือเป็นจำนวนมาก แต่บ่อยครั้งที่ผู้ใช้บริการห้องสมุดไม่ทราบว่าควรจะยืมหนังสือเล่มไหนดี หรือไม่ทราบว่าหนังสือเล่มไหนที่เกี่ยวข้องหรือมีความสัมพันธ์กัน หรือใช้เวลานานในการค้นหาหนังสือที่ต้องการ ด้วยเหตุนี้ทำให้ผู้ใช้บริการไม่ได้รับความสะดวกและไม่สามารถใช้บริการห้องสมุดได้อย่างมีประสิทธิภาพ

แนวคิดในการทำโครงการนี้จึงเกิดขึ้น โดยโครงการนี้จะทำการประยุกต์ใช้การค้นหาคู่ความสัมพันธ์ร่วมกับระบบห้องสมุดเพื่อให้ระบบช่วยแนะนำหนังสือที่มีความเกี่ยวข้องหรือสัมพันธ์กันให้กับผู้ใช้บริการห้องสมุด โดยอาศัยข้อมูลจากประวัติการยืมหนังสือของสมาชิกห้องสมุด

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

ในการดำเนินโครงการนี้ มีวัตถุประสงค์ต่างๆ ดังต่อไปนี้

1. เพื่อศึกษาและประยุกต์ใช้การค้นหาคู่ความสัมพันธ์ร่วมกับระบบห้องสมุด
2. เพื่อพัฒนาระบบค้นหาคู่ความสัมพันธ์เพื่อใช้งานร่วมกับระบบห้องสมุด
3. เพื่อปรับปรุงระบบการสืบค้นข้อมูลของระบบห้องสมุด
4. เพื่ออำนวยความสะดวกในการใช้งานให้ผู้ใช้บริการห้องสมุด
5. เพื่อให้ผู้ใช้บริการห้องสมุดสามารถใช้งานห้องสมุดได้อย่างมีประสิทธิภาพ

1.3 ขอบเขตการดำเนินการ

โครงการนี้ เป็นการประยุกต์ใช้การค้นหาคู่ความสัมพันธ์กับระบบห้องสมุด โดยนำประวัติการยืมหนังสือของสมาชิกห้องสมุดมาสร้างเป็นกฎความสัมพันธ์เพื่อหารายการหนังสือที่มักจะถูกยืมไปด้วยกันเสมอๆ โดยจัดเก็บอยู่ในรูปของโมดูลหนึ่งโมดูลซึ่งเป็นออบเจกต์หนึ่งของ SQL Server 2005 โดยมีขอบเขตในการดำเนินการ ดังนี้

1. การจัดเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์ค้นหาคู่ความสัมพันธ์ของทรัพยากรห้องสมุด
2. การพัฒนาโปรแกรมประยุกต์เพื่อใช้สำหรับการสร้างและจัดการแบบโมดูลหนึ่งโมดูล
3. การพัฒนาเว็บเซอร์วิสเพื่อใช้ในการทำนาคู่ความสัมพันธ์

4. การพัฒนาโปรแกรมประยุกต์แบบเว็บเพื่อให้บริการสืบค้นทรัพยากรห้องสมุด

1.4 ขั้นตอนการดำเนินการ

ในการพัฒนาระบบ มีขั้นตอนในการดำเนินการ ดังนี้

1. ศึกษารูปแบบและขั้นตอนการดำเนินงานเดิมของระบบห้องสมุด
2. วิเคราะห์ปัญหาและอุปสรรคที่เกิดขึ้นจากรูปแบบการทำงานเดิม
3. หาแนวทางในการแก้ปัญหาและอุปสรรคที่ได้จากการวิเคราะห์รูปแบบการทำงานเดิม
4. ศึกษาการพัฒนาโปรแกรมประยุกต์แบบเว็บโดยใช้เฟรมเวิร์ค Ruby on Rails และการพัฒนาเว็บเซอร์วิสโดยใช้เทคโนโลยี .NET
5. วิเคราะห์และออกแบบระบบงาน
6. พัฒนาระบบงาน
7. ทดสอบการใช้งานและแก้ไขปรับปรุงระบบงานที่พัฒนาแล้ว
8. สรุปผลการใช้งานระบบใหม่ที่พัฒนาขึ้น
9. จัดทำเอกสารประกอบโครงการ

1.5 เครื่องมือที่ใช้ในการดำเนินการ

ในการพัฒนาระบบครั้งนี้ได้มีการใช้งานเครื่องมือหรือซอฟต์แวร์ต่างๆ ดังต่อไปนี้

- Microsoft .NET Framework 2.0
- Microsoft Visual Studio 2005
- Microsoft SQL Server 2005
- Ruby on Rails Framework
- NetBeans IDE 6.0
- MySQL
- Internet Information Services (IIS)

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ช่วยอำนวยความสะดวกให้กับผู้ใช้งานระบบห้องสมุดในการสืบค้นทรัพยากรห้องสมุด
2. เพื่อให้ผู้ใช้งานห้องสมุดสามารถใช้งานห้องสมุดได้อย่างมีประสิทธิภาพ
3. สามารถนำไปประยุกต์ใช้กับงานต่างๆ ที่มีความเกี่ยวข้องกับการค้นหา ความสัมพันธ์ เช่น การขายสินค้า ได้ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ดาต้าไมน์นิง

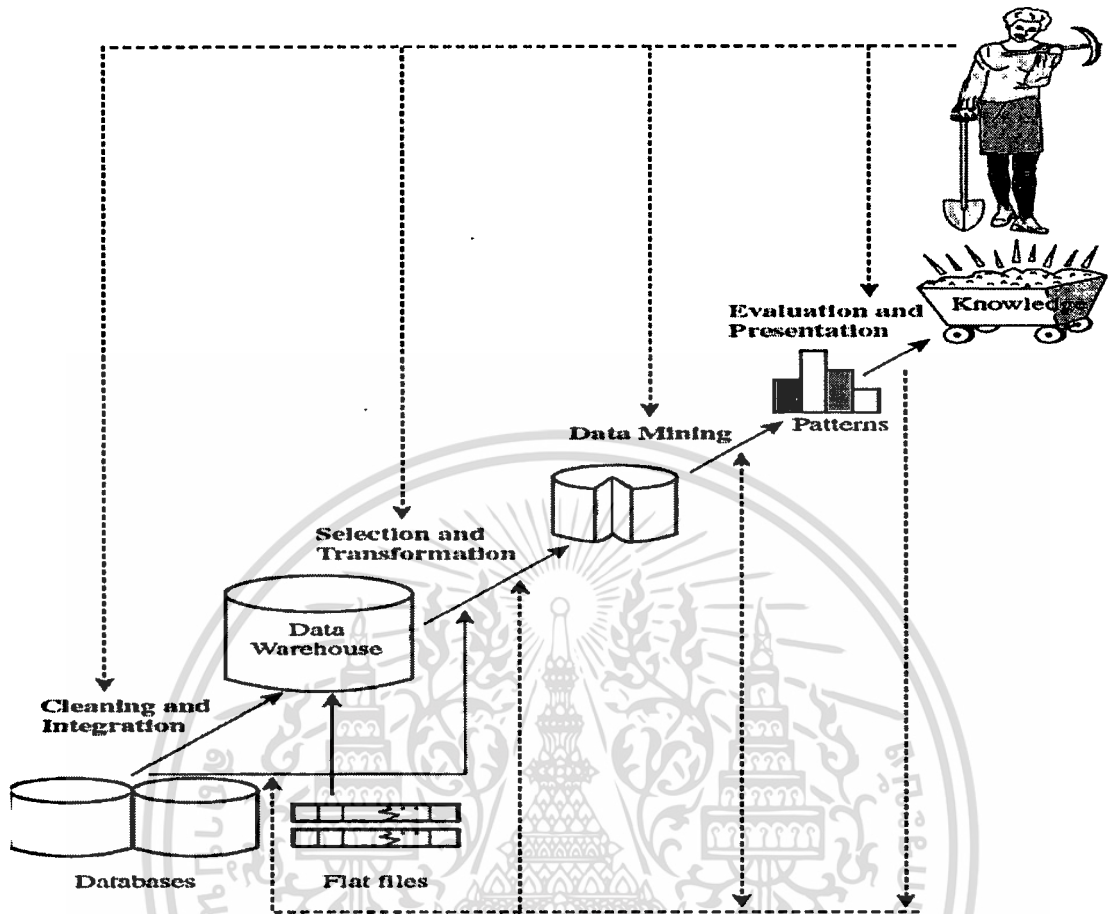
ในบทนี้จะกล่าวถึงทฤษฎีพื้นฐาน และหลักการทั่วไปที่เกี่ยวข้องในการทำโครงการ โดยเนื้อหาในบทนี้จะกล่าวถึง ความรู้พื้นฐานเกี่ยวกับดาต้าไมน์นิง (Data Mining), การค้นหากฎความสัมพันธ์ (Association Rule Discovery), อัลกอริทึม Apriori, SQL Server 2005, OLE DB สำหรับดาต้าไมน์นิง (OLE DB for Data Mining) และอัลกอริทึม Microsoft Association Rules

2.1 ความหมายของดาต้าไมน์นิง

ดาต้าไมน์นิง (Data Mining) หรือการทำเหมืองข้อมูล หมายถึง ขั้นตอนหรือกระบวนการในการค้นหาความรู้ที่มีคุณค่าจากข้อมูลที่มีขนาดใหญ่ บางคนมองว่าดาต้าไมน์นิงมีความหมายเช่นเดียวกับคำว่า Knowledge Discovery from Data หรือ KDD ในขณะที่บางคนก็มองว่า ดาต้าไมน์นิงเป็นขั้นตอนหนึ่งในกระบวนการค้นหาความรู้ (Knowledge Discovery) ขั้นตอนการค้นหาความรู้แสดงได้ดังรูปที่ 2.1 ซึ่งประกอบไปด้วยขั้นตอนต่างๆ ดังนี้

1. การทำความสะอาดข้อมูล (Data cleaning)
2. การรวมข้อมูล (Data integration)
3. การคัดเลือกข้อมูล (Data selection)
4. การแปลงข้อมูล (Data transformation)
5. ดาต้าไมน์นิง
6. การประเมินแบบอย่างข้อมูล (Pattern evaluation)
7. การนำเสนอความรู้ (Knowledge presentation)

ขั้นตอนที่ 1 ถึง 4 เป็นรูปแบบต่างๆ ของการเตรียมข้อมูล (Data preparation) เพื่อให้ข้อมูลพร้อมสำหรับการทำดาต้าไมน์นิง



รูปที่ 2.1 คาด้าไมน์นิ่งเป็นขั้นตอนหนึ่งในกระบวนการค้นหาความรู้

2.2. งานของคาด้าไมน์นิ่ง (Data Mining Tasks)

งานของคาด้าไมน์นิ่งสามารถแบ่งเป็นกลุ่มต่างๆ ได้ดังนี้

2.2.1 การจัดหมวดหมู่ (Classification)

การจัดหมวดหมู่ถือว่าเป็นหนึ่งในงานที่ได้รับความนิยมของคาด้าไมน์นิ่ง ปัญหาทางด้านธุรกิจต่างๆ เช่น การจัดการความเสี่ยง และการจัดแบ่งกลุ่มเป้าหมายของการโฆษณา เป็นต้น โดยปกติแล้วจะเกี่ยวข้องกับการจัดหมวดหมู่

2.2.2 การรวมกลุ่ม (Clustering)

การรวมกลุ่มจะถูกใช้ในการระบุกลุ่มโดยธรรมชาติของข้อมูลบนพื้นฐานของกลุ่มของแอตทริบิวต์ ข้อมูลที่อยู่ในกลุ่มเดียวกันก็มักจะมีค่าของแอตทริบิวต์ที่ใกล้เคียงกัน การรวมกลุ่มจะไม่อาศัยการแบ่งกลุ่มไว้ล่วงหน้า จะไม่มีแอตทริบิวต์ตัวใดตัวหนึ่งที่ใช้ในการกำหนดแนวทางในการบวนการเทรนแบบจำลอง ทุกๆแอตทริบิวต์ที่ถูกอินพุตเข้าไปจะมีความเท่าเทียมกัน

นอกจากนี้ การรวมกลุ่มยังใช้ในการค้นหาความสัมพันธ์ที่ซ่อนอยู่ระหว่างข้อมูล ซึ่งสามารถนำไปใช้ในการทำนายแนวโน้มในอนาคตได้ นอกจากนี้ การรวมกลุ่มยังใช้ในการค้นหาความสัมพันธ์ที่ซ่อนอยู่ระหว่างข้อมูล ซึ่งสามารถนำไปใช้ในการทำนายแนวโน้มในอนาคตได้

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 การค้นหาความสัมพันธ์ (Association)

Association เป็นอีกงานของดาต้าไมน์นิ่งที่ได้รับความนิยม มันเป็นการหาว่าสิ่งใดที่มักจะไปด้วยกันเสมอๆ เช่น สินค้าชนิดใดบ้างที่มักจะขายไปด้วยกันเสมอๆ

2.2.4 การประเมินค่า (Estimation)

การประเมินค่านั้นคล้ายกับการจัดหมวดหมู่ แต่การจัดหมวดหมู่นั้นจะให้ผลลัพธ์ในลักษณะที่ไม่ต่อเนื่อง ในขณะที่การประเมินค่าจะให้ผลลัพธ์ในลักษณะของตัวแปรที่ยังไม่ทราบค่าแต่มีความต่อเนื่อง เช่น การประมาณการของจำนวนลูกค้าในครอบครัว เป็นต้น

2.2.5 การทำนาย (Prediction)

การทำนายนั้นเหมือนกับการจัดหมวดหมู่และการประเมินค่า เว้นเสียแต่ว่าข้อมูลจะถูกจัดหมวดหมู่ให้สอดคล้องกับพฤติกรรมในอนาคตที่ทำนายไว้หรือค่าที่ประเมินไว้ในอนาคต

2.2.6 การหารายละเอียด (Description and Profiling)

บางครั้งวัตถุประสงค์ของดาต้าไมน์นิ่งก็เพื่ออธิบายว่ามีอะไรกำลังเกิดขึ้นในฐานข้อมูลที่ซับซ้อนในทางที่จะช่วยให้เราเข้าใจสิ่งที่ทำให้เกิดข้อมูลเหล่านั้น

2.3 กระบวนการทำดาต้าไมน์นิ่ง

กระบวนการในการทำดาต้าไมน์นิ่งนั้นสามารถแบ่งเป็นขั้นตอนต่างๆ ได้ดังนี้

2.3.1 การกำหนดเป้าหมายทางธุรกิจ (Business Objective)

โดยปกติแล้วสิ่งที่ต้องทำเป็นอันดับแรกในการทำดาต้าไมน์นิ่งก็คือการกำหนดขอบเขตและเป้าหมายในการทำดาต้าไมน์นิ่ง วิเคราะห์ถึงงานหรือเทคนิคที่ต้องใช้ในการทำดาต้าไมน์นิ่งเพื่อให้ได้ผลลัพธ์ที่สอดคล้องกับความต้องการทางธุรกิจ

2.3.2 การรวบรวมข้อมูล (Data Collection)

ในการรวบรวมข้อมูลจากแหล่งต่างๆ อันดับแรกก็ต้องทำการดึงข้อมูลออกมาไว้ที่ฐานข้อมูลหรือดาต้ามาร์ท (Data Mart) ที่จะใช้ในการทำดาต้าไมน์นิ่งเสียก่อน แต่บางครั้งก็โชคดีที่มีคลังข้อมูล (Data Warehouse) ที่สามารถนำมาใช้ในการทำดาต้าไมน์นิ่งอยู่แล้ว อย่างไรก็ตามถ้าข้อมูลในคลังข้อมูลไม่สมบูรณ์เพียงพอ ก็จำเป็นต้องรวบรวมข้อมูลจากแหล่งอื่นๆ มาเพิ่มเติม

2.3.3 การเตรียมข้อมูล (Data Preparation)

โดยทั่วไปแล้วขั้นตอนการเตรียมข้อมูลจะมีการทำความสะอาดและแปลงข้อมูล (Data Cleaning and Transformation) ซึ่งนับได้ว่าเป็นขั้นตอนที่เสียเวลาและใช้ทรัพยากรมากที่สุด

ในการทำค้ำไม้นิ่ง ขั้นตอนนี้ทำไปเพื่อกำจัดค่าที่รบกวนและไม่เกี่ยวข้องกับการทำค้ำไม้นิ่งออกจากชุดข้อมูล

2.3.4 การสร้างแบบจำลอง (Model Building)

หลังจากที่ได้ทำความสะอาดและแปลงข้อมูลเรียบร้อยแล้วเราก็สามารถเริ่มทำการสร้างแบบจำลอง การสร้างแบบจำลองค้ำไม้นิ่งนั้นถือได้ว่าเป็นหัวใจของการทำค้ำไม้นิ่ง ถึงแม้ว่ามันจะใช้เวลาไม่มากเท่ากับขั้นตอนการเตรียมข้อมูลก็ตาม เมื่อเราทราบถึงชนิดของงานของค้ำไม้นิ่งแล้ว มันก็ค่อนข้างง่ายในการที่จะเลือกอัลกอริทึมที่เหมาะสม ในบางกรณีที่เราไม่สามารถทราบได้ว่าอัลกอริทึมแบบไหนที่จะเหมาะสมกับข้อมูลที่สุด ทั้งนี้ก็เพราะว่าความแม่นยำของอัลกอริทึมก็ขึ้นอยู่กับธรรมชาติของข้อมูลด้วย วิธีการที่ถูกต้องก็คือทำการสร้างแบบจำลองจากหลายๆ อัลกอริทึมแล้วทำการเปรียบเทียบความถูกต้องของแต่ละแบบจำลอง นอกจากนี้ถึงแม้ว่าจะใช้เพียงอัลกอริทึมเดียวก็อาจมีความจำเป็นที่จะต้องมีการสร้างแบบจำลองหลายๆ อัน โดยการปรับแต่งพารามิเตอร์ให้แตกต่างกันเพื่อเพิ่มความถูกต้องแม่นยำของแบบจำลอง

2.3.5 การประเมินผลแบบจำลอง (Model Assessment)

ในขั้นตอนการสร้างแบบจำลองเราอาจจะได้ทำการสร้างแบบจำลองหลายๆ อันซึ่งมีอัลกอริทึมหรือการปรับแต่งพารามิเตอร์ที่แตกต่างกัน ในขั้นตอนการประเมินผลแบบจำลองก็ต้องทำการประเมินความถูกต้องแม่นยำของแต่ละแบบจำลองโดยใช้เครื่องมือต่างๆ ซึ่งมันจะใช้แบบจำลองที่ผ่านการเทรนแล้วเพื่อทำนายค่าของชุดข้อมูลที่ใช้ทดสอบ นอกจากนี้แล้วในขั้นตอนการประเมินผลเรายังจำเป็นต้องร่วมกันอภิปรายถึงความหมายของแบบอย่างที่เราค้นพบว่ามันสื่อความหมายว่าอย่างไร มีประโยชน์ต่อธุรกิจหรือไม่ อย่างไร เพราะบางครั้งแบบอย่างที่เราค้นพบก็ไม่ได้มีประโยชน์

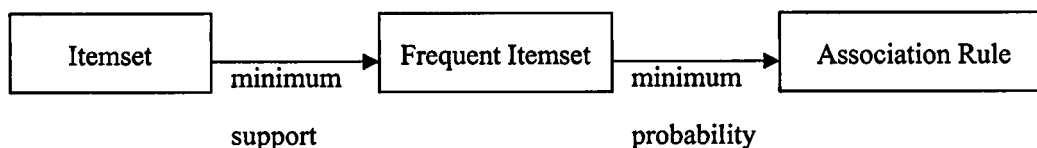
2.4 การค้นหากฎความสัมพันธ์

การค้นหากฎความสัมพันธ์ (Association rule discovery) เป็นการศึกษาเอตทริบิวต์หรือคุณลักษณะที่ “ ไปด้วยกัน ” โดยเป็นวิธีการค้นหากฎของความสัมพันธ์ระหว่างเอตทริบิวต์ 2 เอตทริบิวต์หรือมากกว่า กฎความสัมพันธ์จะปรากฏในรูปของกฎ “ ถ้า ... แล้ว ... ” (If ... Then ...) พร้อมกับค่าสนับสนุน (support) และค่าความเชื่อมั่น (confidence) ของกฎนั้นๆ ยกตัวอย่างเช่นห้างสรรพสินค้าแห่งหนึ่งพบว่า จากจำนวนลูกค้า 1000 คน ในคืนวันพฤหัสบดี พบว่ามีลูกค้าจำนวน 200 คนที่ซื้อผ้าอ้อม และใน 200 คนนั้น มี 50 คนที่ซื้อเบียร์ด้วย ดังนั้นกฎความสัมพันธ์ที่ได้ก็คือ “ถ้าซื้อผ้าอ้อมแล้วซื้อเบียร์” โดยมีค่าสนับสนุนเป็น $50/1000=5\%$ และค่าความเชื่อมั่นเป็น $50/200=25\%$

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ไม่สามารถเผยแพร่หรือใช้โดยไม่ได้รับอนุญาตจากมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1. **ไอเทมเซต (Itemset)** ไอเทมเซตคือ เซตของไอเทม(Item) โดยแต่ละไอเทมก็คือหนึ่งค่าของแอทริบิวต์ ยกตัวอย่างเช่น ในการวิเคราะห์แบบอย่างการซื้อสินค้า ในหนึ่งไอเทมเซตประกอบไปด้วยเซตของสินค้า เช่น เค้ก, คุกกี้ และนม โดยในแต่ละไอเทมเซตจะมีขนาด (size) ซึ่งเท่ากับจำนวนของไอเทมที่อยู่ในไอเทมเซต นั้นๆ เช่น ขนาดของไอเทมเซต {เค้ก, คุกกี้, นม} คือ 3
2. **ฟรีควেন্টไอเทมเซต (Frequent Itemset)** ก็คือไอเทมเซตที่มีโอกาสพบได้บ่อยๆ โดยสามารถวัดโอกาสในการพบโดยใช้ค่าสนับสนุน ซึ่งใช้ในการวัดจำนวนประชากรของไอเทมเซต
3. **ค่าสนับสนุน (Support)** ค่าสนับสนุนของไอเทมเซต {A,B} เกิดจากการรวมกันของจำนวนทรานแซกชันทั้งหมดที่ประกอบด้วยทั้ง A และ B ค่าสนับสนุนน้อยที่สุด (Minimum_Support) เป็นค่าที่จะต้องกำหนดก่อนการประมวลผลแบบจำลอง มันหมายถึงจำนวนขั้นต่ำที่สุดของค่าสนับสนุนที่มีไอเทมเซตก่อนที่จะมีการสร้างกฎขึ้นมา ในอัลกอริทึม Microsoft Association Rules นั้น ถ้ากำหนดค่าสนับสนุนน้อยที่สุดเป็นค่าน้อยกว่า 1 จะถือว่าหน่วยเป็นเปอร์เซ็นต์ เช่น ค่าสนับสนุนน้อยที่สุด = 0.01 หมายถึง 1% แต่ถ้ากำหนดค่าเป็นจำนวนเต็มมันจะหมายถึงค่าสนับสนุนน้อยที่สุดเท่ากับจำนวนนั้นๆ โดยค่าสนับสนุนน้อยที่สุดจะถูกใช้ในการจำกัดจำนวนไอเทมเซต แต่ไม่ได้ใช้ในการจำกัดกฎที่จะถูกสร้างขึ้น
4. **ค่าความน่าจะเป็น (Probability) หรือค่าความเชื่อมั่น (Confidence)** ค่าความน่าจะเป็นหรือเรียกว่าค่าความเชื่อมั่นเป็นคุณสมบัติของกฎความสัมพันธ์ ความน่าจะเป็นของกฎ $A \Rightarrow B$ ถูกคำนวณโดยใช้ค่าสนับสนุนของไอเทมเซต {A, B} หารด้วยค่าค่าสนับสนุนของ {A} โดยมีค่าความเป็นไปได้น้อยที่สุด(Minimum_Probability) เป็นค่าที่ต้องกำหนดก่อนการประมวลผลแบบจำลอง มันหมายความว่าเราสนใจเฉพาะกฎที่มีค่าความน่าจะเป็นสูงกว่าค่าความเป็นไปได้น้อยที่สุด โดยค่าความเป็นไปได้น้อยที่สุดจะส่งผลกระทบต่อกฎแต่ไม่ส่งผลกระทบต่อไอเทมเซต
5. **ค่าความสำคัญ (Importance) หรือเรียกว่าคะแนนความน่าสนใจ(Interesting Score)** โดยค่าความสำคัญสามารถใช้วัดได้ทั้งไอเทมเซตและกฎ ถ้าค่าความสำคัญ ({A, B}) = 1 แสดงว่า A และ B เป็นอิสระต่อกัน ถ้าค่าความสำคัญ ({A, B}) < 1 เป็นความสัมพันธ์เชิงลบหมายความว่าถ้ามี A ก็ไม่น่าจะมี B แต่ถ้าค่าความสำคัญ ({A, B}) > 1 แสดงว่าเป็นความสัมพันธ์เชิงบวก หมายความว่าถ้ามี A ก็น่าจะมี B ด้วย สำหรับในกรณีของกฎ ถ้าค่าความสำคัญ $(A \Rightarrow B) = 0$ แสดงว่าไม่มีความสัมพันธ์ระหว่าง A และ B ถ้าค่าเป็นบวกหมายความว่าค่าความน่าจะเป็นของ B จะสูงขึ้นเมื่อ A เป็นจริง แต่ถ้าค่าเป็นลบหมายความว่าค่าความน่าจะเป็นของ B จะมีค่าต่ำลงเมื่อ A เป็นจริง

การค้นหากฎความสัมพันธ์มีกระบวนการทำงาน 2 ขั้นตอน คือ หา Frequent Itemset ทั้งหมด แล้วสร้างกฎความสัมพันธ์จาก Frequent Itemset ดังรูปที่ 2.2



รูปที่ 2.2 ขั้นตอนการค้นหากฎความสัมพันธ์

2.5 อัลกอริทึม Apriori

อัลกอริทึม Apriori เป็นอัลกอริทึมดั้งเดิมที่นิยมใช้ในการค้นหากฎความสัมพันธ์ ถึงแม้ว่าจะมีอัลกอริทึมใหม่ๆ แต่ส่วนใหญ่แล้วอัลกอริทึมเหล่านี้ก็จะมีพื้นฐานมาจากอัลกอริทึม Apriori ซึ่งรวมถึงอัลกอริทึม Microsoft Association Rules ก็จัดอยู่ในตระกูลอัลกอริทึม Apriori ด้วย อัลกอริทึม Apriori มีกระบวนการทำงาน 2 ขั้นตอน คือ ขั้นตอนการค้นหา Frequent Itemset ทั้งหมด และขั้นตอนการสร้างกฎความสัมพันธ์จาก Frequent Itemset ที่ค้นหาได้

2.5.1 การค้นหา Frequent Itemset

อัลกอริทึม Apriori มีขั้นตอนการทำงานที่ซ้ำๆ กันเรื่อยๆ กล่าวคือ Frequent k-Itemsets จะถูกใช้ในการหา Frequent (k+1)-Itemsets ในอันดับแรกจะทำการค้นหา Frequent 1-Itemsets โดยการแสกนฐานข้อมูลเพื่อนับจำนวนของแต่ละไอเทม และรวบรวมไอเทมที่สอดคล้องตามค่าสนับสนุนน้อยที่สุด (ในที่นี้ใช้ L_1 เป็นสัญลักษณ์แทน Frequent 1-Itemset และ L_k เป็นสัญลักษณ์แทน Frequent k-Itemsets) ผลลัพธ์ที่ได้แทนด้วย L_1 ลำดับถัดมา L_1 ก็จะถูกใช้เพื่อหา L_2 และเซตของ Frequent 2-Itemsets ซึ่งก็จะถูกใช้เพื่อหา L_3 และทำซ้ำลักษณะนี้ต่อไปเรื่อยๆ จนกว่าจะไม่สามารถหา Frequent k-Itemsets ได้อีกต่อไป และเพื่อเพิ่มประสิทธิภาพในการสร้าง Frequent Itemsets คุณสมบัติที่สำคัญของ Apriori ที่เรียกว่า Apriori property ได้ถูกใช้เพื่อลดพื้นที่ในการสืบค้นข้อมูล คุณสมบัติดังกล่าวได้แสดงไว้ที่ด้านล่าง

Apriori property : ทุกๆ เซตที่ไม่ว่างของ Frequent Itemset จะต้อง Frequent ด้วย

จากคำจำกัดความ ถ้าไอเทมเซต I มีจำนวนน้อยกว่าค่าสนับสนุนน้อยที่สุด \min_sup แล้ว I จะไม่ Frequent นั่นคือความน่าจะเป็น $P(I) < \min_sup$. ถ้าไอเทม A ถูกเพิ่มเข้าไปในไอเทมเซต I แล้วไอเทมเซตที่เป็นผลลัพธ์ (หรือ $I \cup A$) ไม่สามารถเกิดบ่อยได้มากกว่า I ดังนั้น $I \cup A$ ก็จะไม่ Frequent ด้วย นั่นก็คือ $P(I \cup A) < \min_sup$

ขั้นตอนการทำงานที่ L_{k-1} ถูกใช้เพื่อหา L_k เมื่อ $k \geq 2$ มี 2 ขั้นตอนคือ Join และ Prune

1. ขั้นตอนการเชื่อมความสัมพันธ์ระหว่าง Itemset (Join step) เป็นขั้นตอนการสร้างแคนดิเดตไอเทมเซต (Candidate Itemsets) หรือเป็น Itemset ตัวเลือกที่จะถูกเลือกไปเป็น Frequent Itemset โดย Candidate 1-Itemset แทนด้วยสัญลักษณ์ C_1 และ Candidate 2-Itemsets แทนด้วยสัญลักษณ์ C_2 ไปจนถึง Candidate k-Itemsets แทนด้วยสัญลักษณ์ C_k โดย C_k จะเกิดจากการเชื่อมความสัมพันธ์ของ L_{k-1} และ L_{k-1} วิธีการเชื่อมความสัมพันธ์ทำโดยการพิจารณาว่าจะเชื่อมความสัมพันธ์จากจำนวน Itemsets เท่าไหร่ ไปเป็นเท่าไหร่ โดยกำหนดให้หา C_k ซึ่งจะเกิดจากการเชื่อมความสัมพันธ์ระหว่าง L_{k-1} กับตัวมันเอง ในการที่จะพิจารณาว่า Itemset ใดสามารถเชื่อมความสัมพันธ์กันได้ ให้พิจารณาที่ Item ที่ 1 ถึง k-2 ของทั้งสอง Itemset หากเหมือนกันก็สามารถเชื่อมความสัมพันธ์กันได้ เช่น $\{I1, I2\}$, $\{1, I3\}$, $\{I2, I3\}$ Itemset ที่ 1 และ 2 สามารถเชื่อมความสัมพันธ์กันได้ เพราะลำดับแรกของไอเทมเหมือนกัน (I1) เมื่อเชื่อมความสัมพันธ์จะได้ $\{I1, I2, I3\}$ และในการเชื่อมความสัมพันธ์ต่อไป Itemset ที่ 1 และ 3 ไม่สามารถเชื่อมความสัมพันธ์ได้ เพราะลำดับแรกของ Item ทั้งสองไม่เหมือนกัน

2. ขั้นตอนการคัด Itemset ทิ้ง (Prune step) เป็นขั้นตอนการคัดสมาชิกใน C_k โดยที่ C_k นั้นเป็น Superset ของ L_k นั่นคือสมาชิกของ C_k อาจจะเป็นหรือไม่เป็น Frequent ก็ได้ แต่ Frequent k-itemsets ทั้งหมดนั้นจะอยู่ใน C_k ทำให้ C_k อาจมีขนาดใหญ่เกินไปและต้องใช้เวลาประมวลผลอย่างมาก ด้วยเหตุนี้เพื่อลดขนาดของ C_k คุณสมบัติของอัลกอริทึม Apriori จึงถูกนำมาใช้งานดังนี้ (k-1)-Itemset ใดๆ ที่ไม่ Frequent จะไม่สามารถเป็นซัพเซตของ Frequent k-Itemset ซึ่งจะได้ว่า ถ้า (k-1)-subset ใดๆ ของ Candidate k-Itemset ไม่ได้อยู่ใน L_{k-1} ดังนั้น Candidate จะไม่สามารถที่จะเป็น Frequent ด้วย จึงสามารถเอาออกจาก C_k ได้

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

รูปที่ 2.3 ฐานข้อมูล D

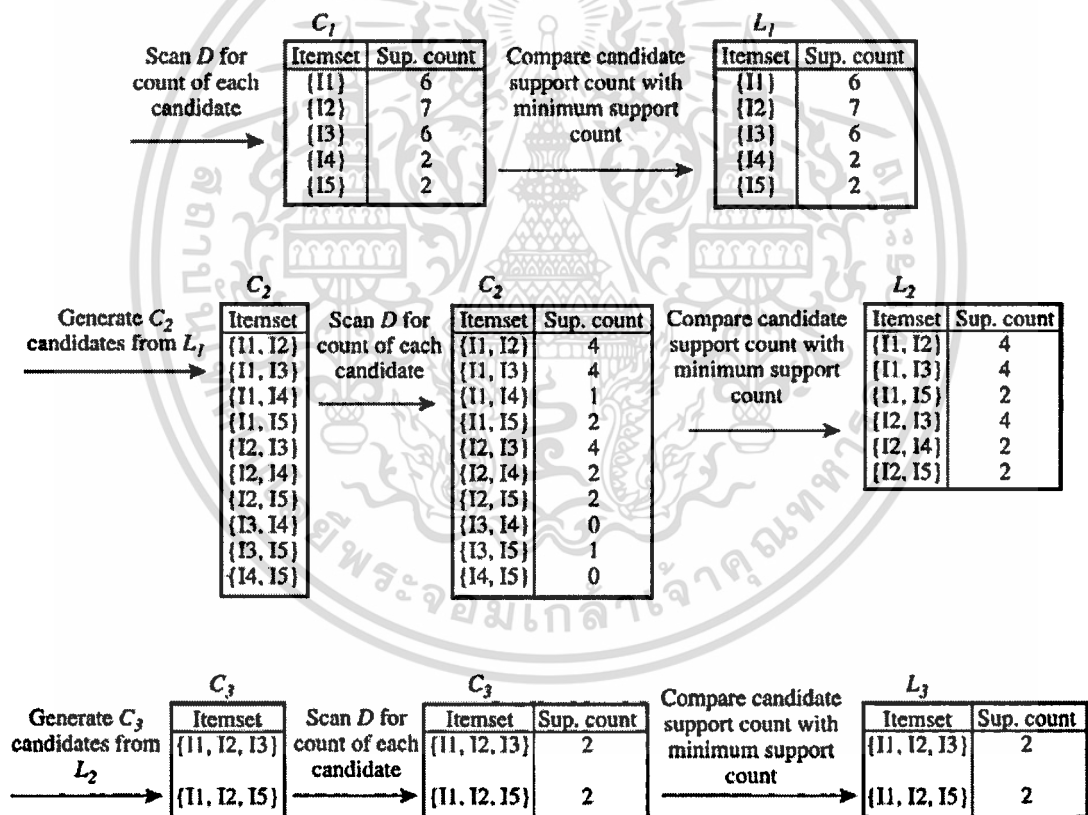
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การหา Frequent Itemset จะกระทำวนซ้ำตามขั้นตอนที่ 1 และ 2 ไปเรื่อย ๆ จนกว่าจะไม่สามารถหา L_k ได้อีก จึงหยุดการหา Frequent Itemset วิธีการโดยละเอียดของการหา Frequent Itemset สามารถแสดงได้ดังตัวอย่างต่อไปนี้

กำหนดให้ฐานข้อมูล D มีจำนวนทรานแซกชัน (Transaction) 9 ทรานแซกชันแสดงดังรูปที่ 2.3 และค่าสนับสนุนน้อยที่สุดมีค่าเท่ากับ 2 และค่าความเชื่อมั่นน้อยที่สุดมีค่า 70 %

จากฐานข้อมูล D ที่มีอยู่สามารถนำรายการที่มีอยู่มาหากฎความสัมพันธ์ของข้อมูลแต่ละรายการ โดยการหา Frequent Itemset แสดงดังรูปที่ 2.4

จากรูปที่ 2.4 แสดงการหา Frequent Itemset จากฐานข้อมูล D เริ่มต้นจากการหา C_1 โดยการสแกนหา 1-Itemset จากฐานข้อมูล D และนับความถี่ของแต่ละ Itemset ในฐานข้อมูลขั้นตอนตามรูปที่ 2.4 สามารถอธิบายได้ดังนี้



รูปที่ 2.4 วิธีการหา Frequent Itemset

1. ขั้นตอนที่ 1 ของอัลกอริทึม แต่ละไอเทมเป็นสมาชิกของเซตของ Candidate 1-Itemsets, C_1 อัลกอริทึมจะทำการสแกนทรานแซกชันทั้งหมดเพื่อนับจำนวนการปรากฏของแต่ละไอเทม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. จาก C_1 เป็น L_1 สามารถหา L_1 ได้จาก Itemset จาก C_1 ที่ผ่านค่าสนับสนุนน้อยที่สุดแล้ว จากตัวอย่างค่าสนับสนุนแต่ละตัวใน Itemset ของ C_1 มีค่ามากกว่าหรือเท่ากับค่าสนับสนุนน้อยที่สุด(ในที่นี้ค่าสนับสนุนน้อยที่สุดมีค่าเป็น 2) ดังนั้นสมาชิกของ C_1 ทุกตัวจึงเป็นสมาชิกของ L_1
3. ขั้นตอนที่ 3 จาก L_1 เป็น C_2 เป็นการหา Candidate 2-Itemsets ซึ่งเกิดจากการเชื่อมความสัมพันธ์ของ L_1 และ L_1 จะได้เป็น C_2 เมื่อได้ Itemset ที่เกิดจากการเชื่อมความสัมพันธ์แล้ว ต้องคัดสมาชิกที่ได้ด้วยการหาเซตย่อยของแต่ละ Itemset ว่ามีอยู่ใน L_1 หรือไม่ เช่น $\{I1, I2\}$ มีเซตย่อยคือ $\{I1\}$, $\{I2\}$ เมื่อดู Item ใน L_1 แล้วมีทั้ง 2 Itemset ดังนั้นจึงเก็บ $\{I1, I2\}$ นี้ไว้ และพิจารณาอย่างนี้ทุก Itemset
4. ขั้นตอนที่ 4 จะเป็นการแสดงทราณแซกชันใน D เพื่อหาความถี่หรือค่าสนับสนุนของแต่ละ Itemset ใน C_2 ที่ปรากฏในฐานข้อมูล ดังที่ได้แสดงไว้ในตารางตรงกลางของแถวที่สองในรูปที่ 2.4
5. ขั้นตอนที่ 5 จาก C_2 เป็น L_2 โดย L_2 เป็น Itemset จาก C_2 ที่ผ่านค่าสนับสนุนน้อยที่สุดแล้ว จากรูปที่ 2.4 พบว่ามีทั้งหมด 4 Itemset ที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนน้อยที่สุด ได้แก่ $\{I1, I4\}$, $\{I3, I4\}$ $\{I3, I5\}$ และ $\{I4, I5\}$ ดังนั้นจึงตัด Itemset เหล่านี้ทิ้ง
6. ขั้นตอนที่ 6 จาก L_2 เป็น C_3 เป็นการหา Candidate 3-Itemsets ซึ่งเกิดจากการเชื่อมความสัมพันธ์ของ L_2 และ L_2 จะได้เป็น C_3 เมื่อได้ Itemset แล้วต้องคัดสมาชิกที่ได้ด้วยการหาเซตย่อยของแต่ละ Itemset ว่ามีอยู่ใน L_2 หรือไม่ จากรูปที่ 2.4 เมื่อเชื่อมความสัมพันธ์ของ L_2 และ L_2 จะได้ Itemset จำนวน 6 Itemset ดังนี้ $\{I1, I2, I3\}$, $\{I1, I2, I5\}$, $\{I1, I3, I5\}$, $\{I2, I3, I4\}$, $\{I2, I3, I5\}$, $\{I2, I4, I5\}$ และนำแต่ละ Itemset มาหาเซตย่อยแล้วตรวจดูว่าเซตย่อยที่ได้ปรากฏอยู่ใน L_2 หรือไม่ เช่น $\{I1, I2, I3\}$ หาเซตย่อยได้ $\{I1, I2\}$, $\{I1, I3\}$ และ $\{I2, I3\}$ เมื่อพบว่าเซตย่อยทั้งหมดปรากฏอยู่ใน L_2 จึงเก็บ Itemset นี้ไว้ และในกรณีที่ต้องตัด Itemset นั้น ๆ ทิ้ง เช่น $\{I1, I3, I5\}$ หาเซตย่อยได้ $\{I1, I3\}$, $\{I1, I5\}$ และ $\{I3, I5\}$ เมื่อตรวจดูใน L_2 แล้วพบว่าไม่มี $\{I3, I5\}$ ปรากฏอยู่ดังนั้นจึงตัด $\{I1, I3, I5\}$ ออก ทำอย่างนี้ไปเรื่อย ๆ จนกว่าจะหมดทุก Itemset จะพบว่ามีจำนวน 2 Itemset เท่านั้นที่มีคุณสมบัติครบคือ $\{I1, I2, I3\}$ และ $\{I1, I2, I5\}$
7. ขั้นตอนที่ 7 จะเป็นการแสดงทราณแซกชันใน D เพื่อหา L_3 จากรูปที่ 2.4 พบว่าทุก Itemset มีค่าสนับสนุนเท่ากับค่าสนับสนุนน้อยที่สุดจึงเก็บ Itemset เหล่านี้ไว้

8. จาก L_3 นำมาเชื่อมความสัมพันธ์เพื่อจะได้ C_4 จะได้ $\{I1, I2, I3, I5\}$ และหาเซตย่อยของ Itemset นี้จะได้ $\{I1, I2, I3\}$, $\{I1, I2, I5\}$, $\{I1, I3, I5\}$ และ $\{I2, I3, I5\}$ เมื่อดูเซตย่อยนี้ใน L_3 ปรากฏว่าใน L_3 ไม่มี Itemset ที่เป็นเซตย่อยทั้งหมด ดังนั้นจึงไม่มี C_4 เกิดขึ้น อัลกอริทึมจึงหยุดการหา Frequent Itemset

Algorithm: Apriori. Find frequent itemsets using an iterative level-wise approach based on candidate generation.

Input:

- ▣ D , a database of transactions;
- ▣ min_sup , the minimum support count threshold.

Output: L , frequent itemsets in D .

Method:

```

(1)  $L_1 = \text{find\_frequent\_1-itemsets}(D)$ ;
(2) for ( $k = 2; L_{k-1} \neq \emptyset; k++$ ) {
(3)    $C_k = \text{apriori\_gen}(L_{k-1})$ ;
(4)   for each transaction  $t \in D$  { // scan  $D$  for counts
(5)      $C_t = \text{subset}(C_k, t)$ ; // get the subsets of  $t$  that are candidates
(6)     for each candidate  $c \in C_t$ 
(7)        $c.\text{count}++$ ;
(8)   }
(9)    $L_k = \{c \in C_k | c.\text{count} \geq min\_sup\}$ 
(10) }
(11) return  $L = \cup_k L_k$ ;

procedure apriori_gen( $L_{k-1}$ : frequent ( $k-1$ )-itemsets)
(1) for each itemset  $l_1 \in L_{k-1}$ 
(2)   for each itemset  $l_2 \in L_{k-1}$ 
(3)     if ( $l_1[1] = l_2[1] \wedge l_1[2] = l_2[2] \wedge \dots \wedge l_1[k-2] = l_2[k-2] \wedge l_1[k-1] < l_2[k-1]$ ) then {
(4)        $c = l_1 \bowtie l_2$ ; // join step: generate candidates
(5)       if has_infrequent_subset( $c, L_{k-1}$ ) then
(6)         delete  $c$ ; // prune step: remove unfruitful candidate
(7)       else add  $c$  to  $C_k$ ;
(8)     }
(9) return  $C_k$ ;

procedure has_infrequent_subset( $c$ : candidate  $k$ -itemset;
                                 $L_{k-1}$ : frequent ( $k-1$ )-itemsets); // use prior knowledge
(1) for each ( $k-1$ )-subset  $s$  of  $c$ 
(2)   if  $s \notin L_{k-1}$  then
(3)     return TRUE;
(4) return FALSE;

```

รูปที่ 2.5 อัลกอริทึม Apriori

2.5.2 การสร้างกฎความสัมพันธ์

การสร้างกฎความสัมพันธ์จาก Frequent Itemset จะนำ L_2 ถึง L_k มาสร้างกฎความสัมพันธ์ และในตัวอย่างจะนำ L_2 และ L_3 มาสร้างกฎความสัมพันธ์

วิธีการสร้างกฎความสัมพันธ์ ทำโดยนำ L_2 ถึง L_k (ในตัวอย่างคือ L_3) มาหาเซตย่อยแต่ละ Itemset และนำเซตย่อยที่ได้มาสร้างเป็นกฎความสัมพันธ์ จาก Frequent Itemset L ทำการหาเซตย่อยของ Frequent Itemset L และทุกเซตย่อย s ของ L ยกเว้นเซตว่างจะได้กฎ “ $s \Rightarrow (L-s)$ ”

จาก L_2 ที่ Itemset แรก $\{I1, I2\}$ นำมาหาเซตย่อยได้ $\{I1\}$, $\{I2\}$ และ $\{I1, I2\}$ เมื่อนำมาสร้างกฎจะได้ $I1 \Rightarrow I2$, $I2 \Rightarrow I1$ แต่ไม่สามารถสร้าง $I1 \Rightarrow I1 \wedge I2$ ได้ เพราะทางด้านซ้ายขวาจะมี $I1$ เหมือนกันไม่ได้ และสำหรับ Itemset อื่น ๆ ก็สร้างกฎด้วยวิธีนี้เช่นกัน ในทำนองเดียวกันนำ L_3 มาสร้างกฎความสัมพันธ์ ที่ Itemset หาเซตย่อยได้ $\{I1\}$, $\{I2\}$, $\{I3\}$, $\{I1, I2\}$ $\{I1, I3\}$ $\{I2, I3\}$ และ $\{I1, I2, I3\}$ เมื่อนำมาสร้างกฎความสัมพันธ์จะได้ $I1 \Rightarrow I2 \wedge I3$, $I2 \Rightarrow I1 \wedge I3$, $I3 \Rightarrow I1 \wedge I2$, $I1 \wedge I2 \Rightarrow I3$, $I1 \wedge I3 \Rightarrow I2$ และ $I2 \wedge I3 \Rightarrow I1$ เมื่อสร้างกฎความสัมพันธ์จาก Frequent Itemset แล้ว กฎความสัมพันธ์ที่ได้ยังไม่สามารถนำไปใช้งานได้ ต้องนำไปหาค่าความเชื่อมั่นก่อน กฎที่สามารถนำไปใช้งานได้ต้องผ่านค่าความเชื่อมั่นน้อยที่สุดก่อน จากตัวอย่างได้กำหนดให้ค่าความเชื่อมั่นน้อยที่สุดมีค่า 70% กฎความสัมพันธ์ที่สามารถนำไปใช้งานได้ต้องมีค่าความเชื่อมั่นของกฎมากกว่าหรือเท่ากับ 70% จากตัวอย่างจะได้กฎความสัมพันธ์ที่นำไปใช้งานได้ 6 กฎ คือ $I5 \Rightarrow I1$, $I4 \Rightarrow I2$, $I5 \Rightarrow I2$, $I5 \Rightarrow I1 \wedge I2$, $I1 \wedge I5$ และ $I2 \wedge I5 \Rightarrow I1$

2.6 SQL Server 2005

SQL Server 2005 เป็นระบบจัดการฐานข้อมูลซึ่งได้รับการออกแบบให้เป็นโซลูชันในการจัดการและวิเคราะห์ข้อมูลของบริษัทไมโครซอฟต์ ที่ได้รับการเพิ่มความปลอดภัย, ความสามารถในการขยายระบบ และความต่อเนื่องในการให้บริการ ให้กับการจัดเก็บข้อมูลและแอปพลิเคชันวิเคราะห์ข้อมูลขององค์กร ขณะเดียวกันก็ง่ายต่อการพัฒนา, นำไปใช้งานและบริหารจัดการมากยิ่งขึ้น SQL Server 2005 นั้นนอกจากจะมีบริการพื้นฐานที่จำเป็นสำหรับระบบจัดการฐานข้อมูลแล้ว มันยังมีบริการในลักษณะ Business Intelligence ซึ่งจะช่วยให้องค์กรยุคใหม่สามารถนำมาใช้งานทั้งในด้านการดำเนินงานและการสนับสนุนการตัดสินใจของผู้บริหาร เพื่อให้สามารถดำเนินธุรกิจได้อย่างมีประสิทธิภาพภายใต้สภาพแวดล้อมที่มีการแข่งขันสูงอย่างเช่นในปัจจุบัน

2.6.1 SQL Server Business Intelligence

บริการทางด้าน Business Intelligence ของ SQL Server 2005 ประกอบไปด้วยบริการต่างๆ ดังต่อไปนี้

- SQL Server Integration Services (SSIS) ซึ่งเคยเป็นที่รู้จักกันในชื่อว่า Data Transformation Services (DTS) ใช้ในการรวบรวมข้อมูลจากแหล่งต่างๆ, แปลงข้อมูล และนำข้อมูลเข้าสู่ระบบฐานข้อมูล
- SQL Server Reporting Services (SSRS) ใช้ในการสร้างรายงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- SQL Server Analysis Services (SSAL) ใช้ในการทำ OLAP และค้ำ้าไมน์นึ่ง

นอกจากนี้ SQL Server 2005 ยังมีเครื่องมือช่วยในการพัฒนาที่ทำงานในแบบกราฟิก ซึ่งจะช่วยให้การพัฒนาแอปพลิเคชันในลักษณะ Business Intelligence นั้นสามารถทำได้ง่ายขึ้น นั่นก็คือ Microsoft Business Intelligence Development Studio

2.6.2 ค้ำ้าไมน์นึ่งของ SQL Server (SQL Server Data Mining)

SQL Server 2005 มีเทคนิคในการทำค้ำ้าไมน์นึ่งที่หลากหลาย โดยมีอัลกอริทึมที่สนับสนุน ดังนี้

- Microsoft Decision Trees Algorithm
- Microsoft Clustering Algorithm
- Microsoft Naive Bayes Algorithm
- Microsoft Association Algorithm
- Microsoft Sequence Clustering Algorithm
- Microsoft Time Series Algorithm
- Microsoft Neural Network Algorithm
- Microsoft Logistic Regression Algorithm
- Microsoft Linear Regression Algorithm

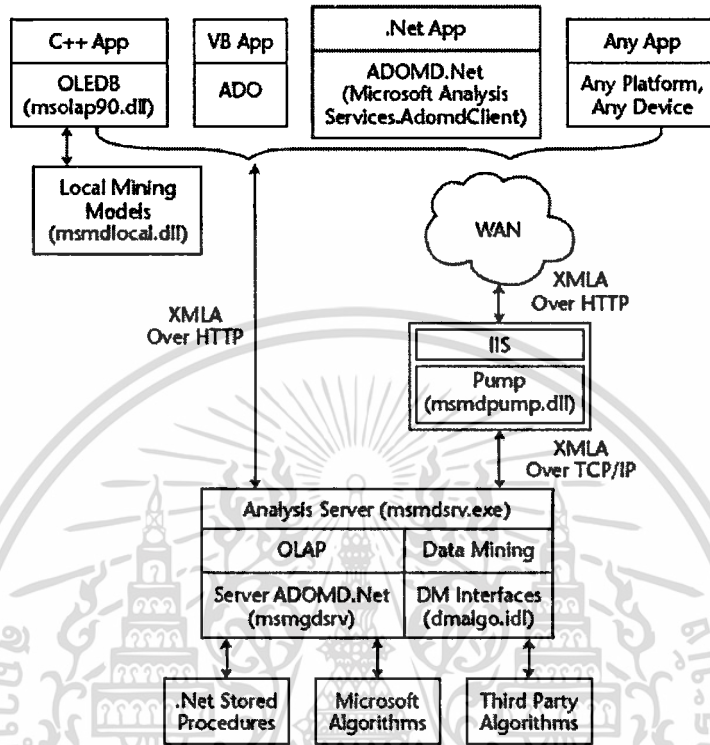
Analysis Services ของ SQL Server 2005 มีสถาปัตยกรรมแบบโคลเอนต์-เซิร์ฟเวอร์ โดยในส่วนของโคลเอนต์นั้นทำงานได้ทั้งผ่าน OLE DB และ ADOMD.Net และมีการสื่อสารกับ Analysis Services โดยใช้ XML for Analysis Services (XMLA) สถาปัตยกรรมค้ำ้าไมน์นึ่งของ SQL Server 2005 นั้นสามารถแสดงได้ดังรูปที่ 2.6

2.7 OLE DB สำหรับค้ำ้าไมน์นึ่ง (OLE DB for Data Mining)

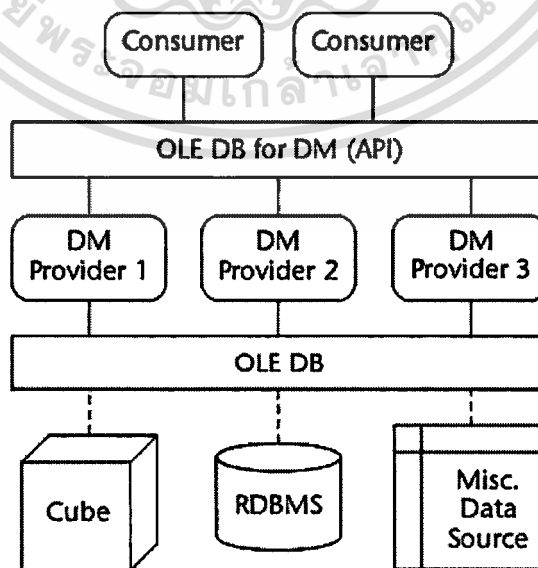
การอินทิเกรตผลิตภัณฑ์ค้ำ้าไมน์นึ่งต่างๆ เข้ากับแอปพลิเคชันของผู้ใช้งานนั้นทำได้ยาก เนื่องจากไม่มีภาษามาตรฐานของค้ำ้าไมน์นึ่ง ผลิตภัณฑ์จากแต่ละผู้ผลิตก็มีรูปแบบของตนเอง ไม่สามารถทำงานติดต่อกับผลิตภัณฑ์จากผู้ผลิตอื่นได้ การพัฒนาแอปพลิเคชันทางด้านค้ำ้าไมน์นึ่งจึงผูกติดอยู่กับแต่ละผู้ผลิต

OLE DB สำหรับค้ำ้าไมน์นึ่ง (OLE DB for Data Mining) เป็นความพยายามของไมโครซอฟต์ที่จะสร้างภาษามาตรฐานของค้ำ้าไมน์นึ่ง โดยมันเป็นส่วนเพิ่มขยายของ OLE DB และสนับสนุนการทำค้ำ้าไมน์นึ่งบน OLE DB Data Providers ต่างๆ โดยมีเป้าหมายเพื่อให้อัลกอริทึมทางด้านค้ำ้าไมน์นึ่งจากบริษัทซอฟต์แวร์ต่างๆ สามารถนำไปใช้ในโปรแกรมประยุกต์

ต่างๆ ได้ง่ายขึ้น OLE DB สำหรับค้าไมน์นิ่งนั้นมีต้นกำเนิดมาจาก OLE DB และ SQL โดยมีแกนหลักอยู่ที่ภาษา Data Mining Extensions (DMX) โดยมีสถาปัตยกรรมพื้นฐานดังรูปที่ 2.7



รูปที่ 2.6 สถาปัตยกรรมค้าไมน์นิ่งของ SQL Server 2005



รูปที่ 2.7 สถาปัตยกรรมพื้นฐานของ OLE DB for Data Mining

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่สามารถนำข้อมูลไปใช้ประโยชน์ด้านการค้าไมน์นิ่งโดยไม่ได้รับอนุญาตจากมหาวิทยาลัยฯ หากฝ่าฝืนโดยไม่แจ้งให้มหาวิทยาลัยฯ ทราบจะดำเนินการฟ้องดำเนินคดีตามกฎหมายที่เกี่ยวข้อง

หลักการพื้นฐานของ OLE DB สำหรับการค้าไมนิ่ง ก็คือ การค้าไมนิ่งเป็นการวิเคราะห์ เคส (Case) โดยเคสก็คือเอนติตีพื้นฐานของข้อมูล ประกอบไปด้วยกลุ่มของแอตทริบิวต์ โดยค่าที่เป็นไปได้ของแอตทริบิวต์เรียกว่า state บางเคสอาจจะเป็นแบบง่าย ๆ ยกตัวอย่างเช่น การใช้ข้อมูลของลูกค้ามาวิเคราะห์ความเสี่ยงในการให้เครดิต แต่ละเคสก็คือลูกค้าแต่ละราย ซึ่งก็คือแต่ละแถวของตารางลูกค้า หรือในบางเคสอาจจะมี ความซับซ้อนขึ้น ยกตัวอย่างเช่น การวิเคราะห์พฤติกรรม การซื้อของลูกค้า โดยใช้ข้อมูลของลูกค้ามาประกอบกับประวัติการซื้อสินค้า ในกรณีนี้แต่ละเคสก็คือลูกค้าแต่ละคนรวมถึงผลิตภัณฑ์ที่ลูกค้าผู้นั้นได้ซื้อไปด้วย เคสแบบนี้จะเรียกว่า nested case

Case key เป็นแอตทริบิวต์ที่ใช้ระบุแต่ละเคสซึ่งไม่ซ้ำกัน โดยปกติแล้วก็คือไพรมารีคีย์ของตารางนั่นเอง บ่อยครั้งที่อัลกอริทึมจะไม่นำ case key ไปใช้ในการวิเคราะห์หาแบบอย่างของข้อมูล

สำหรับ Nested key นั้น นับได้ว่าเป็นแอตทริบิวต์ที่มีความสำคัญมาก nested key ไม่ได้ถูกใช้เป็นตัวระบุแต่ละเคส แต่มันมีข้อมูลที่เป็นประโยชน์เกี่ยวกับแบบอย่างของข้อมูล ยกตัวอย่างเช่น ถ้าเราต้องการวิเคราะห์พฤติกรรม การซื้อของลูกค้า nested key ก็คือผลิตภัณฑ์ นั่นเอง

Case table เป็นตารางที่บรรจุข้อมูลของเคสซึ่งสัมพันธ์กับส่วนที่แบนราบของเคส สำหรับ nested table ก็คือตารางที่มีข้อมูลที่สัมพันธ์กับส่วนที่เป็นแบบ nested โดยปกติแล้ว nested table ก็คือตารางทรานแซกชัน ยกตัวอย่างเช่น ประวัติการซื้อสินค้า nested table สามารถรวมเข้ากับ case table โดยใช้ case key

คอลัมน์ในแบบจำลองไมนิ่งคล้ายกับคอลัมน์ในฐานข้อมูลเชิงสัมพันธ์ แบบจำลองสามารถมีคอลัมน์ได้ 4 ประเภทขึ้นอยู่กับการใช้งาน คือ key, input, predictable หรือเป็นทั้ง input และ predictable โดยทั่วไปแล้วเป้าหมายของแบบจำลองก็คือคอลัมน์ predictable

โครงสร้างของคอลัมน์นั้นเป็นได้ 2 แบบ คือ สเกลาร์และตาราง แต่ละคอลัมน์ที่เป็นแบบสเกลาร์นั้นจะมีค่าเพียงค่าเดียวสำหรับแต่ละเรคคอร์ด ยกตัวอย่างเช่น เพศ และอายุ สำหรับคอลัมน์ที่เป็นแบบตารางนั้นมันจะฝังตารางไว้ในคอลัมน์ ยกตัวอย่างดังรูปที่ 2.8 ตารางการซื้อสินค้า เป็นคอลัมน์แบบตาราง มันบรรจุตารางที่ประกอบไปด้วย 2 คอลัมน์ คือ ผลิตภัณฑ์ และ จำนวนที่ซื้อ

2.8 อัลกอริทึม Microsoft Association Rules

ดังที่ได้กล่าวไปแล้วในตอนต้นว่าอัลกอริทึม Microsoft Association Rules นั้นมีพื้นฐานการทำงานมาจากอัลกอริทึม Apriori โดยอัลกอริทึมส่วนใหญ่ นั้นมักจะหยุดการทำงานเมื่อหาไอเทมเซตและกฎความสัมพันธ์ได้สำเร็จ แต่สำหรับอัลกอริทึม Microsoft Association Rules นั้นมันยังสามารถทำนายผลจากกฎที่สร้างขึ้นได้ด้วย โดยปกติแล้วผลจากการทำนายจะเป็นเซตของไอเทม

ที่แนะนำเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Cust.ID	Gender	Income	Marital Status	Purchases	
				Product	Quantity
1	Male	23000	Single	Milk	1
				Cheese	1
				Beer	2
2	Female	79200	Married	Milk	8
				Pepsi	6
				Cake	1
3	Male	42000	Married	Cheese	2
				Juice	2

รูปที่ 2.8 ตารางเคสแบบ Nested

อย่างไรก็ตาม ไม่ใช่ทุกๆ ไอเทมเซตจะปรากฏในกฎ อัลกอริทึม Microsoft Association มีวิธีการในการทำนายผลดังนี้

1. กำหนดรายการของไอเทมมาให้ หากกฎทั้งหมดที่ด้านซ้ายของกฎสอดคล้องกับทั้งหมดหรือสับเซตของรายการ ไอเทม แล้วใช้กฎเหล่านั้นเพื่อทำนาย
2. ถ้าไม่มีกฎที่เหมาะสมหรือมีน้อยเกินไป ก็ใช้ค่าทางสถิติเพื่อเลือก n ไอเทมใดๆ ที่ได้รับความนิยมมากที่สุด
3. เรียงลำดับไอเทมจากขั้นตอนที่ 1 และ 2 ตามลำดับค่าความน่าจะเป็นจากมากไปหาน้อย

บทที่ 3

การวิเคราะห์และออกแบบระบบงาน

ในบทนี้จะกล่าวถึงขั้นตอนกระบวนการต่างๆ ในการวิเคราะห์และออกแบบระบบงานของโครงการประยุกต์ใช้การค้นหากฎความสัมพันธ์ร่วมกับระบบห้องสมุด โดยจะมีการกล่าวถึงลักษณะการทำงานของระบบงานปัจจุบัน, ปัญหาที่พบในระบบงานปัจจุบัน, การศึกษาความต้องการของระบบงานใหม่, ยุทธศาสตร์ของระบบงานใหม่, คลาสไดอะแกรมของระบบงานใหม่, ซีควเอนซ์ไดอะแกรมของระบบงานใหม่ และการออกแบบฐานข้อมูลของระบบงานใหม่

3.1 ลักษณะการทำงานของระบบงานปัจจุบัน

ปัจจุบันระบบห้องสมุดที่ใช้งานอยู่จะเป็นลักษณะการทำงานผ่านเว็บ โดยแบ่งระบบห้องสมุดออกเป็น 2 ส่วนหลักๆ คือ ระบบที่ให้บริการสำหรับบุคคลทั่วไป และระบบสำหรับเจ้าหน้าที่บรรณารักษ์ สำหรับระบบที่ให้บริการสำหรับบุคคลทั่วไปนั้นจะให้บริการในการสืบค้นข้อมูลสารสนเทศของห้องสมุดพร้อมทั้งแสดงรายละเอียดของทรัพยากรเหล่านั้นจากฐานข้อมูล นอกจากนี้แล้วยังให้บริการตรวจสอบประวัติการยืม-คืนของสมาชิกอีกด้วย ส่วนระบบสำหรับเจ้าหน้าที่บรรณารักษะนั้นจะให้บริการในส่วนของการยืม-คืน และการจัดการทรัพยากรต่างๆ ของห้องสมุด

3.2 ปัญหาที่พบในระบบงานปัจจุบัน

ระบบงานในปัจจุบันมีปัญหาหรือสิ่งที่จะต้องปรับปรุงให้ดีขึ้น ดังต่อไปนี้

- ในส่วนที่ให้บริการสำหรับบุคคลทั่วไปนั้นยังไม่มีส่วนของระบบที่ช่วยแนะนำข้อมูลให้กับผู้ใช้งานเกี่ยวกับทรัพยากรห้องสมุดที่น่าจะมีความเกี่ยวข้องหรือมีความสัมพันธ์กับทรัพยากรห้องสมุดที่ผู้ให้บริการกำลังสนใจอยู่ ซึ่งจะช่วยอำนวยความสะดวกในการสืบค้นข้อมูล และช่วยให้ผู้ให้บริการห้องสมุดสามารถให้บริการได้อย่างมีประสิทธิภาพ
- ระบบปัจจุบันยังไม่มีส่วนที่จะช่วยให้ผู้ให้บริการห้องสมุดสามารถแสดงความคิดเห็น พร้อมทั้งกำหนดระดับคะแนนความพึงพอใจ(Rating)ที่มีต่อทรัพยากรห้องสมุด ซึ่งจะช่วยให้ผู้ใช้งานคนอื่นๆ สามารถใช้ในการพิจารณาเลือกทรัพยากรห้องสมุดนั้นๆ ได้ด้วย

3.3 การศึกษาความต้องการของระบบงานใหม่

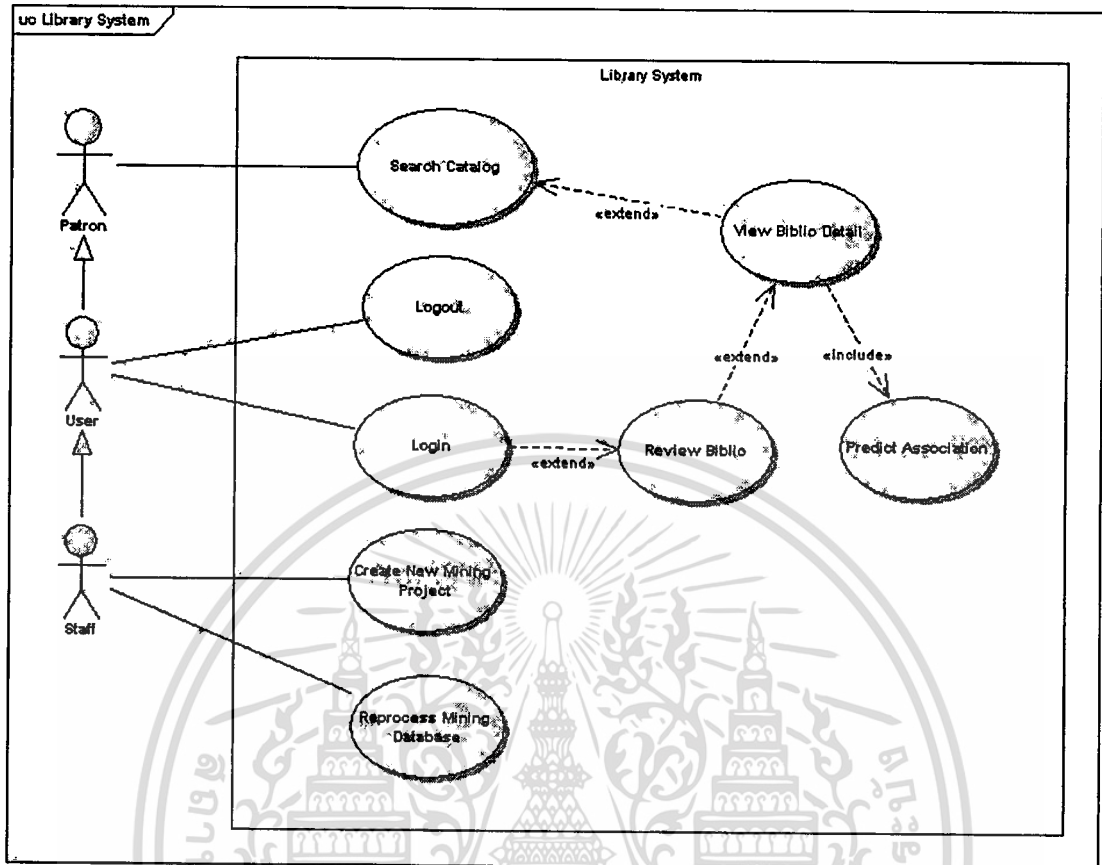
จากการศึกษาระบบ สามารถสรุปความต้องการของระบบงานใหม่ได้ดังนี้

- ผู้ใช้บริการสามารถค้นหาและแสดงรายละเอียดของทรัพยากรห้องสมุดได้
- ระบบสามารถสร้างและจัดการแบบจำลองการค้า ไม่นิ่งได้
- ระบบสามารถแนะนำรายการของทรัพยากรห้องสมุดที่น่าจะมีความเกี่ยวข้องหรือมีความสัมพันธ์กันกับทรัพยากรห้องสมุดที่ผู้บริการกำลังพิจารณาอยู่
- ผู้บริการห้องสมุดสามารถแสดงความคิดเห็นพร้อมทั้งกำหนดระดับคะแนนความพึงพอใจ (Rating) ที่มีต่อทรัพยากรห้องสมุดได้

3.4 ยูสเคสของระบบงานใหม่

ยูสเคสซึ่งใช้อธิบายการทำงานที่เกิดขึ้นในระบบ ดังแสดงในรูปที่ 3.1 ประกอบไปด้วย

- Search Catalog เป็นยูสเคสที่ใช้ในการสืบค้นข้อมูลทรัพยากรห้องสมุด
- View Biblio Detail เป็นยูสเคสที่ใช้ในการแสดงรายละเอียดของทรัพยากรห้องสมุด
- Predict Association เป็นยูสเคสที่ใช้ในการทำนายกฎความสัมพันธ์
- Review Biblio เป็นยูสเคสที่ใช้ในการเขียนแสดงความคิดเห็นและกำหนดระดับคะแนนให้กับทรัพยากรห้องสมุด
- Login เป็นยูสเคสที่ใช้ในการล็อกอินเข้าสู่ระบบ
- Logout เป็นยูสเคสที่ใช้ในการออกจากระบบ
- Create New Mining Project เป็นยูสเคสที่ใช้ในการสร้างและประมวลผลฐานข้อมูล ไม่นิ่งใหม่
- Reprocess Mining Database เป็นยูสเคสที่ใช้ในการประมวลผลฐานข้อมูล ไม่นิ่งที่มีอยู่แล้ว



รูปที่ 3.1 ยูสเคสโคอะแกรมของระบบงานใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 รายละเอียดประกอบยูสเคส Search Catalog

Use Case Name: Search Catalog	ID: UC1
Brief Description: ค้นหารายการทรัพยากรห้องสมุดตามพารามิเตอร์ที่กำหนด	
Primary Actor: Patron	
Trigger: ผู้ใช้งานทำการค้นหารายการทรัพยากรห้องสมุด	
Relationships: Include: Extend:	
Normal Flow of Events: <ol style="list-style-type: none"> 1. ผู้ใช้งานกำหนดพารามิเตอร์ที่ต้องการใช้ในการค้นหาข้อมูล 2. ระบบทำการค้นหาข้อมูล 3. ระบบแสดงข้อมูลรายการที่พบพร้อมทั้งแสดงรายละเอียดต่างๆ ได้แก่ ชื่อเรื่อง, ชื่อผู้แต่ง, ปีที่พิมพ์ และสถานภาพการยืม 	
Alternate/Exceptional Flows: <ol style="list-style-type: none"> 2a. ถ้าไม่พบข้อมูล ระบบทำการแจ้งให้ลูกค้าทราบ และจบการทำงานของยูสเคสนี้ 	

ตารางที่ 3.2 รายละเอียดประกอบยูสเคส View Biblio Detail

Use Case Name: View Biblio Detail	ID: UC2
Brief Description: แสดงรายละเอียดของทรัพยากรห้องสมุด ระดับคะแนน(rating) เฉลี่ยที่ได้รับจากผู้ให้บริการห้องสมุด	
Primary Actor:	
Trigger: ผู้ใช้งานต้องการดูรายละเอียดของทรัพยากรห้องสมุด	
Relationships:	
Include:	Predict Association
Extend:	Search Catalog
Normal Flow of Events:	
<ol style="list-style-type: none"> 1. ผู้ใช้งานต้องการดูรายละเอียดของทรัพยากรห้องสมุด 2. ระบบจะดำเนินการตาม UC3 (Predict Association) 3. ระบบแสดงรายละเอียดของทรัพยากรห้องสมุด ได้แก่ รูป(ถ้ามี), ชื่อเรื่อง, หมายเลข ISBN, ปีที่พิมพ์, สำนักพิมพ์, และจำนวนหน้าทั้งหมด เป็นต้น พร้อมทั้งแสดงรายการทรัพยากรห้องสมุดที่ได้รับการทำนากฎความสัมพันธ์ และความคิดเห็นและระดับคะแนนจากผู้ใช้งาน 	
Alternate/Exceptional Flows:	
3a. ถ้าไม่มีกฎความสัมพันธ์ที่เหมาะสม ระบบจะไม่แสดงกฎความสัมพันธ์	

ตารางที่ 3.3 รายละเอียดประกอบยูสเคส Predict Association

Use Case Name: Predict Association	ID: UC3
Brief Description: ทำนายกฎความสัมพันธ์	
Primary Actor:	
Trigger: ผู้ใช้งานต้องการดูรายละเอียดของทรัพยากรห้องสมุด	
Relationships: Include: Extend:	
Normal Flow of Events: 1. รับพารามิเตอร์ที่ต้องใช้ในการทำนายกฎความสัมพันธ์มาจาก UC2 (View Biblio Detail) 1. ระบบทำการทำนายกฎความสัมพันธ์ 2. ส่งผลลัพธ์ให้กับ UC2	
Alternate/Exceptional Flows:	

ตารางที่ 3.4 รายละเอียดประกอบยูสเคส Review Biblio

Use Case Name: Review Biblio	ID: UC4
Brief Description: ผู้ใช้งานที่ล็อกอินเข้าระบบแล้วสามารถเขียนแสดงความคิดเห็นและกำหนดระดับคะแนน (rating) ให้กับทรัพยากรห้องสมุด	
Primary Actor:	
Trigger: ผู้ใช้งานต้องการแสดงความคิดเห็นและกำหนดระดับคะแนน (rating) ให้กับทรัพยากรห้องสมุด	
Relationships: Include: Extend: View Catalog Detail	
Normal Flow of Events: 1. ผู้ใช้งานต้องการแสดงความคิดเห็นและกำหนดระดับคะแนน (rating) ให้กับทรัพยากรห้องสมุด 2. ระบบแสดงหน้าจอให้ผู้ใช้งานแสดงความคิดเห็นและกำหนดระดับคะแนน	
Alternate/Exceptional Flows: 1a. ถ้าหากว่าผู้ใช้งานยังไม่ได้ล็อกอินเข้าสู่ระบบ ให้ระบบดำเนินการตาม UC5(Login)	

ตารางที่ 3.5 รายละเอียดประกอบยูสเคส Login

Use Case Name: Login	ID: UC5
Brief Description: ผู้ใช้งานป้อนข้อมูล ชื่อผู้ใช้งานและรหัสผ่านเพื่อล็อกอินเข้าสู่ระบบ	
Primary Actor: User	
Trigger: ผู้ใช้งานต้องการล็อกอินเข้าสู่ระบบ	
Relationships: Include: Extend: Review Biblio	
Normal Flow of Events: 1. ผู้ใช้งานป้อนข้อมูล ชื่อผู้ใช้งานและรหัสผ่านเพื่อล็อกอินเข้าสู่ระบบ 2. ระบบจะทำการตรวจสอบชื่อผู้ใช้งานและรหัสผ่านกับฐานข้อมูล 3. ระบบจะแจ้งให้ทราบว่าผู้ใช้งานได้ล็อกอินเรียบร้อยแล้ว	
Alternate/Exceptional Flows: 2a. ถ้าไม่มีชื่อผู้ใช้งานดังกล่าวหรือรหัสผ่านไม่ถูกต้อง ระบบจะแสดงหน้าจอล็อกอินใหม่ อีกครั้ง	

ตารางที่ 3.6 รายละเอียดประกอบยูสเคส Logout

Use Case Name: Logout	ID: UC6
Brief Description: ผู้ใช้งานล็อกเอาต์ออกจากระบบ	
Primary Actor: User	
Trigger: ผู้ใช้งานต้องการออกจากระบบ	
Relationships: Include: Extend:	
Normal Flow of Events: 1. ผู้ใช้งานต้องการออกจากระบบ 2. ระบบทำการล็อกเอาต์ผู้ใช้ออกจากระบบ และแสดงข้อความแจ้งให้ผู้ใช้ทราบ	
Alternate/Exceptional Flows:	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 รายละเอียดประกอบยูสเคส Create New Mining Project

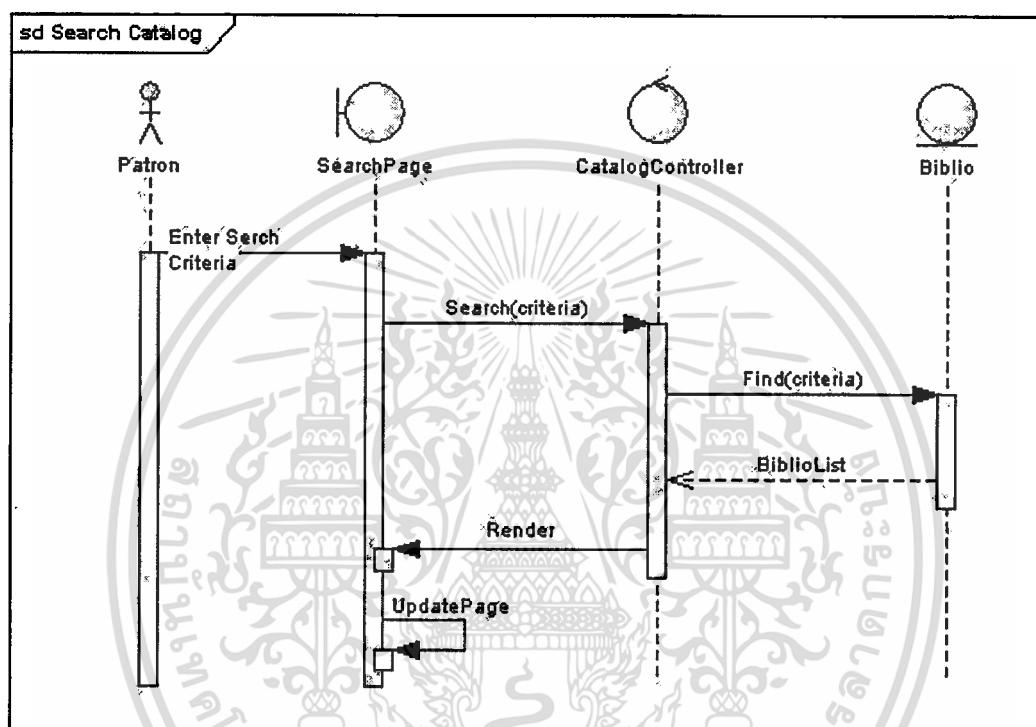
Use Case Name: Create New Mining Project	ID: UC7
Brief Description: เจ้าหน้าที่สร้างฐานข้อมูลและแบบจำลองไมน์นิ่งขึ้นมาใหม่ แล้วทำการประมวลผลฐานข้อมูล	
Primary Actor: Staff	
Trigger: เจ้าหน้าที่ที่ต้องการสร้างฐานข้อมูลและแบบจำลองไมน์นิ่งขึ้นมาใหม่	
Relationships: Include: Extend:	
Normal Flow of Events: <ol style="list-style-type: none"> 1. เจ้าหน้าที่กำหนดพารามิเตอร์ต่างๆ และยืนยันข้อมูล 2. ระบบทำการตรวจสอบความถูกต้องของข้อมูล 3. ระบบสร้างฐานข้อมูลใหม่ 4. ระบบทำการเชื่อมต่อกับแหล่งข้อมูล 5. ระบบสร้าง Mining Structure ใหม่ 6. ระบบสร้าง Mining Model ใหม่ 7. ระบบทำการ Process Model 	
Alternate/Exceptional Flows: <ol style="list-style-type: none"> 2a. ถ้าข้อมูลไม่ถูกต้อง ระบบแจ้งให้ผู้ใช้ป้อนข้อมูลใหม่ 3a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 4a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 5a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 6a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 7a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 	

ตารางที่ 3.8 รายละเอียดประกอบยูสเคส Reprocess Mining Database

Use Case Name: Reprocess Mining Database	ID: UC8
Brief Description: เจ้าหน้าที่กำหนดพารามิเตอร์และทำการประมวลผลแบบจำลองไมน์นิ่งที่ได้สร้างไว้แล้ว	
Primary Actor: Staff	
Trigger: เจ้าหน้าที่ที่ต้องการเปลี่ยนค่าของพารามิเตอร์ หรือต้องการประมวลผลแบบจำลองไมน์นิ่งอีกครั้ง	
Relationships: Include: Extend:	
Normal Flow of Events: <ol style="list-style-type: none"> 1. เจ้าหน้าที่กำหนดพารามิเตอร์ที่ต้องการแก้ไข และยืนยันข้อมูล 2. ระบบทำการตรวจสอบความถูกต้องของข้อมูล 3. ระบบทำการเชื่อมต่อกับแหล่งข้อมูล 4. ระบบจะทำการประมวลผลฐานข้อมูล 	
Alternate/Exceptional Flows: <ol style="list-style-type: none"> 2a. ถ้าข้อมูลไม่ถูกต้อง ระบบแจ้งให้ผู้ใช้ป้อนข้อมูลใหม่ 3a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 4a. ถ้าระบบไม่สามารถทำงานได้สำเร็จ ระบบจะแจ้งความผิดพลาดที่เกิดขึ้น 	

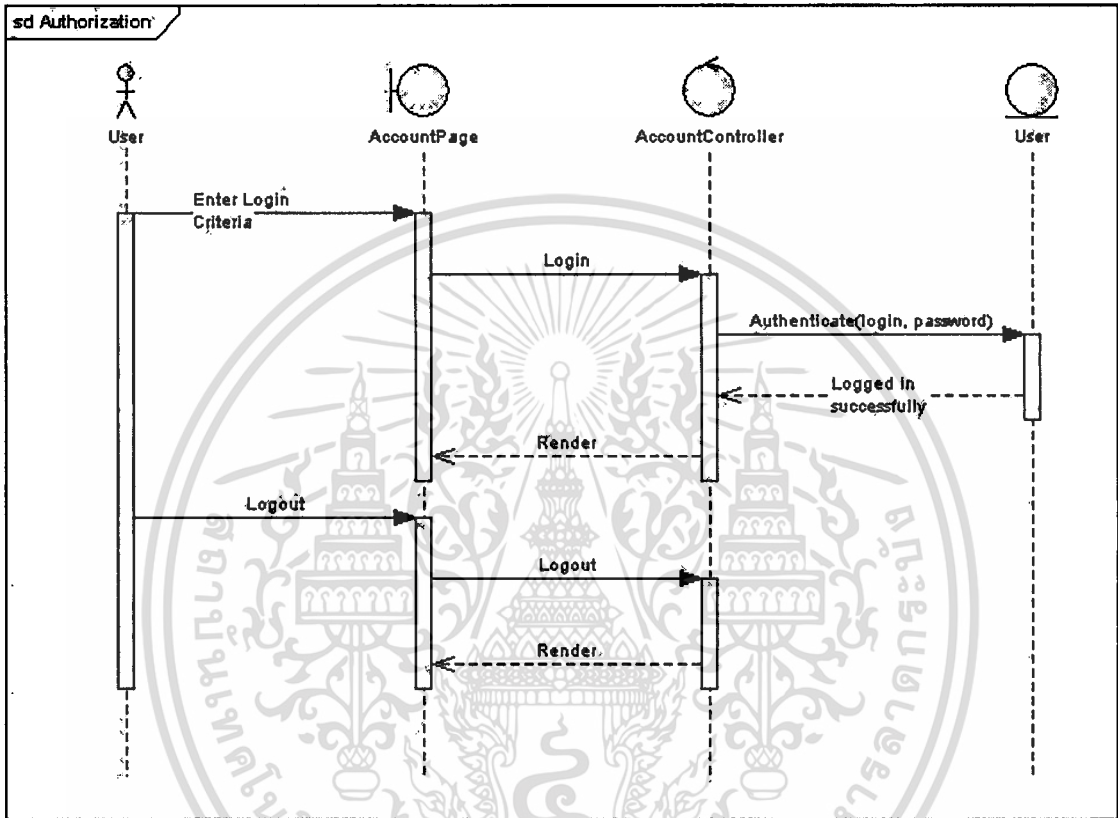
3.5 ซีเควนซ์ไคอะแกรมของระบบงานใหม่

ซีเควนซ์ไคอะแกรมของยูสเคส Search Catalog ดังแสดงในรูปที่ 3.2 เป็นซีเควนซ์ไคอะแกรมที่แสดงการทำงานในระหว่างที่ผู้ใช้งานต้องการสืบค้นข้อมูล โดยผู้ใช้งานจะทำการกำหนดค่าพารามิเตอร์ต่างๆ ของข้อมูลที่ต้องการสืบค้นที่หน้าสืบค้นข้อมูล ซึ่งจะเรียกใช้งานคลาส CatalogController เพื่อค้นหาข้อมูลจากฐานข้อมูลและแสดงผลลัพธ์ให้กับผู้ใช้



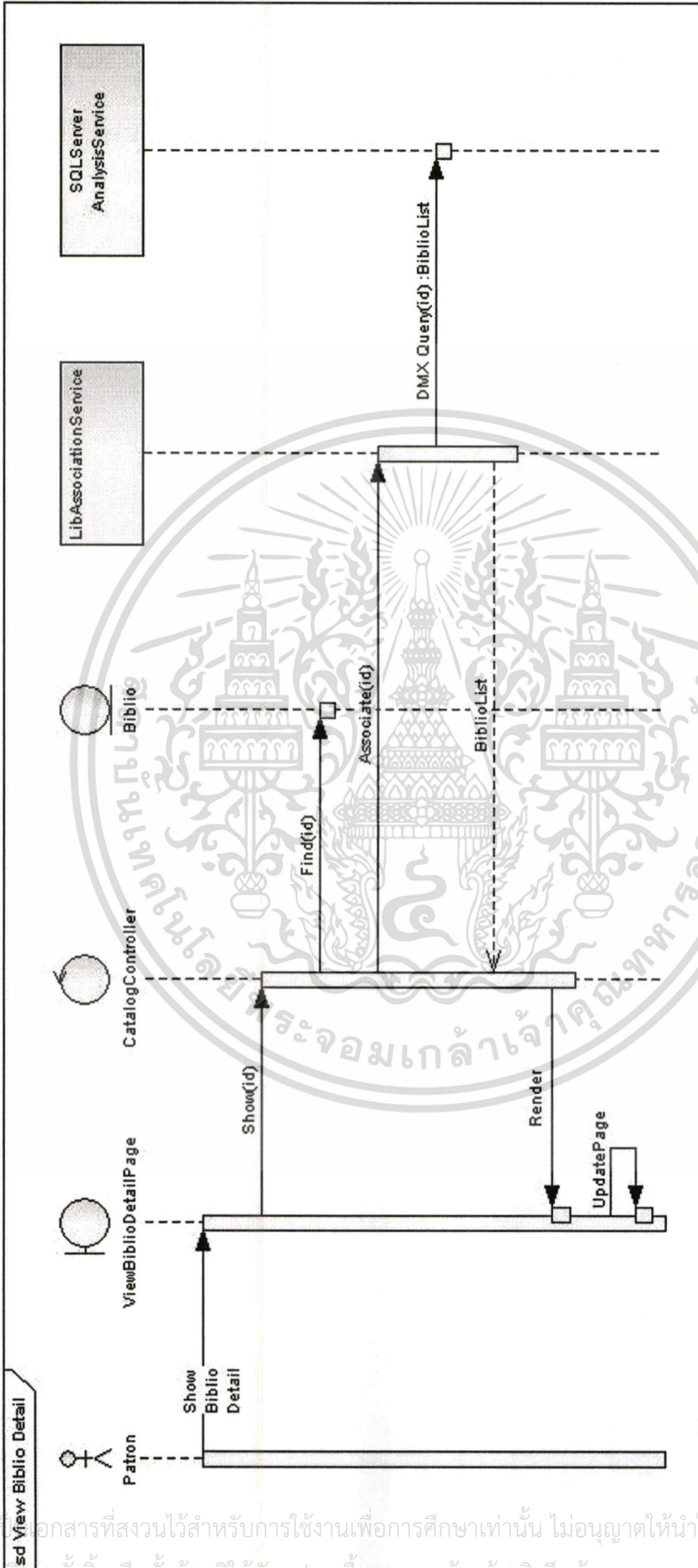
รูปที่ 3.2 ซีเควนซ์ไคอะแกรมของยูสเคส Search Catalog

ซีควเอนซ์ไดอะแกรม Authorization ดังแสดงในรูปที่ 3.3 จะแสดงขั้นตอนการทำงานในระหว่างที่ผู้ใช้งานระบบต้องการล็อกอินเข้าสู่ระบบและการออกจากระบบ ผู้ใช้งานจะทำการป้อนข้อมูลล็อกอินและรหัสผ่านที่หน้าจอล็อกอิน แล้วคลาส AccountController จะทำการตรวจสอบข้อมูลกับฐานข้อมูล ถ้าล็อกอินและรหัสผ่านถูกต้องก็จะอนุญาตให้เข้าสู่ระบบ



รูปที่ 3.3 ซีควเอนซ์ไดอะแกรมของยูสเคส Login และ Logout

ซีควเอนซ์ไดอะแกรมของยูสเคส View Biblio Detail ดังแสดงในรูปที่ 3.4 จะแสดงขั้นตอนการทำงานในระหว่างที่ผู้ใช้งานต้องการทราบรายละเอียดของทรัพยากรห้องสมุด โดยคลาส CatalogController จะแสดงรายละเอียดต่างๆ จากฐานข้อมูล พร้อมทั้งเรียกใช้บริการเว็บเซอร์วิสที่ใช้ในการทำนายกฎความสัมพันธ์และแสดงผลลัพธ์ที่ได้



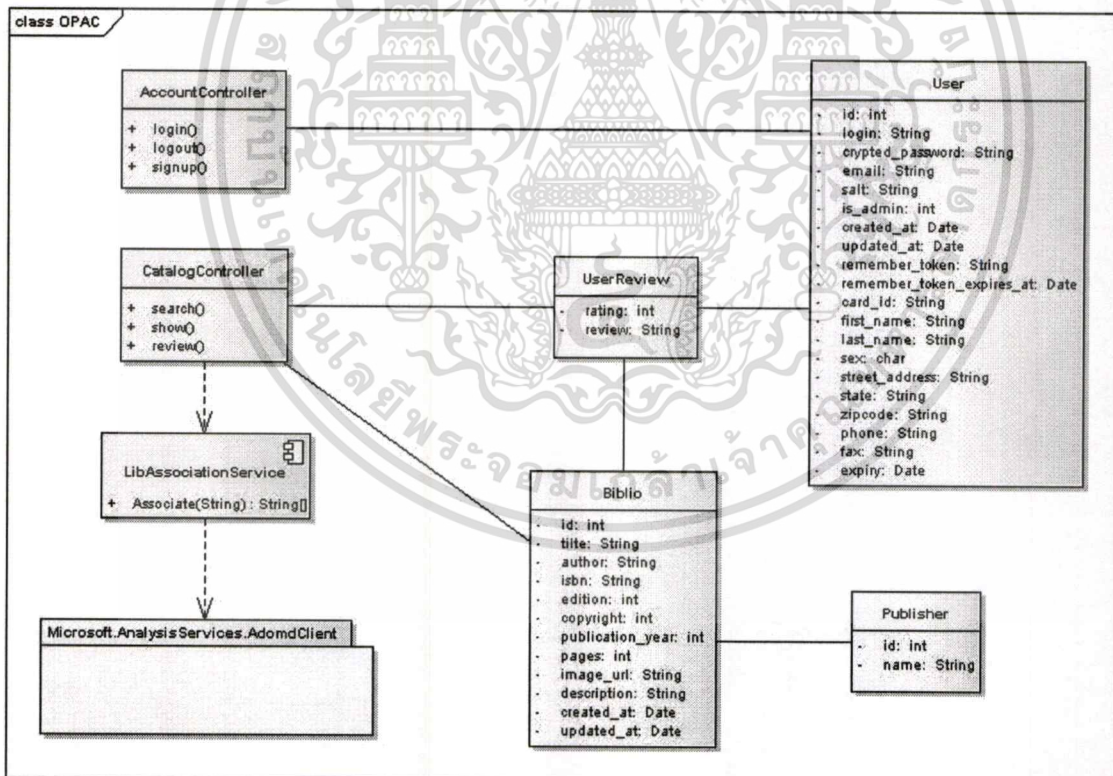
รูปที่ 3.4 ซีเควนซ์ไดอะแกรมของยูสเคส View Biblio Detail

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตีพิมพ์ลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6 คลาสไดอะแกรมของระบบงานใหม่

คลาสดิอะแกรมของระบบเว็บห้องสมุดดังที่แสดงในรูปที่ 3.5 ประกอบไปด้วย

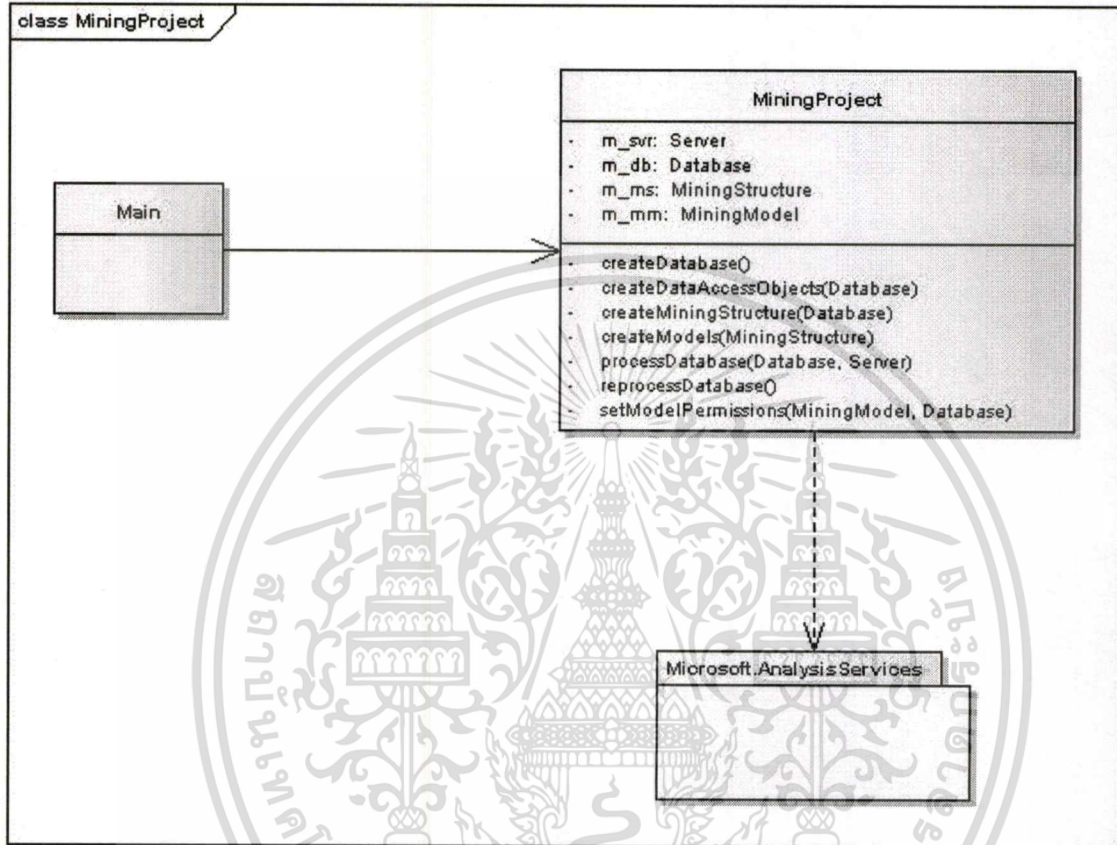
- คลาส User ทำหน้าที่จัดการสถานะต่างๆ ของสมาชิกห้องสมุด
- คลาส Biblio ทำหน้าที่จัดการสถานะต่างๆ ของทรัพยากรห้องสมุด
- คลาส Publisher ทำหน้าที่จัดการสถานะต่างๆ ของสำนักพิมพ์
- คลาส UserReview ทำหน้าที่จัดการสถานะต่างๆ ของการแสดงความคิดเห็นของผู้ใช้งาน
- คลาส AccountController ทำหน้าที่ควบคุมการทำงานในส่วนที่เกี่ยวข้องกับการจัดการสมาชิกห้องสมุด
- คลาส CatalogController ทำหน้าที่ควบคุมการทำงานในส่วนที่เกี่ยวข้องกับการค้นหาข้อมูล, การแสดงรายละเอียดของทรัพยากรห้องสมุด และการแสดงความคิดเห็นของสมาชิกห้องสมุด



รูปที่ 3.5 คลาสดิอะแกรมของเว็บระบบห้องสมุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คลาสไดอะแกรมของการจัดการค้ำไมน์นิ่งดังแสดงในรูปที่ 3.6 นั้น ใช้ในการสร้างและประมวลผลฐานข้อมูลไมน์นิ่งใน Analysis Services ของ SQL Server 2005



รูปที่ 3.6 คลาสไดอะแกรมของการจัดการค้ำไมน์นิ่ง

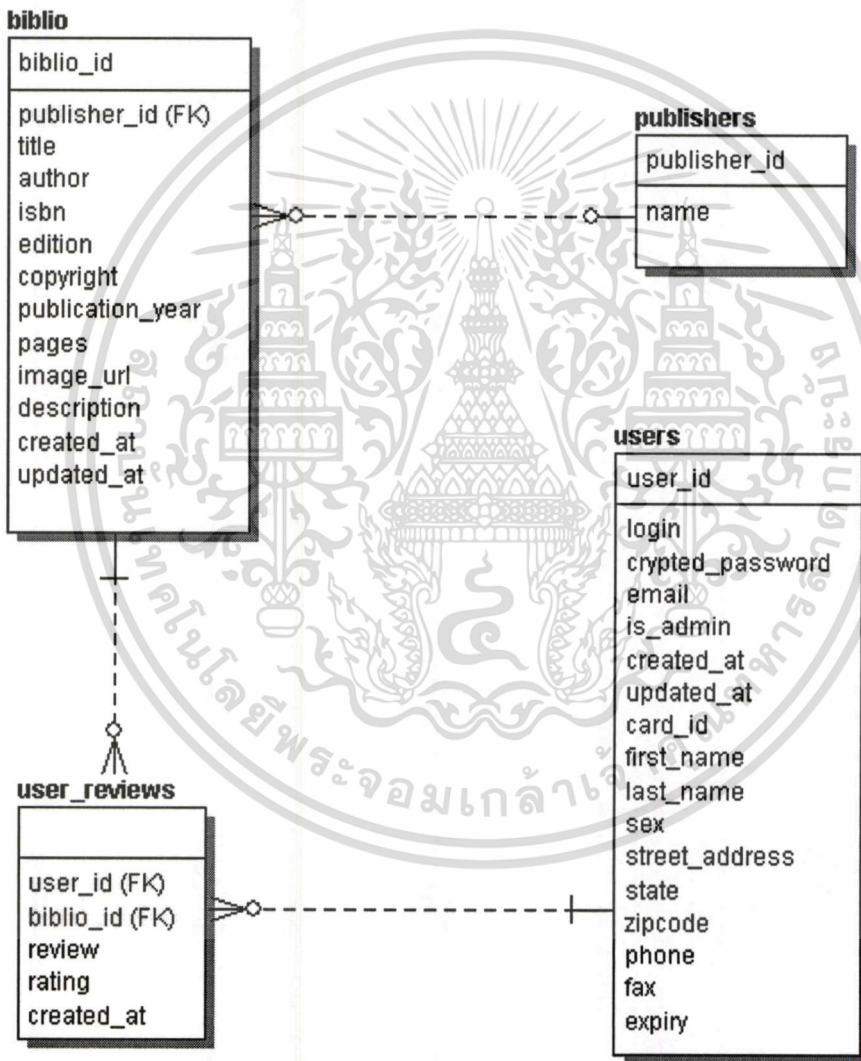
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7 การออกแบบฐานข้อมูลของระบบงานใหม่

ฐานข้อมูลของระบบเว็บห้องสมุดดังแสดงในรูปที่ 3.7 ประกอบไปด้วยเอนทิตีต่างๆ ดังนี้

- biblio จัดเก็บข้อมูลในส่วนที่เกี่ยวกับทรัพยากรห้องสมุด
- publishers จัดเก็บข้อมูลที่เกี่ยวข้องกับสำนักพิมพ์
- users จัดเก็บข้อมูลที่เกี่ยวข้องกับข้อมูลสมาชิก
- user_reviews จัดเก็บข้อมูลที่เกี่ยวข้องกับการแสดงความคิดเห็นของใช้งาน

โดยมีลำดับดัชนีนาารดังแสดงในตารางที่ 3.9 ถึงตารางที่ 3.12



รูปที่ 3.7 อีอาร์ไดอะแกรมของระบบเว็บห้องสมุด

ตารางที่ 3.9 ตาราง biblio

Attribute	Type	Size	Description	Key	Reference Table
BIBLIO_ID	Int	11	รหัสของทรัพยากร	PK	
PUBLISHER_ID	Int	11	รหัสของสำนักพิมพ์	FK	publishers
TITLE	Varchar	60	ชื่อหนังสือ		
AUTHOR	Varchar	60	ชื่อผู้แต่ง		
ISBN	Varchar	13	ISBN		
EDITION	Smallint	3	EDITION		
COPYRIGHT	Smallint	6	ปีที่สงวนลิขสิทธิ์		
PUBLICATION_YEAR	Smallint	6	ปีที่พิมพ์		
PAGES	Smallint	6	จำนวนหน้า		
IMAGE_URL	Varchar	255	ยูอาร์แอลของรูป		
DESCRIPTION	Text		รายละเอียด		
CREATED_AT	Datetime		เวลาที่บันทึกข้อมูล		
UPDATED_AT	Datetime		เวลาที่แก้ไขข้อมูล		

ตารางที่ 3.10 ตาราง publishers

Attribute	Type	Size	Description	Key	Reference Table
PUBLISHER_ID	Int	11	รหัสของสำนักพิมพ์	PK	
NAME	Varchar	60	ชื่อสำนักพิมพ์		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.11 ตาราง users

Attribute	Type	Size	Description	Key	Reference Table
USER_ID	Int	11	รหัสผู้ใช้งาน	PK	
LOGIN	Varchar	40	ชื่อผู้ใช้งาน		
CRYPTED_PASSWORD	Varchar	40	รหัสผ่านที่เข้ารหัส		
EMAIL	Varchar	80	อีเมล		
IS_ADMIN	Boolean		ผู้ดูแลระบบหรือไม่		
CREATED_AT	Datetime		เวลาที่บันทึกข้อมูล		
UPDATED_AT	Datetime		เวลาที่แก้ไขข้อมูล		
CARD_ID	Varchar	6	หมายเลขบัตร		
FIRST_NAME	Varchar	40	ชื่อ		
LAST_NAME	Varchar	40	นามสกุล		
SEX	Char	1	เพศ		
STREET_ADDRESS	Varchar	50	ที่อยู่		
STATE	Varchar	40	จังหวัด หรือรัฐ		
ZIPCODE	Varchar	8	รหัสไปรษณีย์		
PHONE	Varchar	40	หมายเลขโทรศัพท์		
FAX	Varchar	40	หมายเลขแฟกซ์		
EXPIRY	Date		วันหมดอายุ		

ตารางที่ 3.12 ตาราง user_reviews

Attribute	Type	Size	Description	Key	Reference Table
USER_ID	Int	11	รหัสผู้ใช้งาน	FK	user
BIBLIO_ID	Int	11	รหัสของทรัพยากร	FK	biblio
RATING	Int	1	ระดับคะแนน		
CREATED_AT	Datetime		เวลาที่บันทึกข้อมูล		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนของคุณสมบัติที่ใช้ในการทำดาต้าไมนิ่งนั้นจะมีอีอาร์ไดอะแกรมดังแสดงในรูปที่ 3.8 และมีดาต้าดิกชันนารีดังตารางที่ 3.13 โดยเอนติตี Users นั้นจะเก็บข้อมูลเกี่ยวกับสมาชิกห้องสมุด ส่วนเอนติตี LibCir นั้นจะเก็บข้อมูลประวัติการยืมหนังสือของสมาชิก



รูปที่ 3.8 อีอาร์ไดอะแกรมของข้อมูลที่ใช้ในการค้นหาความสัมพันธ์

ตารางที่ 3.13 ตาราง LibCir

Attribute	Type	Size	Description
user_id	Varchar	11	รหัสผู้ใช้งาน
bib_id	Varchar	11	รหัสของทรัพยากรห้องสมุด

บทที่ 4

การพัฒนาระบบงาน

4.1 ความต้องการของการพัฒนาระบบ

ความต้องการของระบบทางด้านฮาร์ดแวร์และซอฟต์แวร์ที่จำเป็นต้องใช้ในการพัฒนาระบบงาน มีดังนี้

1. ความต้องการทางด้านฮาร์ดแวร์ ประกอบด้วย

- เครื่องคอมพิวเตอร์ ความเร็วหน่วยประมวลผล 2 GHz
- หน่วยความจำ 1 GB
- ฮาร์ดดิสก์ขนาด 120 GB
- ดีวีดีรอม
- จอมอนิเตอร์
- แป้นพิมพ์
- เมาส์

2. ความต้องการทางด้านซอฟต์แวร์ ประกอบด้วย

- ระบบปฏิบัติการ Microsoft Windows XP Professional Service Pack 2
- .NET Framework 2.0
- Ruby on Rails Framework
- โปรแกรม Microsoft SQL Server 2005
- โปรแกรม MySQL
- โปรแกรม Visual Studio 2005
- โปรแกรม NetBeans 6.0
- โปรแกรม Internet Information Services(IIS)
- โปรแกรม Apache

4.2 การทำดาต้าไมนิ่ง

ในขั้นตอนการทำดาต้าไมนิ่งนั้น มีขั้นตอนการดำเนินการ ดังนี้

4.2.1 กำหนดวัตถุประสงค์

วัตถุประสงค์ในการทำดาต้าไมนิ่งร่วมกับระบบห้องสมุดก็เพื่อประยุกต์ใช้การ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ทางการค้า
ค้นหาความสัมพันธ์ร่วมกับระบบห้องสมุด เพื่อให้ระบบช่วยแนะนำหนังสือที่น่าจะมีความ
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกี่ยวข้องกับสัมพันธ์กันให้กับผู้ใช้งาน ซึ่งจะช่วยอำนวยความสะดวกให้กับผู้ใช้งานห้องสมุดและช่วยให้ผู้ใช้งานสามารถใช้งานระบบห้องสมุดได้อย่างมีประสิทธิภาพ สามารถค้นหาหนังสือที่มีความเกี่ยวข้องกับสัมพันธ์กันได้อย่างรวดเร็ว

4.2.2 การรวบรวมข้อมูล

ในขั้นตอนการรวบรวมข้อมูลนั้น ข้อมูลที่จะใช้ในการทำค้ำไมน์นิ่งนำมาจากประวัติการยืมหนังสือของห้องสมุดแห่งหนึ่ง

4.2.3 การเตรียมข้อมูล

ในขั้นตอนการเตรียมข้อมูลจะทำการจัดเตรียมข้อมูลเป็น 2 ตาราง คือ ตาราง Users และ ตาราง LibCir โดยตาราง Users นั้นจะเก็บข้อมูลเกี่ยวกับสมาชิกห้องสมุด มีแอตทริบิวต์ user_id เป็นไพรมารีคีย์ ส่วนตาราง LibCir นั้นจะเก็บข้อมูลประวัติการยืม โดยมีแอตทริบิวต์ user_id และ bib_id ดังแสดงในรูปที่ 3.8

4.2.4 การทำค้ำไมน์นิ่ง

ในขั้นตอนการทำค้ำไมน์นิ่งนั้น ในการพัฒนาระบบครั้งนี้จะทำการพัฒนาโปรแกรมประยุกต์เพื่อใช้ในการสร้างและประมวลผลแบบจำลองค้ำไมน์นิ่ง

4.3 การพัฒนาโปรแกรม

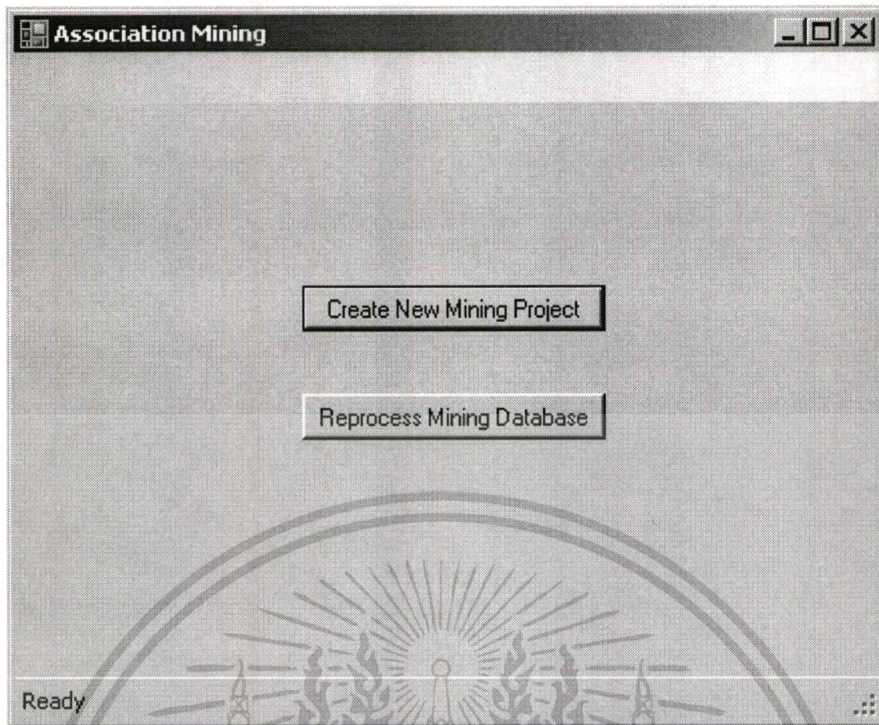
ในการพัฒนาระบบในครั้งนี้จะพัฒนาโปรแกรมขึ้นมาโดยแบ่งเป็น 3 ส่วนหลักๆ คือ

- โปรแกรมสร้างและประมวลผลแบบจำลองค้ำไมน์นิ่ง
- เว็บเซอร์วิสสำหรับทำนายกฎความสัมพันธ์
- โปรแกรมประยุกต์แบบเว็บสำหรับให้บริการสืบค้นข้อมูล

4.4 โปรแกรมสร้างและประมวลผลแบบจำลองค้ำไมน์นิ่ง

สำหรับโปรแกรมที่ใช้ในการสร้างและประมวลผลแบบจำลองค้ำไมน์นิ่งนั้น จะพัฒนาเป็นแบบวินโดวส์ฟอร์มโดยใช้เทคโนโลยี .NET โดยเมื่อเรียกใช้งานโปรแกรมนั้นจะมีเมนูหลักดังรูปที่

4.1



รูปที่ 4.1 เมนูหลักของโปรแกรมจัดการดาต้าไมนิง

ถ้าผู้ใช้งานคลิกที่ปุ่ม Create New Mining Project จะเข้าสู่โปรแกรมในส่วนของการสร้างและประมวลผลฐานข้อมูลใหม่ ดังรูปที่ 4.2

Create New Mining Project

Analysis Service

Analysis Service Server: localhost

New Database: LibAnalysis

Existing Database: [Empty]

Data Source

Data Source Server: localhost

Initial Catalog: [Empty]

Mining Model

Mining Structure Name: LibCir

Mining Model Name: LibCirAssoc

Minimum Probability: [Empty]

Minimum Support: [Empty]

Create and Process New Database Reprocess Database

Ready

รูปที่ 4.2 การสร้างและประมวลผลฐานข้อมูลใหม่

แต่ถ้าหากว่าผู้ใช้งานคลิกที่ปุ่ม Reprocess Mining Database จะเข้าสู่โปรแกรมในส่วนของการประมวลผลฐานข้อมูลที่ได้สร้างไว้แล้ว ดังรูปที่ 4.3

Reprocess Mining Model

Analysis Service

Analysis Service Server: localhost

New Database: LibAnalysis

Existing Database: [Empty]

Data Source

Data Source Server: localhost

Initial Catalog: [Empty]

Mining Model

Mining Structure Name: LibCir

Mining Model Name: LibCirAssoc

Minimum Probability: [Empty]

Minimum Support: [Empty]

Buttons: Create and Process New Database, Reprocess Database

Status: Ready

รูปที่ 4.3 การประมวลผลฐานข้อมูลที่ได้สร้างไว้แล้ว

4.5 เว็บเซอร์วิสสำหรับทำนายกฎความสัมพันธ์

โปรแกรมที่ใช้ทำนายกฎความสัมพันธ์นั้นจะพัฒนาเป็นลักษณะเว็บเซอร์วิสโดยใช้เทคโนโลยี .NET ทำงานร่วมกับระบบจัดการฐานข้อมูล SQL Server 2005 และ Internet Information Services (IIS) โดยจะมีไฟล์ Web Service Description Language (WSDL) ที่สามารถเข้าถึงได้โดยผ่านทาง <http://localhost/LibAssociationService/LibAssociationService.asmx?WSDL> ดังที่แสดงในรูปที่ 4.4

```

- <wsdl:definitions targetNamespace="http://mydomain.int/">
- <wsdl:types>
- <s:schema elementFormDefault="qualified" targetNamespace="http://mydomain.int/">
- <s:element name="Associate">
- <s:complexType>
- <s:sequence>
- <s:element minOccurs="0" maxOccurs="1" name="bibid" type="s:string"/>
- </s:sequence>
- </s:complexType>
- </s:element>
- <s:element name="AssociateResponse">
- <s:complexType>
- <s:sequence>
- <s:element minOccurs="0" maxOccurs="1" name="AssociateResult" type="tns:ArrayOfString"/>
- </s:sequence>
- </s:complexType>
- </s:element>
- <s:complexType name="ArrayOfString">
- <s:sequence>
- <s:element minOccurs="0" maxOccurs="unbounded" name="string" nillable="true" type="s:string"/>
- </s:sequence>
- </s:complexType>
- </s:schema>
- </wsdl:types>

```

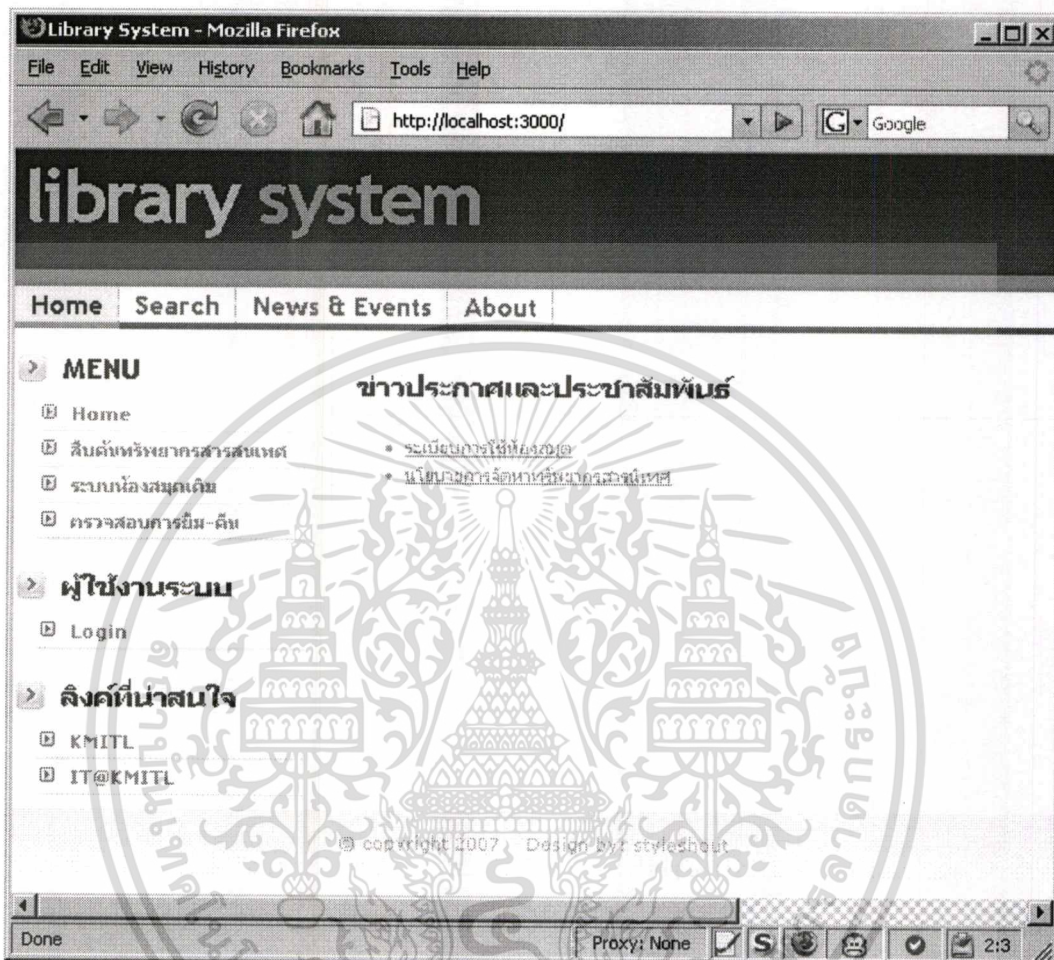
รูปที่ 4.4 WSDL ของเว็บเซอร์วิสที่ใช้ทำนายภูมิกวามสัมพันธ์

4.6 ระบบสืบค้นข้อมูลห้องสมุดผ่านเว็บ

สำหรับระบบสืบค้นข้อมูลห้องสมุดนั้นจะให้บริการหลักในส่วนของการสืบค้นข้อมูลของห้องสมุด แสดงรายละเอียดของทรัพยากรห้องสมุด และเรียกใช้เว็บเซอร์วิสที่ทำนายภูมิกวามสัมพันธ์

สำหรับระบบนี้จะพัฒนาในลักษณะของโปรแกรมประยุกต์แบบเว็บ โดยใช้ภาษา Ruby ทำงานอยู่บนเฟรมเวิร์ค Ruby On Rails ร่วมกับระบบจัดการฐานข้อมูล MySQL Server โดยมีส่วนประกอบต่างๆ ดังนี้

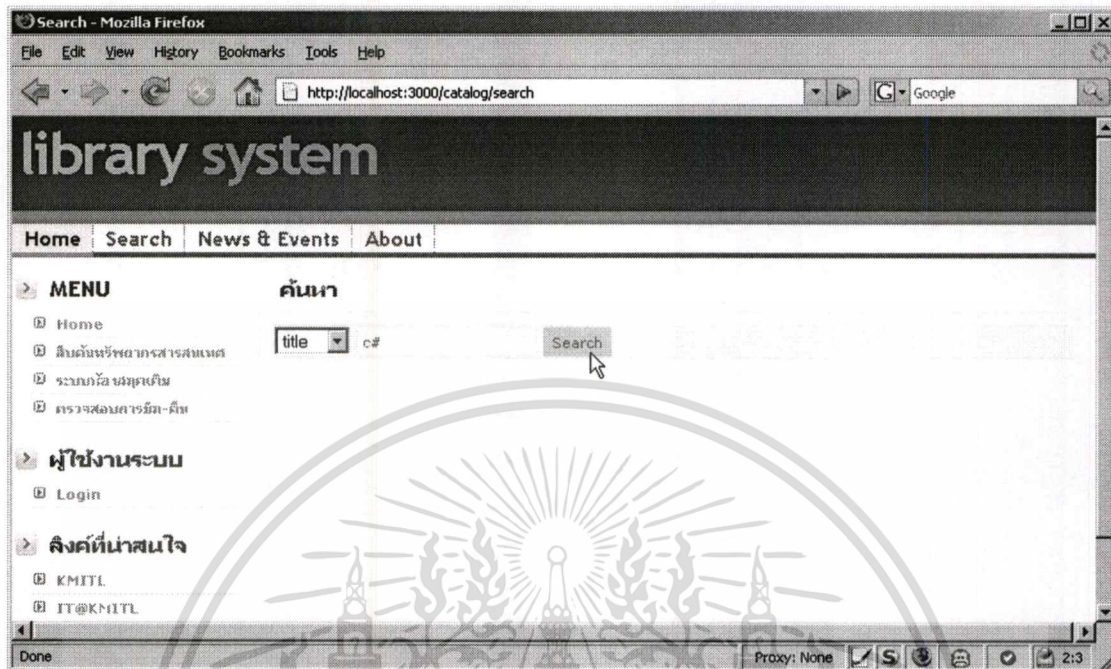
1. หน้าโฮมเพจจะมีลักษณะดังแสดงในรูปที่ 4.5



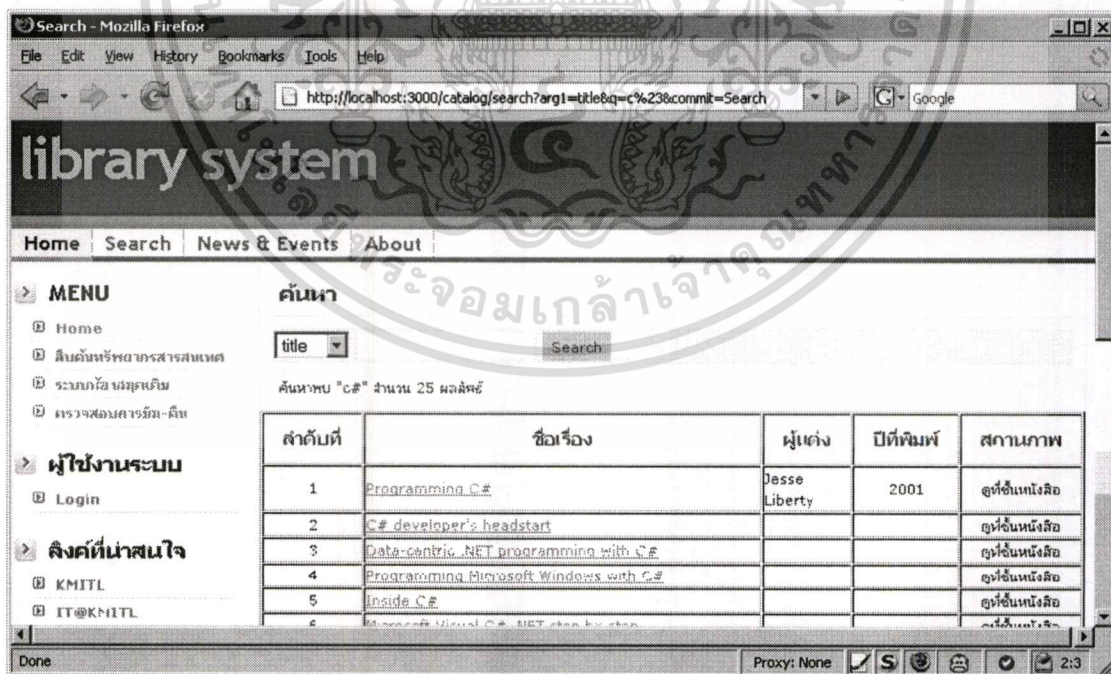
รูปที่ 4.5 หน้าโฮมเพจของเว็บระบบห้องสมุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ส่วนของการสืบค้นข้อมูล ดังแสดงในรูปที่ 4.7



รูปที่ 4.6 ส่วนของการสืบค้นข้อมูล



รูปที่ 4.7 หน้าแสดงผลพัทธ์จากการสืบค้นข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The screenshot shows a Mozilla Firefox browser window with the address bar displaying `http://localhost:3000/catalog/show/2001015951`. The page content includes a navigation menu on the left, a book cover for 'Programming C#' by Jesse Liberty, and a list of related books. A large watermark of a university seal is overlaid on the page.

MENU

- Home
- สืบค้นทรัพยากรสารสนเทศ
- ระบบห้องสมุดเดิม
- ตรวจสอบการยืม-คืน

ผู้ใช้งานระบบ

ยินดีต้อนรับคุณ chart

- Logout

ลิงค์ที่น่าสนใจ

- KNITL
- IT@KNITL

Programming C#

ชื่อเรื่อง : Programming C#
 ผู้แต่ง : Jesse Liberty
 ISBN : 0596001177
 ปีที่พิมพ์ : 2001
 สำนักพิมพ์ : O'reilly
 จำนวนหน้า : 680
 Average User Rating : 4.5 /5.0 (แสดงความคิดเห็น)

ผู้ที่ยืมหนังสือเล่มนี้มักจะยืมหนังสือ ดังต่อไปนี้ด้วย

- Professional C# แต่งโดย Simon Robinson พิมพ์เมื่อปี 2002
- MCAD/MCSD self-paced training kit แต่งโดย xxxx พิมพ์เมื่อปี xxxx
- เจียรรี่ ASP.NET ด้วยภาษา C# แต่งโดย xxxx พิมพ์เมื่อปี xxxx
- สถาปัตยกรรมและโปรโตคอลอินเทอร์เน็ต แต่งโดย xxxx พิมพ์เมื่อปี xxxx
- เบสิคของ TCP/IP และโปรโตคอลของอินเทอร์เน็ต แต่งโดย xxxx พิมพ์เมื่อปี xxxx

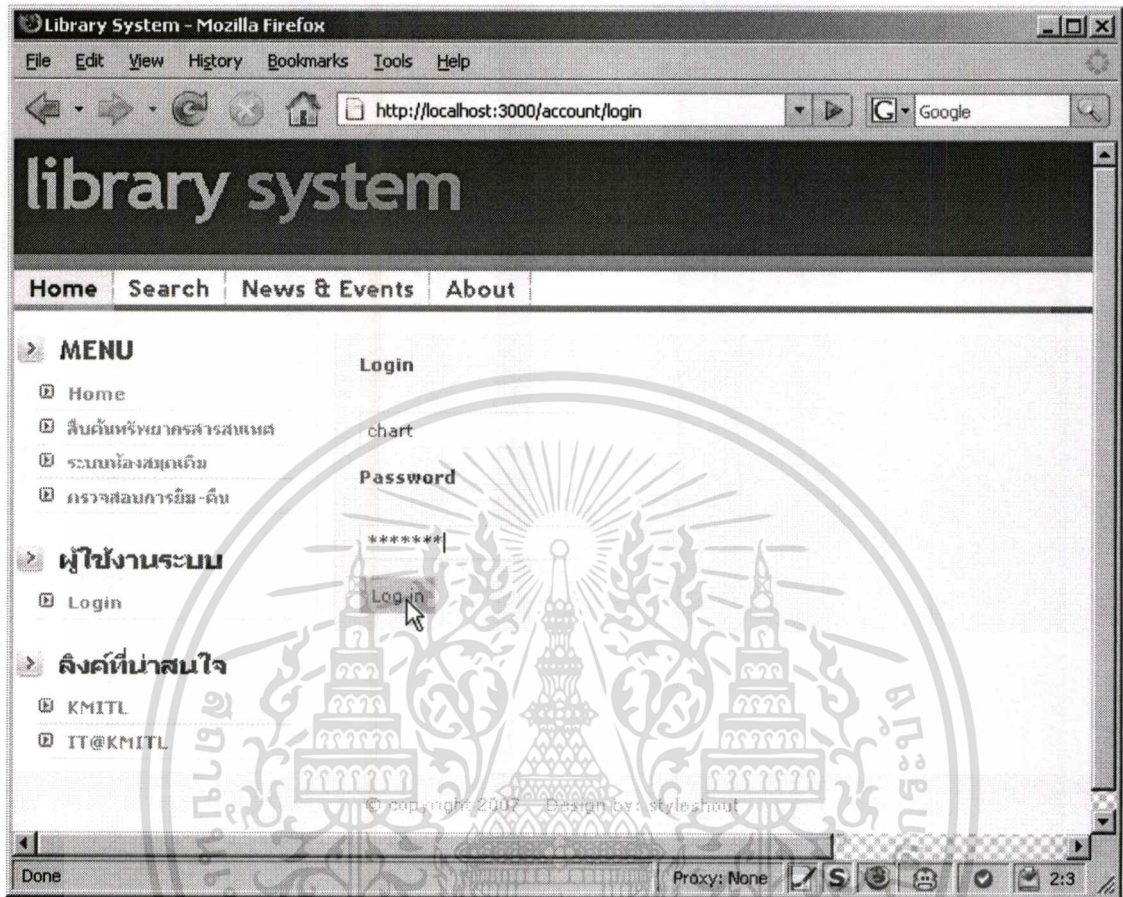
ความคิดเห็นของผู้ใช้งาน

suchart เมื่อ Mon Sep 17 14:19:41 +0700 2007
 คะแนนที่ได้ : 4
 ความคิดเห็น : ดีครับ แนะนำครับ

รูปที่ 4.8 หน้าแสดงรายละเอียดของทรัพยากรห้องสมุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

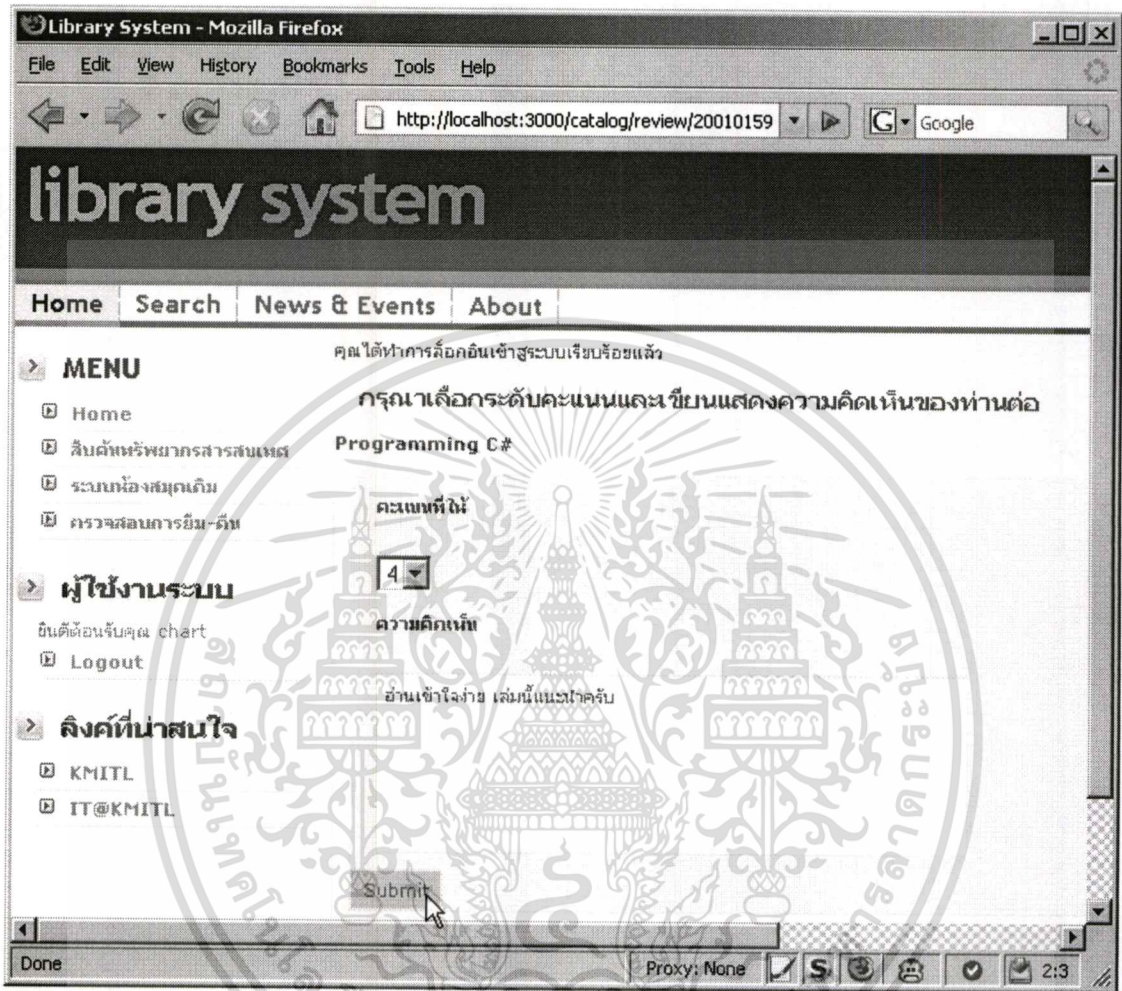
3. ส่วนของการล็อกอินเข้าสู่ระบบ ดังแสดงในรูปที่ 4.9



รูปที่ 4.9 หน้าแสดงส่วนของการล็อกอินเข้าสู่ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ส่วนของการเขียนแสดงความคิดเห็น และกำหนดระดับคะแนนความพึงพอใจของผู้ใช้งานต่อทรัพยากรห้องสมุด ดังแสดงในรูปที่ 4.10



รูปที่ 4.10 ส่วนของการเขียนแสดงความคิดเห็นของผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการดำเนินงาน

โครงการพัฒนาระบบนี้ เป็นโครงการที่จัดทำขึ้นเพื่อเป็นการประยุกต์ใช้การค้นหากฎ ความสัมพันธ์ (Association Rules Discovery) ร่วมกับระบบห้องสมุดโดยใช้ความสามารถใน ส่วน คำค้นหาของระบบจัดการฐานข้อมูล SQL Server 2005 เพื่อช่วยในการแนะนำรายการ ทรัพยากรห้องสมุดที่มีความสัมพันธ์กันให้กับผู้ที่ใช้บริการสืบค้นข้อมูลของระบบห้องสมุด

5.1 สรุปผลการดำเนินงาน

ในการพัฒนาระบบครั้งนี้ได้ทำการพัฒนาโปรแกรมขึ้นมา 3 ส่วน ได้แก่

- โปรแกรมประยุกต์เพื่อใช้ในการสร้างและประมวลผลแบบจำลองคำค้นหาบบ ระบบจัดการฐานข้อมูล SQL Server 2005 โปรแกรมนี้จะช่วยให้ผู้ที่เกี่ยวข้อง สามารถสร้าง, แก้ไขพารามิเตอร์ และประมวลผลแบบจำลองค้นหาได้อย่าง สะดวกและง่ายดาย
- เว็บเซอร์วิสที่ให้บริการในการทำนายกฎความสัมพันธ์ การพัฒนาเป็นเว็บ เซอร์วิสจะทำให้การพัฒนาโปรแกรมประยุกต์ที่มาเรียกใช้บริการได้มีความ ยืดหยุ่นมากขึ้น สามารถทำงานข้ามแพลตฟอร์มได้เป็นอย่างดี
- โปรแกรมประยุกต์แบบเว็บในการให้บริการสืบค้นข้อมูลแก่ผู้ใช้บริการห้องสมุด ซึ่งจะเรียกใช้บริการเว็บเซอร์วิสที่สร้างขึ้น จะช่วยอำนวยความสะดวกให้กับ ผู้ใช้บริการห้องสมุดและช่วยให้ผู้ใช้บริการสามารถใช้บริการห้องสมุดได้อย่างมี ประสิทธิภาพมากยิ่งขึ้น

5.2 ข้อเสนอแนะ

1. ถ้าเป็นไปได้ ควรจะมีการนำข้อมูลอื่นๆ ของผู้ใช้งาน เช่น เพศ, อายุ และคณะหรือ ภาควิชาที่ศึกษา เป็นต้น มาใช้ในการศึกษาและวิเคราะห์ข้อมูลด้วย อาจจะช่วยให้การ ทำนายมีความถูกต้องแม่นยำมากยิ่งขึ้น
2. ระบบนี้ยังสามารถนำไปประยุกต์ใช้กับต่างๆที่มีความเกี่ยวข้องกับการค้นหากฎ ความสัมพันธ์ เช่น การขายสินค้า ได้อีกด้วย

บรรณานุกรม

สุภาภรณ์ บุตรดีวงศ์. 2549. “อัลกอริทึมใหม่ในการจำแนกประเภทเอกสารโดยใช้กฎ
ความสัมพันธ์.” วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Daniel T. Larose. 2005. **Discovering Knowledge in Data : An Introduction to Data Mining.**
New Jersey : John Wiley & Sons, Inc.

Jiawei Han and Micheline Kamber. 2006. **Data Mining: Concepts and Techniques.** 2nd ed.
Morgan Kaufmann.

ZhaoHui Tang and Jamie MacLennan. 2005. **Data Mining with SQL Server 2005.** Indiana :
Wiley Publishing, Inc.

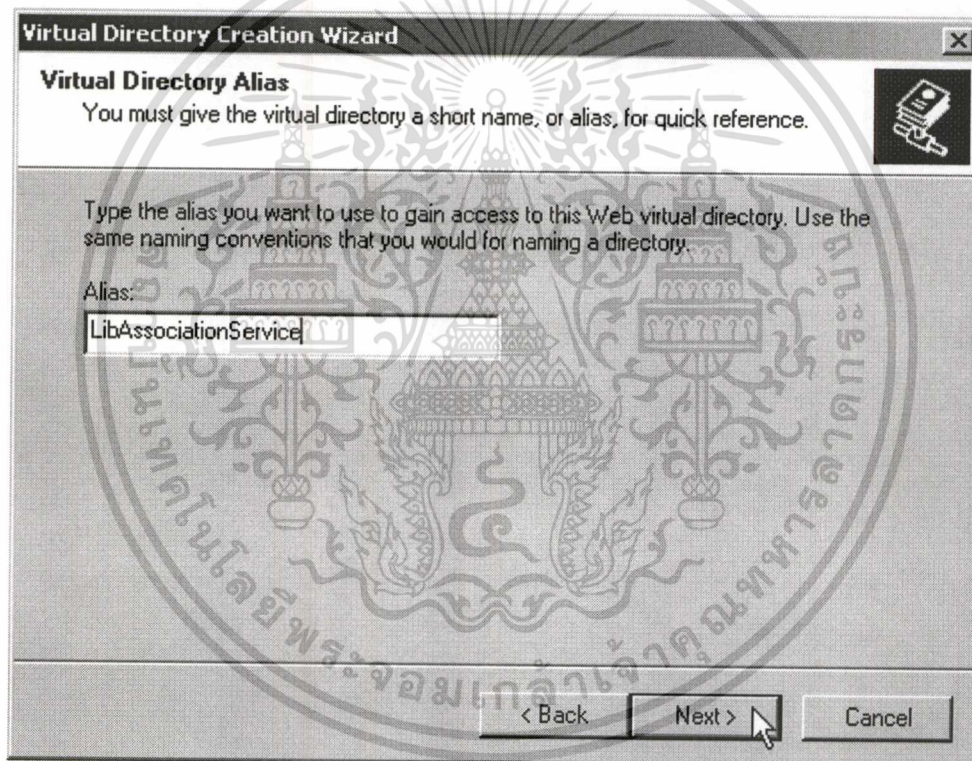


ภาคผนวก ก

คู่มือการติดตั้งระบบ

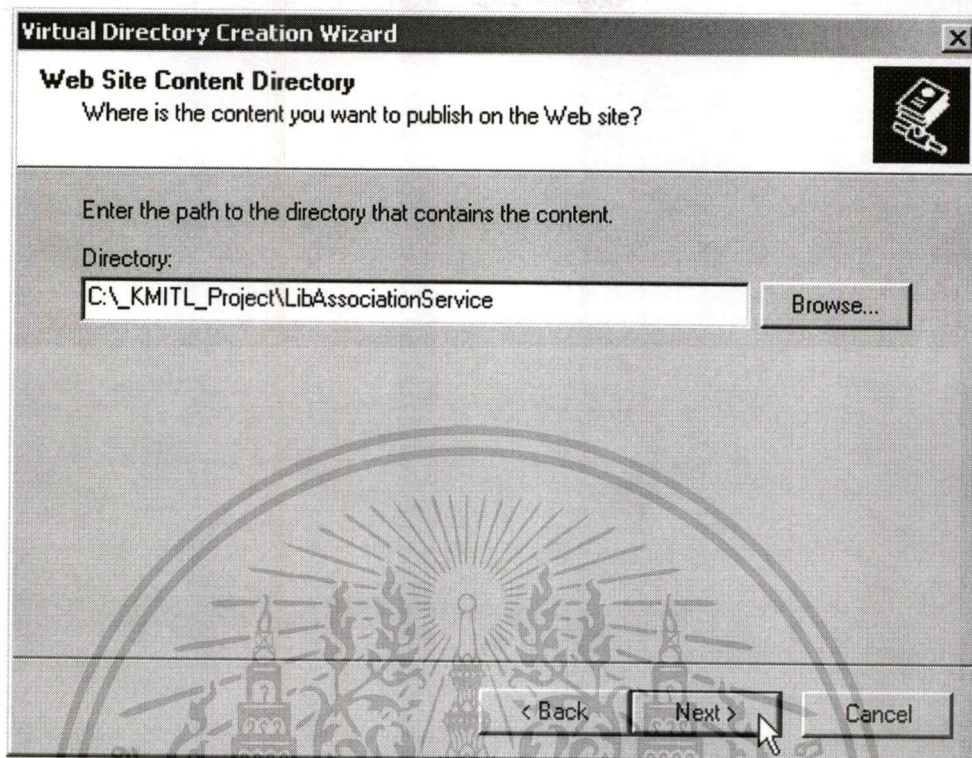
การติดตั้งเว็บเซอร์วิสที่ใช้ทำนายกฏความสัมพันธ์

1. เรียกใช้งาน Internet Information Services จาก Start -> Settings -> Control Panel -> Administrative Tools
2. คลิกขวาที่ Default Web Site เลือก New -> Virtual Directory คลิก Next
3. กำหนดค่า Virtual Directory Alias เป็น LibAssociationService ดังรูปที่ 1ก แล้วคลิก Next



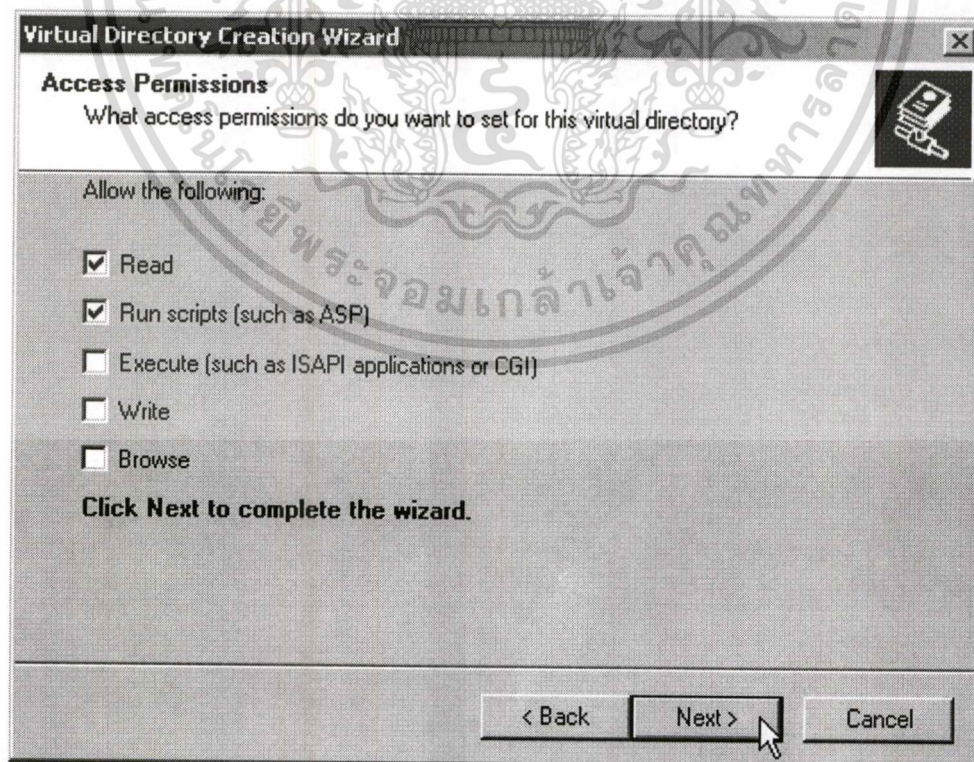
รูปที่ 1ก กำหนด Virtual Directory Alias

4. เลือกไปยังไดเรกทอรี LibAssociationService ดังรูปที่ 2ก แล้วคลิก Next



รูปที่ 2ก กำหนดตำแหน่งที่เก็บเว็บไซต์

5. กำหนด Access Permission ตามค่าปกติ คลิก Next แล้วคลิก Finish



รูปที่ 3ก กำหนด Access Permissions

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การติดตั้ง Ruby on Rails Web Application

ในที่นี้จะแสดงการติดตั้งบนระบบปฏิบัติการลินุกซ์ Ubuntu

1. ทำการอัปเดตระบบ

```
sudo apt-get update
```

```
sudo apt-get upgrade
```

2. ติดตั้งแพคเกจที่จำเป็นในการคอมไพล์โปรแกรม

```
sudo apt-get install build-essential
```

3. ติดตั้งชุดแพคเกจ Ruby

```
sudo apt-get install ruby1.8 ruby1.8-dev ri1.8 rdoc1.8 irb1.8 libopenssl-ruby1.8  
libreadline-ruby1.8
```

4. ติดตั้ง MySQL และไลบรารีที่จำเป็น

```
apt-get install mysql-server libmysql-ruby1.8
```

5. ติดตั้งและอัปเดต rubygem

```
wget http://rubyforge.org/frs/download.php/20989/rubygems-0.9.4.tgz
```

```
tar xvzf rubygems-0.9.4.tgz
```

```
cd rubygems-0.9.4/
```

```
sudo ruby setup.rb
```

```
sudo gem update --system
```

6. ติดตั้ง rails และอื่นๆ ผ่าน gem ด้วยคำสั่งดังนี้

```
sudo gem install rails --include-dependencies
```

```
sudo gem install mongrel --include-dependencies
```

```
sudo gem install mongrel_cluster --include-dependencies
```

7. ติดตั้ง subversion

```
sudo apt-get install subversion
```

8. ติดตั้ง apache

```
apt-get install apache2 apache2.2-common apache2-mpm-prefork apache2-utils
```

```
mkdir /var/www/html
```

```
cd /var/www/html
```

9. ทำการสำเนาโปรแกรมประยุกต์มายังไดเรกทอรี

```
cp -R library /var/www/html
```

```
sudo chown -R www-data:www-data library
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สร้าง database เข้าไปด้วยคำสั่ง

```
mysql -u root -p
```

ได้ prompt เป็น mysql แล้วป้อนข้อมูลว่า

```
mysql> create database library_production;
```

```
mysql> quit
```

10. ต่อไปเป็นการปรับแต่งให้ mongrel ทำงานเร็วขึ้น โดยการให้ apache ส่งผ่านคำสั่งไปยัง mongrel ให้สร้างเพิ่ม /etc/apache2/sites-available/library

```
<VirtualHost *:80>
```

```
DocumentRoot /var/www/html/library/public
```

```
<Directory "/var/www/html/library/public">
```

```
Options FollowSymLinks
```

```
AllowOverride None
```

```
Order allow,deny
```

```
Allow from all
```

```
</Directory>
```

```
RewriteEngine On
```

```
RewriteCond %{DOCUMENT_ROOT}%{REQUEST_FILENAME} !-f
```

```
RewriteRule ^/(.*)$ balancer://mongrel_cluster%{REQUEST_URI} [P,QSA,L]
```

```
<Proxy *>
```

```
Order allow,deny
```

```
Allow from all
```

```
</Proxy>
```

```
<Proxy balancer://mongrel_cluster>
```

```
BalancerMember http://localhost:8000
```

```
BalancerMember http://localhost:8001
```

```
BalancerMember http://localhost:8002
```

```
</Proxy>
```

```
</VirtualHost>
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้วใช้คำสั่งต่างๆ ดังนี้

```
a2enmod rewrite
a2enmod proxy_http
a2enmod proxy_balancer
/etc/init.d/apache2 force-reload
sudo a2dissite 000-default
sudo sudo a2ensite library
/etc/init.d/apache2 force-reload
sudo mongrel_rails cluster::configure
```

11. แก้ไขเพิ่ม `/var/www/html/library/config/mongrel_cluster.yml` มีข้อความดังนี้

```
user: www-data
cwd: /var/www/html/library
port: "8000"
environment: production
group: www-data
address: localhost
pid_file: log/mongrel.pid
servers: 3
```

12. สั่งให้ mongrel ทำงานด้วยคำสั่ง

```
cd /var/www/html/library
mongrel_rails cluster::start
```

ทดสอบผลงาน โดยทำงานบน server ตัวนี้ เปิดเว็บ `http://localhost` จะต้องได้หน้าโฮมเพจเสร็จแล้วให้ยกเลิกการทำงาน ด้วยการใส่คำสั่งดังนี้

```
cd /var/www/html/library
mongrel_rails cluster::stop
```

13. ถึงตอนนี้ทุกอย่างพร้อมใช้แล้ว แต่ยังใช้ได้เฉพาะ localhost ในการใช้งานจริงที่เป็นชื่อหรือหมายเลขไอพี ให้ทำการแก้ไขเพิ่มข้อมูลที่เคยแก้ไขมา จาก localhost เป็นชื่อ เช่น `yourhost.com` เพิ่มที่ต้องแก้ไขคือ

```
/etc/apache2/sites-available/library และ
/var/www/html/library/config/mongrel_cluster.yml
```

14. สั่งให้ทำงานทุกครั้งที่บูทเครื่องจะต้องสร้างเพิ่มสั่งทำงานขึ้นมาก่อนชื่อ `/root/mysite-start.sh` มีข้อความว่า

เอกสารนี้เป็นเอกสารที่ให้บริการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
#!/bin/sh
```

```
cd /var/www/html/library
```

```
mongrel_rails cluster::start
```

15. แก้ไขเพิ่มชื่อ /etc/rc.local

เพิ่มบรรทัดข้อความว่า /bin/sh /root/mysite-start.sh

ไว้เป็นบรรทัดสุดท้าย



ภาคผนวก ข

คู่มือการใช้งานระบบ

การใช้งานโปรแกรมจัดการฐานข้อมูลไมน์นิง

การใช้งาน โปรแกรมจัดการฐานข้อมูลไมน์นิงเพื่อสร้างและประมวลผลฐานข้อมูลไมน์นิง สามารถแบ่งการใช้งานได้ 2 ลักษณะ คือ การสร้างและประมวลผลฐานข้อมูลไมน์นิงใหม่ และการประมวลผลฐานข้อมูลไมน์นิงที่มีอยู่แล้ว

1. หน้าจอหลักหลังจากเรียกใช้โปรแกรม แสดงดังรูปที่ 1ข



รูปที่ 1ข หน้าจอหลัก

2. ถ้าต้องการสร้างฐานข้อมูลไมน์นิงขึ้นมาใหม่ ให้เลือก Create New Mining Project จะแสดงหน้าจอดังรูปที่ 2ข

Create New Mining Project

Analysis Services

Analysis Services Server: localhost

New Database: LibAnalysis

Existing Database: [Dropdown]

Data Source

Data Source Server: localhost

Initial Catalog: LibCir

Mining Model

Mining Structure Name: LibCir

Mining Model Name: LibCirAssoc

Minimum Probability: 0.5

Minimum Support: 5

Buttons: Create and Process New Database, Reprocess Database

Status: Ready

รูปที่ 2x หน้าจอในการสร้างฐานข้อมูลใหม่

โดยมีค่าพารามิเตอร์ต่างๆ ที่ต้องกำหนด ดังนี้

พารามิเตอร์	ความหมาย
Analysis Services	
Analysis Services Server	ชื่อเครื่องที่ SQL Server 2005 Analysis Services ทำงานอยู่
New Database	ชื่อฐานข้อมูลไมน์นิ่งที่จะสร้างขึ้นใหม่
Data Source	
Data Source Server	ชื่อเครื่องที่เก็บข้อมูลที่จะใช้ในการประมวลผล
Initial Catalog	ชื่อฐานข้อมูลที่จะใช้ในการประมวลผล
Mining Model	
Mining Structure Name	ชื่อของ Mining Structure
Mining Model Name	ชื่อของ Mining Model
Minimum Probability	ค่า Minimum Probability หรือค่า Confidence
Minimum Support	ค่า Minimum Support

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ห้ามการใช้นอกเหนือจากที่ระบุไว้ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ถ้าต้องการประมวลผลฐานข้อมูลไมน์นิ่งที่มีอยู่แล้ว ที่หน้าจอแสดงในรูปที่ 1x ให้เลือก Reprocess Mining Database จะแสดงดังรูปที่ 3x และในช่อง Existing Database เลือกฐานข้อมูลไมน์นิ่งที่ต้องการ ส่วนค่าอื่นๆ ก็มีความหมายเช่นเดียวกับ ข้อ 2 เมื่อกำหนดค่าต่างๆ เรียบร้อยแล้วก็คลิก Reprocess Database เพื่อทำการประมวลผล

The screenshot shows the 'Reprocess Mining Database' dialog box with the following configuration:

- Analysis Services:**
 - Analysis Services Server: localhost
 - New Database: (empty)
 - Existing Database: LibAnalysis
- Data Source:**
 - Data Source Server: localhost
 - Initial Catalog: LibCir
- Mining Model:**
 - Mining Structure Name: LibCir
 - Mining Model Name: LibCirAssoc
 - Minimum Probability: 0.5
 - Minimum Support: 5

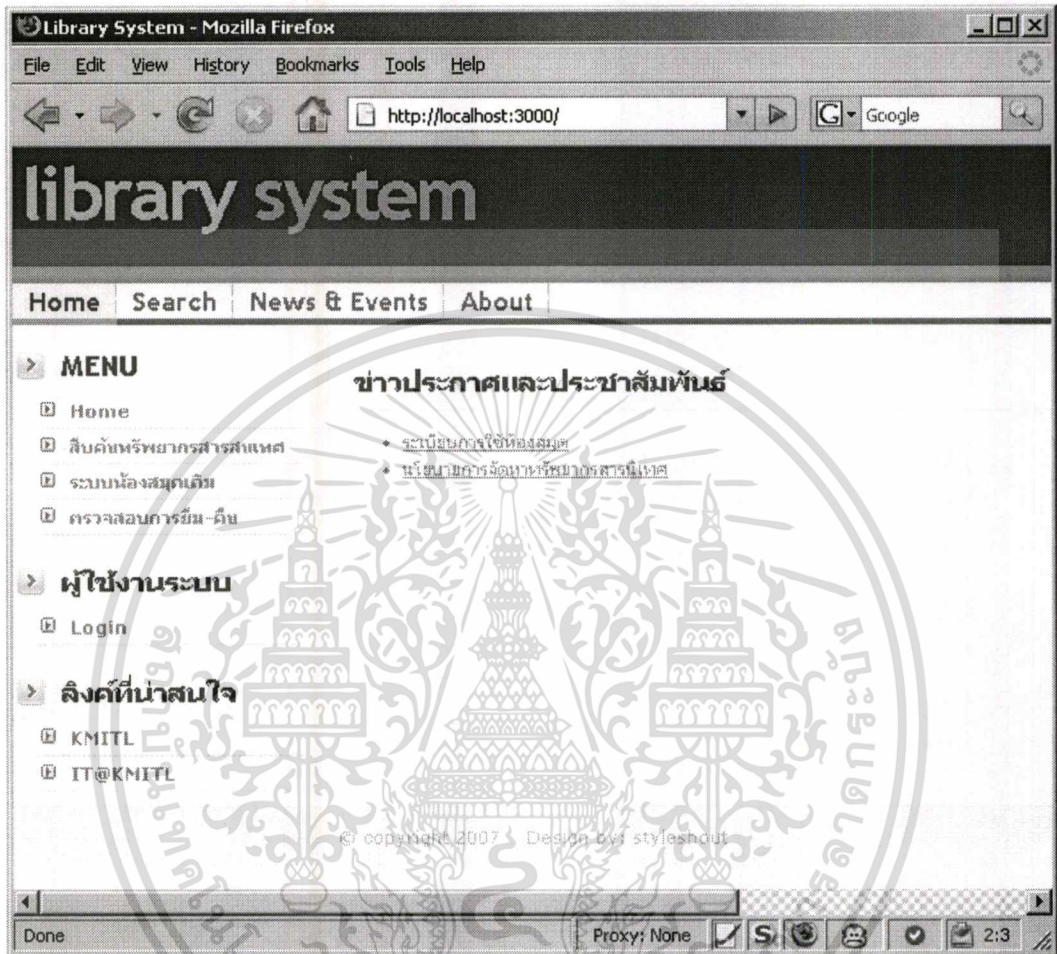
Buttons: 'Create and Process New Database' and 'Reprocess Database' (selected).

Status: Ready

รูปที่ 3x หน้าจอในการประมวลผลฐานข้อมูลที่มีอยู่แล้ว

การใช้งานระบบสืบค้นข้อมูลห้องสมุดผ่านเว็บ

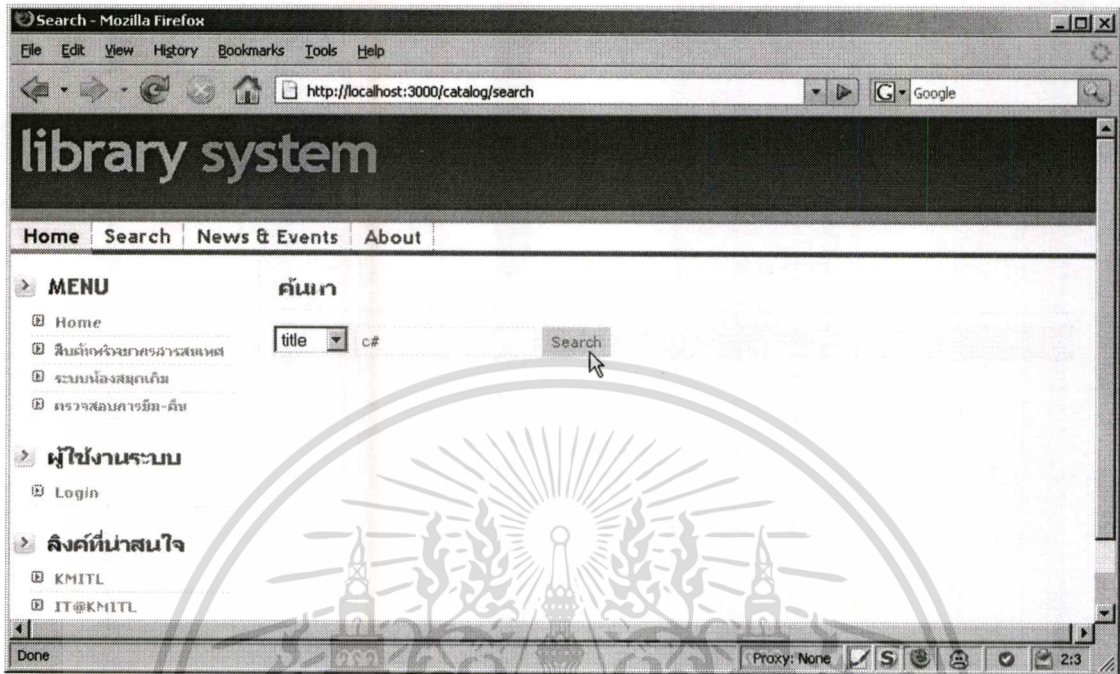
1. หน้าโฮมเพจจะมีลักษณะดังแสดงในรูปที่ 4ข



รูปที่ 4ข หน้าโฮมเพจของเว็บระบบห้องสมุด

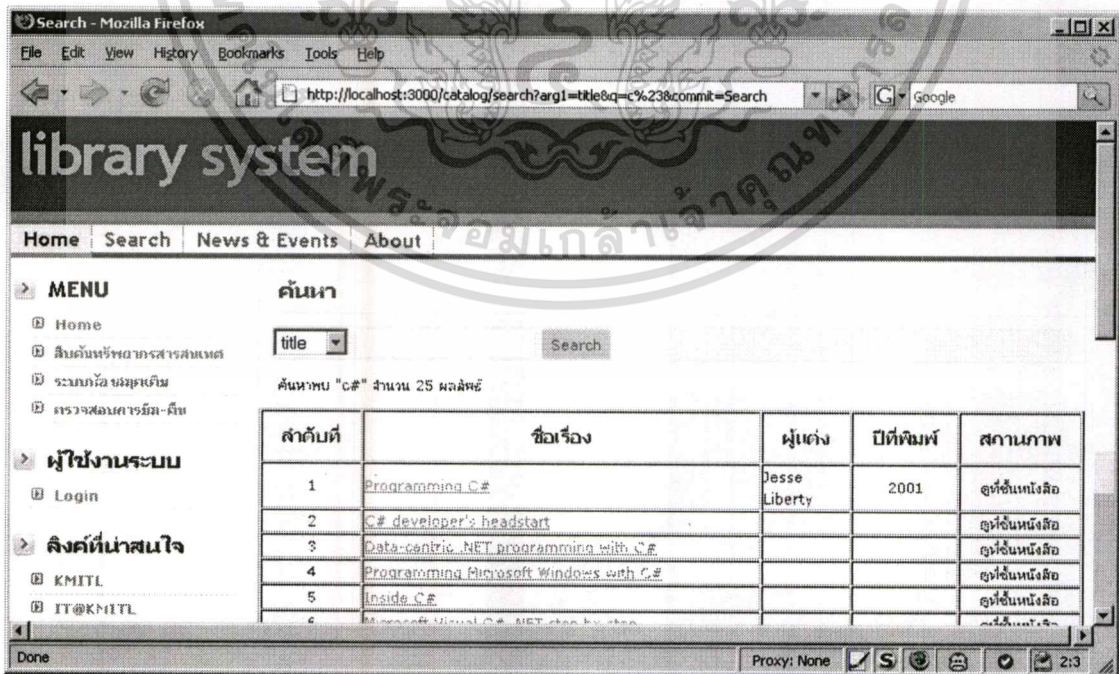
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ถ้าต้องการสืบค้นข้อมูลให้คลิกที่เมนู Search จะแสดงดังรูปที่ 5x ให้เลือกประเภทของการค้นหา และป้อนข้อมูลที่ต้องการค้นหา แล้วคลิก Search



รูปที่ 5x ส่วนของการสืบค้นข้อมูล

ระบบจะแสดงผลพืชรจากการค้นหา ดังแสดงในรูปที่ 6x



รูปที่ 6x หน้าแสดงผลพืชรจากการสืบค้นข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

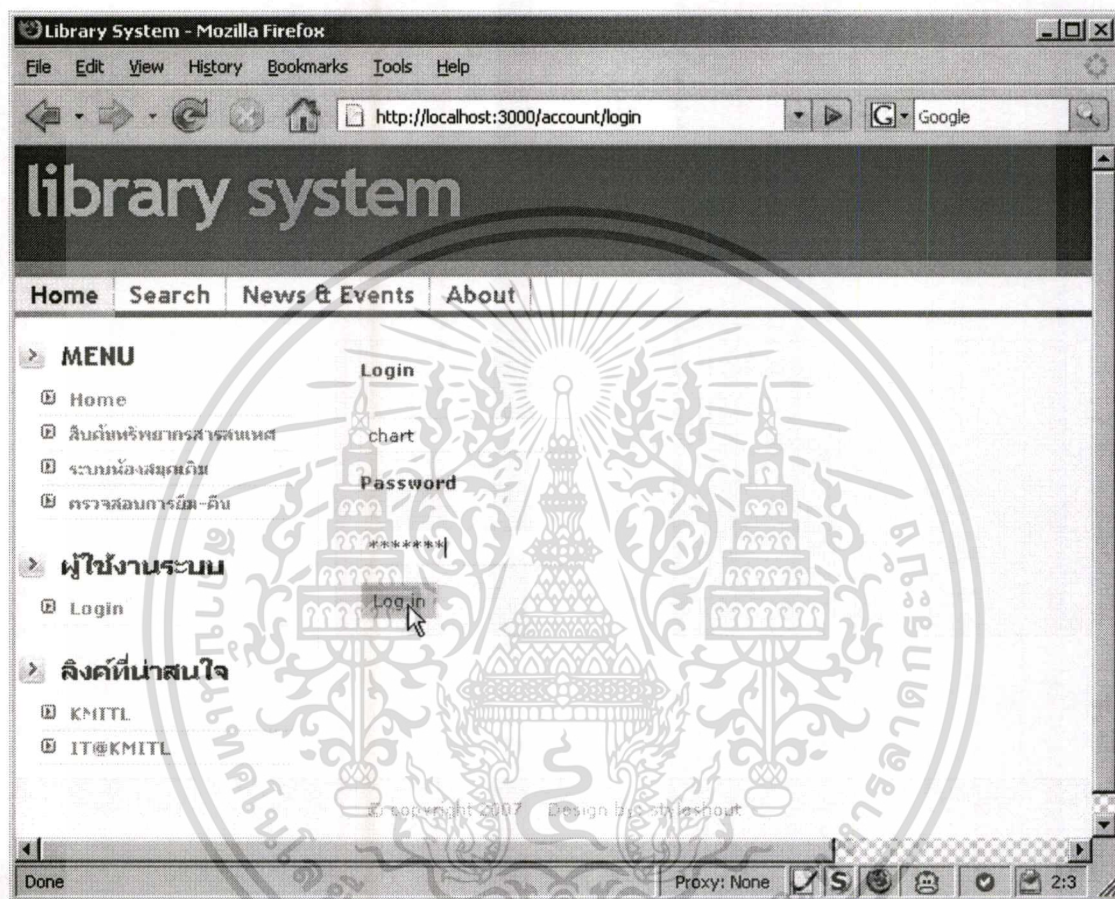
3. ถ้าต้องการดูรายละเอียดของทรัพยากรห้องสมุด ให้คลิกที่ลิงค์ของชื่อทรัพยากรห้องสมุด ระบบจะแสดงรายละเอียดของทรัพยากรห้องสมุด, ระดับคะแนนเฉลี่ย และความคิดเห็นของผู้ใช้งาน พร้อมทั้งเรียกใช้บริการเว็บเซอร์วิสเพื่อทำนากฎความสัมพันธ์ และแสดงรายการของทรัพยากรห้องสมุดที่มีความสัมพันธ์กับทรัพยากรห้องสมุดที่ผู้ใช้งานกำลังพิจารณาอยู่ ดังแสดงผลดังรูปที่ 7ข

The screenshot shows a Mozilla Firefox browser window displaying a web application titled "Programming C# - Mozilla Firefox". The address bar shows the URL "http://localhost:3000/catalog/show/2001015951". The page content includes a navigation menu on the left, a central book cover for "Programming C#" by Jesse Liberty, and a list of related books. The book cover features a stork carrying a bundle, with the text "Programming C#" and "O'REILLY". Below the cover, the book details are listed: "ชื่อเรื่อง : Programming C#", "ผู้แต่ง : Jesse Liberty", "ISBN : 0596001177", "ปีที่พิมพ์ : 2001", "สำนักพิมพ์ : O'reilly", "จำนวนหน้า : 680", and "Average User Rating : 4.5 /5.0 (แสดงความคิดเห็น)". A section titled "ผู้ที่ยืมหนังสือเล่มนี้มักจะยืมหนังสือ ดังต่อไปนี้ด้วย" lists several related books with their authors and publication years. At the bottom, there is a section for "ความคิดเห็นของผู้ใช้งาน" (User Reviews) showing a review by "suchart" on "Mon Sep 17 14:19:41 +0700 2007" with a rating of 4 and the text "ความคิดเห็น : ดีครับ แนะนำครับ".

รูปที่ 7ข หน้าแสดงรายละเอียดของทรัพยากรห้องสมุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

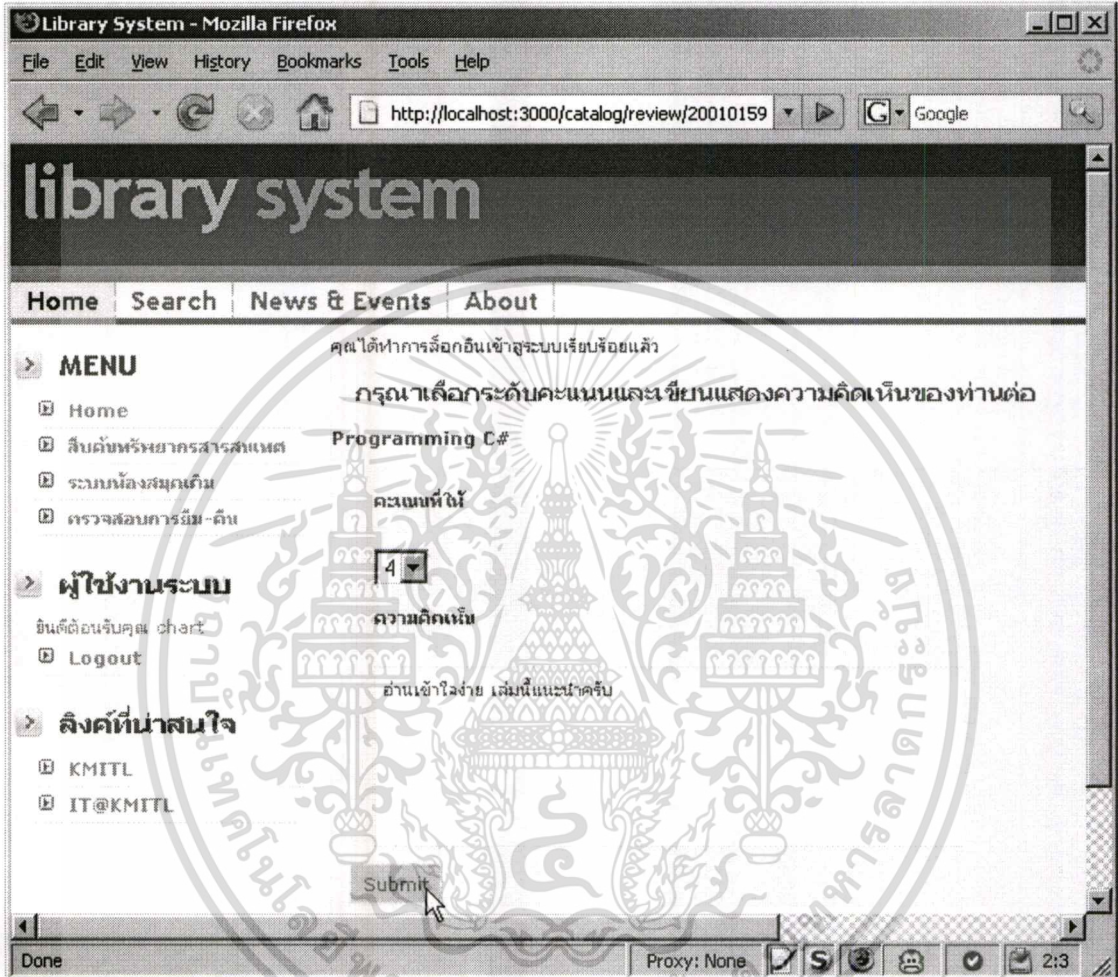
4. ถ้าต้องการเขียนแสดงความคิดเห็นและกำหนดระดับคะแนนให้กับ
ทรัพยากรห้องสมุด จากรูปที่ 7ฯ ให้คลิกที่ลิงค์ “แสดงความคิดเห็น” ถ้าผู้ใช้งานยัง
ไม่ได้ล็อกอินเข้าสู่ระบบ ระบบจะแสดงหน้าจอส่วนของการล็อกอินเข้าสู่ระบบ ดัง
แสดงในรูปที่ 8ฯ ให้ป้อนค่าชื่อผู้ใช้งานและรหัสผ่าน แล้วคลิก Log in



รูปที่ 8ฯ หน้าแสดงส่วนของการล็อกอินเข้าสู่ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. เลือกระดับคะแนนความพึงพอใจของผู้ใช้งานต่อทรัพยากรห้องสมุดจากระดับ 1 ถึง 5 (มีความพอใจน้อยที่สุดไปยังมากที่สุด) และเขียนแสดงความคิดเห็น ดังแสดงในรูปที่ 9๗ แล้วคลิก Submit



รูปที่ 9๗ ส่วนของการเขียนแสดงความคิดเห็นของผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ นายสุชาติ แก้วยศ
 วันเดือนปีเกิด 19 ธันวาคม พ.ศ. 2519
 ประวัติการศึกษา
 มัธยมศึกษา โรงเรียนสุรวิทยาคาร จังหวัดสุรินทร์
 ปริญญาตรี จุฬาลงกรณ์มหาวิทยาลัย คณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรม
 สิ่งแวดล้อม ปีการศึกษา 2542
 ประวัติการทำงาน
 พ.ศ. 2550 ผู้ดูแลระบบคอมพิวเตอร์
 อีเมล suchart_ky@yahoo.com



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้