ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

ระบบวิเคราะห์ความสัมพันธ์ของข้อมูลหลายระดับ

Multi Level Association Rule Discovery System

โดย

เกศินี บุญช่วย

รหัส 45061604

อาจารย์ที่ปรึกษา

ผศ. ดร. วรพจน์ กรีสุระเดช

b11700944
i12sm662x

วัน เดือน ปี 15 ก.พ. 2550
เลขทะเบียน 02254
เลขเรียกหนังสือ วพ ก ๓๓๔ร ๒๕๔๗
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

H002254

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ภาคเรียนที่ 2 ปีการศึกษา 2547
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	ระบบวิเคราะห์ความสัมพันธ์ของข้อมูลหลายระดับ
นักศึกษา	นางสาว เกศินี บุญช่วย
อาจารย์ที่ปรึกษา	ผศ. ดร.วรพจน์ กรีสุระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2547

บทคัดย่อ

ในการวิเคราะห์ข้อมูลโดยการทำดาต้าไมนิ่งในหลายรูปแบบ ช่วยให้ผู้วิเคราะห์ได้ข้อมูลซึ่งเป็นประโยชน์ และสามารถนำไปใช้งานได้ ไม่ว่าเพื่อทำการคาดการณ์เหตุการณ์ต่างๆ และนำมาวางแผนในการจัดการ หรือการอธิบายลักษณะของสิ่งใดสิ่งหนึ่ง และวางแผนกลยุทธ์ต่างๆเพื่อทำให้เกิดผลประโยชน์จากลักษณะของสิ่งนั้น

การค้นหาความสัมพันธ์ของข้อมูลเป็นที่รู้จักกันดีในรูปแบบหนึ่งของการทำดาต้าไมนิ่ง ผู้วิเคราะห์สามารถใช้ประโยชน์จากความสัมพันธ์ที่ได้จากการวิเคราะห์ เช่น ในการนำไปใช้ด้านการตลาดเพื่อวิเคราะห์พฤติกรรมของผู้บริโภค เพื่อวางแผนการตลาดให้เหมาะสมกับพฤติกรรมผู้บริโภค และเพื่อให้การวิเคราะห์ข้อมูลโดยใช้ดาต้าไมนิ่งได้ข้อมูลที่มีประโยชน์ต่อผู้วิเคราะห์มากยิ่งขึ้น การค้นหาความสัมพันธ์จึงมีการพัฒนาต่อไปเพื่อให้สามารถหาความสัมพันธ์ในข้อมูลที่มีการจัดกลุ่มเป็นระดับชั้น และมองหาความสัมพันธ์ในแต่ละระดับชั้นได้

การค้นหาความสัมพันธ์ของมูล ซึ่งมีลักษณะความสัมพันธ์ของข้อมูลที่เป็นระดับชั้น เกิดจากการจัดกลุ่มของข้อมูลตามคุณสมบัติต่างในข้อมูลนั้นๆ ทำให้ได้มาซึ่งข้อมูลที่มีการจัดหมวดหมู่เป็นระดับชั้นตามขอบเขตการวิเคราะห์ที่กำหนดไว้โดยผู้วิเคราะห์ และสามารถวิเคราะห์ความสัมพันธ์ในแต่ละชั้นของข้อมูล หรือแม้กระทั้งความสัมพันธ์ในระหว่างระดับชั้นของข้อมูลได้ โดยมีพื้นฐานมาจากการทำดาต้าไมนิ่งในรูปแบบของ Association Rule และตลอดจนการนำอัลกอริทึม Apriori เข้ามาใช้ในการพัฒนา ในการทำดาต้าไมนิ่งในลักษณะดังกล่าวจะเป็นไปตามหลักการที่เรียกว่า Multiple Level Association Rules ซึ่งผู้ใช้งานจะได้รายละเอียดในการวิเคราะห์มากขึ้นจากเดิม และนำไปใช้ประโยชน์ได้มากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Multi Level Association Rule Discovery System
Student	Miss. Kesinee Boonchuay
Advisor	Asst. Prof. Dr. Worapoj Kreesuradej
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic Year	2004

ABSTRACT

Data analysis by using Data Mining in several types will assist the analyzer to get the beneficial and useful data such as event forecasting for management planning, explanation of the thing and strategic planning.

Searching for data relation is well known in a type of data mining. The analyzer can use the benefit of the relation from this analysis such as in term of marketing to analyze the customer behavior for marketing plan. To be more beneficial to the analyzer of using data mining, the Searching of relation has to be continuous development to search for data relation that has been hierarchies.

The searching of data relation that has data relation in term of Hierarchical is from grouping of data according to qualification. It is to get the data that has been arranged in type and level according the scope of analysis that has been set by the analyzer and analyze the relation of each step of the data even the relation between levels of each data by based from data mining of Association Rule and using of Apriori Algorithm.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงงานการพัฒนาระบบการทำดาต้าไมนิ่งในข้อมูลหลายระดับได้รับความสนับสนุนเป็นอย่างดีจากหลายๆฝ่าย ไม่ว่าจะเป็นในด้านการให้คำแนะนำ คำปรึกษา และกำลังใจ ในการศึกษาวิเคราะห์และออกแบบระบบงาน เพื่อให้เป็นไปตามวัตถุประสงค์ที่วางไว้ ผู้จัดทำจึงใคร่ขอขอบพระคุณบุคคลดังต่อไปนี้

1. ดร.วรพจน์ กรีสุระเดช
2. บิดา มารดา และญาติพี่น้องที่ให้กำลังใจ
3. เพื่อนๆทุกคนที่คอยให้คำปรึกษา และกำลังใจในการทำงาน

น.ส. เกศินี บุญช่วย

17 กุมภาพันธ์ 2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญภาพ VIII

บทที่

1. บทนำ 1

1.1 ความเป็นมาของโครงการ 1

1.2 วัตถุประสงค์ของโครงการ 1

1.3 ขอบเขตของโครงการ 2

1.4 ขั้นตอนการศึกษา 2

2. Data mining with Association Rule 3

2.1 กระบวนการในการทำดาต้าไมนิ่ง 3

2.2 Association Rules 4

2.3 Apriori Algorithm 5

3. Multiple Level Association Rules 9

3.1 นิยาม Multiple Level Association Rules 11

3.2 ขั้นตอนการทำ Multiple Level Association Rules 12

3.3 อัลกอริทึม ML-T2 16

3.4 ดาต้าไมนิ่งในโครงสร้างผสม 17

3.5 ดาต้าไมนิ่งระหว่างระดับชั้น 18

4. วิเคราะห์และออกแบบระบบ 20

4.1 ฐานข้อมูลในระบบ 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

5. การพัฒนาระบบงาน และหน้าจอการทำงาน 26
5.1 สภาพแวดล้อมการพัฒนาระบบ 26
5.2 การจัดเตรียมสภาพแวดล้อมในการพัฒนาระบบ 26
5.3 ขั้นตอนการทำงานในการทำไมนิ่ง 27
5.4 การพัฒนาระบบและหน้าจอการทำงาน 30
6. สรุปการพัฒนาระบบ 51
6.1 ผลการวิเคราะห์และออกแบบระบบการทำคาต้าไมนิ่งแบบ Multiple Level Association Rule 51
6.2 ประโยชน์ที่คาดว่าจะได้รับ 51
6.3 แนวทางการพัฒนาระบบเพิ่มเติม 52
บรรณานุกรม 53
ประวัติผู้เขียน 54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

2.1 รายการสินค้าที่มีการซื้อจากห้างสรรพสินค้าแห่งหนึ่ง 4

2.2 ฐานข้อมูล 6

2.3 สร้างสมาชิก 6

2.4 นับจำนวนสมาชิก 6

3.1 ข้อมูลรายการสินค้า 12

3.2 ข้อมูลการขายสินค้า 12

3.3 ข้อมูลที่ถูกสร้างใหม่จากการจัดกลุ่ม 13

3.4 Encoded transaction Table 14

3.5 การหา Large Itemsets ที่ Level 1 14

3.6 การหา Large Itemsets ที่ Level 2 15

3.7 การหา Large Itemsets ที่ Level 3 16

3.8 การหาทำไมนิ่งในโครงสร้างแบบผสม 18

3.9 การทำไมนิ่งระหว่างระดับชั้นใน Level ที่ 2 19

3.10 การทำไมนิ่งระหว่างระดับชั้นใน Level ที่ 3 19

4.1 ข้อมูลรายละเอียดของงาน 21

4.2 ความสัมพันธ์และการเรียกใช้ข้อมูลในการไมนิ่ง 22

4.3 ความสัมพันธ์และการเรียกใช้ข้อมูลในการไมนิ่ง 23

4.4 ข้อมูลของ Large Itemsets 24

4.5 ข้อมูลของกฎที่มีความน่าสนใจ 25

5.1 EncodeData 28

5.2 TransactionData 28

5.3 Transactions 28

5.4 LargeItemData 29

5.5 RuleData 30

5.6 โครงสร้างของเอกสาร XML ที่ใช้เป็น Transaction 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.7 ตัวอย่างของเอกสาร XML ที่ใช้เป็น Transaction 46
5.8 ตัวอย่างของเอกสาร XML ที่ใช้เป็น Item List 47
5.9 รูปแบบเอกสาร CSV สำหรับ Transaction 47
5.10 รูปแบบเอกสาร CSV สำหรับ Item List 58
5.11 ตัวอย่างของเอกสาร CSV ที่ใช้เป็น Transaction 49
5.12 ตัวอย่างของเอกสาร CSV ที่ใช้เป็น Item List 50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้า

รูปที่

2.1 กระบวนการทำดาต้าไมนิ่ง 3
3.1 แสดงระดับชั้นของการจัดกลุ่ม 13
4.1 แบบจำลองความสัมพันธ์ของฐานข้อมูล 20
5.1 โครงสร้างข้อมูลในระหว่างทำงาน 27
5.2 หน้าหลักของระบบ 31
5.3 หน้าจอแสดงการสร้างงานใหม่จากฐานข้อมูล SQL 33
5.4 หน้าจอสร้างการติดต่อไปยังฐานข้อมูล SQL หรือ Connection String 34
5.5 การกำหนดส่วนของ Transaction 35
5.6 การสร้าง Query String 37
5.7 การกำหนดส่วนของ Item 38
5.8 การกำหนดส่วนของ Mapping และ Level 39
5.9 การกำหนดค่า Minimum Support และ Minimum Confidence 41
5.10 แสดงผลการทำไมนิ่ง 42
5.11 Graph View 43
5.12 ข้อมูลประกอบกราฟ 43
5.13 การสร้างงานใหม่จากเอกสาร XML 44
5.14 การสร้างงานใหม่จากเอกสาร CSV 48

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของโครงการ

ในปัจจุบันมีการนำระบบสารสนเทศเข้ามาช่วยในการจัดการข้อมูลมากขึ้น ทำให้ข้อมูลต่างๆ ถูกรวบรวม และจัดเก็บอย่างมีระบบมากกว่าในอดีต ระบบสารสนเทศไม่ได้เพียงแต่มีบทบาทในการดูแล หรือจัดการในการจัดเก็บข้อมูลเท่านั้น การนำข้อมูลที่มีอยู่มาใช้ให้เกิดประโยชน์ ก็เป็นสิ่งสำคัญ ในการนำข้อมูลที่มีอยู่มาวิเคราะห์อย่างจริงจัง จะทำให้ผู้วิเคราะห์ได้รับรู้ถึงข้อมูลที่ซ้อนเร้นอยู่ในข้อมูลเหล่านั้น หรือเรียกว่าการเกิดของข้อมูลที่มีรูปแบบ การค้นหารูปแบบเหล่านี้จะทำให้ผู้วิเคราะห์ทราบถึงข้อมูลซึ่งสามารถนำมาใช้ประโยชน์ได้มากมาย ซึ่งทำให้เกิดการพัฒนาในการค้นหารูปแบบของข้อมูลซึ่งสามารถนำมาใช้ประโยชน์ได้ การทำดาต้าไมนิ่งก็เป็นวิธีการหนึ่งในการค้นหาข้อมูลที่ซ้อนอยู่ หรือค้นหารูปแบบการเกิดของข้อมูลนั้นๆ

การทำดาต้าไมนิ่งมีวิธีการต่างๆมากมายที่ถูกพัฒนาขึ้นมาเพื่อให้เหมาะกับรูปแบบ และความต้องการของผู้วิเคราะห์ รูปแบบหนึ่งซึ่งเป็นที่นิยมนั้นคือการหาความสัมพันธ์ของข้อมูล หรือที่เรียกกันว่า Association Rule เป็นการหารูปแบบความสัมพันธ์ระหว่างกันของสิ่งๆใดสิ่งหนึ่งต่ออีกสิ่งหนึ่ง เช่น ความสัมพันธ์ในการซื้อสินค้าชนิดหนึ่ง ควบคู่กับการซื้อสินค้าอีกชนิดหนึ่ง การวิเคราะห์ความสัมพันธ์ลักษณะนี้ ทำให้ผู้วางแผนการตลาดสามารถนำผลการวิเคราะห์นำไปใช้งานในการวางแผนการตลาดสำหรับสินค้า การจัดแผนการสนับสนุนการตลาดให้เหมาะสมได้

จากการวิเคราะห์ความสัมพันธ์ของข้อมูล เพื่อให้ผู้วิเคราะห์สามารถวิเคราะห์ข้อมูลได้มากขึ้น สามารถนำผลการวิเคราะห์ไปใช้ประโยชน์เพิ่มมากขึ้น การนำหลักการของ Multiple Level Association Rule มาใช้ในการหาความสัมพันธ์เป็นอีกวิธีการหนึ่ง ที่ทำให้ผู้วิเคราะห์ในรับข้อมูลจากการวิเคราะห์มากขึ้น และนำผลการวิเคราะห์ไปใช้ประโยชน์ได้มากขึ้นจากเดิม

1.2 วัตถุประสงค์ของโครงการ

1. เพื่อศึกษาขั้นตอน และวิธีการทำดาต้าไมนิ่ง ในแบบ Multiple Level Association Rules
2. เพื่อเรียนรู้ ในการนำผลลัพธ์จากการไมนิ่งโดยใช้ Multiple Level Association Rules ไปใช้งานได้อย่างเหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เพื่อมองหาแนวทางในการพัฒนา Multiple Level Association Rules ที่เหมาะสมในอนาคตต่อไป

1.3 ขอบเขตของโครงการ

โครงงานนี้มุ่งพัฒนาระบบที่ทำการวิเคราะห์โดยใช้ Multiple Level Association Rule มีขอบเขตดังนี้

1. สามารถพัฒนาโปรแกรม ตามทฤษฎี และอัลกอริทึม ได้ถูกต้อง
2. สามารถใช้กับข้อมูลทั่วไปได้
3. สามารถนำเสนอผลลัพธ์ให้กับผู้ใช้งานนำไปใช้ประโยชน์ได้อย่างถูกต้อง

1.4 ขั้นตอนการศึกษา

1. กำหนดหัวข้อ เป้าหมาย และวัตถุประสงค์ ตลอดจนขอบเขตของโครงงาน
2. ศึกษาทฤษฎี บทความ งานวิจัย และหนังสือที่เกี่ยวข้องกับหลักการ Multiple Level Association Rules
3. ศึกษาเทคนิคต่างๆในการพัฒนาโปรแกรม
4. พัฒนาโปรแกรมต่างที่ได้ศึกษาข้อมูลมา
5. ทำการทดสอบโปรแกรม
6. ปรับปรุงและแก้ไขข้อผิดพลาด
7. จัดทำเอกสารประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

Data Mining with Association Rules

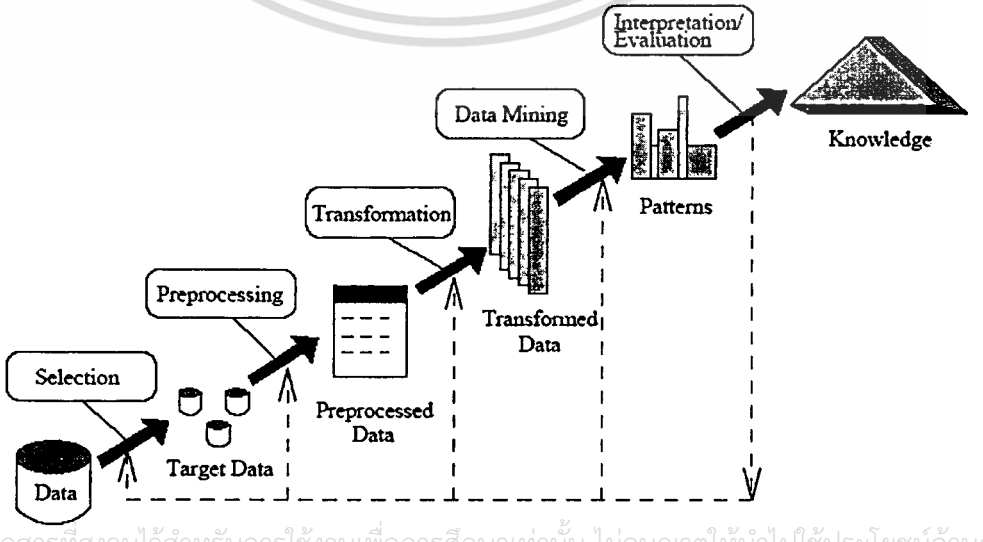
การทำดาต้าไมนิ่ง หรือในบางครั้งเรียกว่า Knowledge Discovery in Database (KDD) คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อหารูปแบบ และความสัมพันธ์ในข้อมูลนั้นๆ ซึ่งสามารถนำมาใช้ประโยชน์จากรูปแบบหรือความสัมพันธ์ดังกล่าวได้

เป้าหมายในการทำดาต้าไมนิ่งมีเป้าหมายหลักใหญ่ๆ 2 ข้อ คือ

1. การทำดาต้าไมนิ่งเพื่อการพยากรณ์ (Predictive data mining)
2. การทำดาต้าไมนิ่งเพื่ออธิบายข้อมูลที่มีอยู่ (Descriptive data mining) เป็นการหาความสัมพันธ์ หรือการจัดหากลุ่มของข้อมูลซึ่งไม่ใช่ลักษณะของการทำนาย

ตัวอย่างการนำไปใช้วิเคราะห์ด้านการตลาด เช่น การวิเคราะห์ข้อมูลลูกค้า การวิเคราะห์ผลกระทบด้านการตลาด การวิเคราะห์แนวโน้มตลาด การใช้ดาต้าไมนิ่งเข้ามาช่วยในการวิเคราะห์เพื่อให้สามารถทำการวิเคราะห์ข้อมูลซึ่งมีขนาดใหญ่ได้อย่างมีประสิทธิภาพโดยเลือกใช้รูปแบบการวิเคราะห์ที่เหมาะสม ซึ่งการวิเคราะห์เหล่านี้ ช่วยให้ผู้ดูแลด้านการตลาดสามารถนำข้อมูลที่ได้จากการวิเคราะห์มาใช้ในการวางแผนการตลาดต่อไปได้

2.1. กระบวนการในการทำดาต้าไมนิ่ง



รูปที่ 2.1 กระบวนการทำดาต้าไมนิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการในการทำดาต้าไมนิ่งเริ่มจากผู้วิเคราะห์กำหนดวัตถุประสงค์ในการทำดาต้าไมนิ่งว่าผู้ทำการวิเคราะห์มีความต้องการทราบอะไรจากข้อมูลที่มีอยู่ แล้วจึงเข้ากระบวนการดังนี้

- Selection ทำการเลือกข้อมูลจากฐานข้อมูลที่มีอยู่ ซึ่งเลือกข้อมูลที่เกี่ยงข้องกับสิ่งที่ต้องการวิเคราะห์มา
- Preprocessing ทำการพิจารณาข้อมูลว่าถูกต้องหรือไม่โดยดูจากความเป็นไปได้ของข้อมูล
- Transformation ทำการแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมทำหรับการทำดาต้าไมนิ่ง
- Data Mining ทำการไมนิ่งข้อมูล
- Interpretation Evaluation หลังจากทำการไมนิ่งแล้ว ผู้วิเคราะห์จะต้องทำการพิจารณารูปแบบที่ได้จากการไมนิ่ง พิจารณาความเป็นไปได้ของการเกิดรูปแบบดังกล่าว เพื่อให้รูปแบบที่ได้เป็นรูปแบบที่สามารถนำไปใช้ประโยชน์ได้ รูปแบบบางอย่างที่ได้จากการทำดาต้าไมนิ่งอาจจะนำไปใช้ประโยชน์ไม่ได้ ซึ่งอาจมาจากปัจจัยอย่างอื่น ซึ่งผู้วิเคราะห์จะต้องทำการพิจารณาก่อนนำไปใช้งาน

2.2 Association Rules

การค้นหาความสัมพันธ์ของข้อมูล สามารถนำไปช่วยในการตัดสินใจในธุรกิจได้ เช่นการค้นหาความสัมพันธ์ของการซื้อสินค้าของผู้บริโภค (Market Basket Analysis) สามารถนำผลการวิเคราะห์มาใช้ในการวางแผนการตลาด หรือการกำหนดกลยุทธ์ทางด้านการขาย ซึ่งการจัดเก็บข้อมูลการซื้อสินค้าของผู้บริโภคจะถูกจัดเก็บลงบนฐานข้อมูลขนาดใหญ่ ผู้วิเคราะห์จะทำการวิเคราะห์หาความสัมพันธ์จากข้อมูลที่ถูกจัดเก็บไว้ โดยยกตัวอย่างจากข้อมูลของลูกค้าที่ซื้อสินค้าในห้างสรรพสินค้าแห่งหนึ่งดังนี้

ตารางที่ 2.1 รายการสินค้าที่มีการซื้อจากห้างสรรพสินค้าแห่งหนึ่ง

TID	Items
001	ขนมปัง, นม, น้ำปลา
002	ขนมปัง, นม, ยาสีฟัน
003	ยาสีฟัน, น้ำปลา, นม
004	ขนมปัง, นม, แยม
005	บะหมี่, ข้าวสาร, ขนมปัง, นม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากข้อมูลการซื้อสินค้าในตารางที่ 2.1 เราสามารถนำมาหาความสัมพันธ์แบบ X => Y หรือความสัมพันธ์ของการซื้อสินค้า X แล้วจะซื้อสินค้า Y ด้วย โดยให้ X และ Y แทนด้วยสินค้าต่างชนิดกันได้ว่า หากต้องการทราบความสัมพันธ์ของลูกค้าที่ซื้อขนมปังแล้วจะซื้อนมด้วยมีความสัมพันธ์กันอย่างไร เพื่อทางห้างสรรพสินค้าจะนำผลการวิเคราะห์ที่ได้นำไปวางแผนการสนับสนุนทางการตลาดให้กับลูกค้าโดยนำขนมปังและนมมาขายควบคู่กันในราคาพิเศษ เพื่อเพิ่มยอดขายสินค้าให้กับทางห้างสรรพสินค้าต่อไปได้

ในการหาความสัมพันธ์เพื่อสร้างกฎความสัมพันธ์ที่เรียกว่า Association Rules นั้นจะเป็นการหาความสัมพันธ์บนพื้นฐานของค่า Support และค่า Confidence โดยสามารถอธิบายได้ดังนี้

กำหนด Association Rule ในรูป X => Y

- X = กลุ่มของ Items
- Y = กลุ่มของ Items
- $X \subseteq I, Y \subseteq I$ และ $X \cap Y = \varnothing$
- C = ค่า Confidence
- S = ค่า Support แทนด้วยสัญลักษณ์ σ
- C% ได้มาจาก $P(X \cap Y) / P(X)$ หรือเมื่อ Transaction นั้น ประกอบด้วย X แล้วจะประกอบด้วย Y คิดเป็นเปอร์เซ็นต์มากกว่า C% แสดงว่ามีความน่าเชื่อถือที่ X มีความสัมพันธ์กับ Y
- S% ได้มาจาก $P(X \cup Y)$ หรือเมื่อ Transaction นั้นประกอบด้วย X หรือ Y มีค่าคิดเป็นเปอร์เซ็นต์มากกว่า S% แสดงว่าข้อมูลดังกล่าวมีค่าความถี่มากเพียงพอในการนำมาพิจารณาความสัมพันธ์

2.3 Apriori Algorithm

การแก้ปัญหา และ อัลกอริทึมที่รู้จักและมีการใช้งานกันอย่างแพร่หลายก็คือ Apriori Algorithm โดยขั้นตอนการทำงานจะมีสองขั้นตอนคือ เริ่มจากการสร้างสมาชิกจากข้อมูลที่มีอยู่และทำการนับจำนวนสมาชิกทั้งหมดเพื่อเลือกข้อมูลที่มีจำนวนที่นับได้สูงกว่าค่า Support เริ่มแรกของ Algorithm นี้ก็คือการนับ 1-itemsets ซึ่งก็คือ itemsets ซึ่งมาจากการนำ item แต่ละ item ที่ไม่ซ้ำกันมาสร้างเป็น itemsets และ หานำจำนวนการที่ Transaction ที่ประกอบไปด้วย Item ดังกล่าวที่มีค่า Support มากกว่าค่า Minimum Support ที่กำหนดไว้ และนำมาสร้างชุดข้อมูลใหม่ได้ว่า large 1-itemsets หลังจากนั้นเริ่มสร้าง 2-itemsets โดยนำ itemsets จาก large 1-itemsets มารวมกันเพื่อให้อยู่ในรูปของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$L_k * L_k = \{X \cup Y \text{ where } X, Y \in L_k, |X \cap Y| = k-1\}$$

เป็นการสร้าง candidate ซึ่งก็คือการคาดการณ์ว่า Itemsets ที่เกิดจากการรวมกันนี้เป็น Large Itemsets ที่ในแต่ละ set ประกอบด้วย 2 item แล้วทำกานับจำนวน 2-itemsets ทั้งหมดที่มีและทำการคำนวณหาว่า itemsets ใดที่เป็น large 2-itemsets โดยเทียบกับค่า minimum Support ดังเช่นที่ทำการหา large 1-itemsets หลังจากนั้นก็ทำการรวม large 2-itemsets โดยรูปแบบเดียวกับการรวม large 1-itemsets เพื่อให้เป็น candidate ของ 3-itemsets ทำการนับและคำนวณหา large 3–itemsets ซึ่งจะมีการทำงานเป็นอย่างนี้ไปเรื่อยๆจนกระทั้งไม่สามารถหาค่า large itemsets ได้แล้ว

กำหนดให้ L_k เป็นชุดของ large k-itemsets ยกตัวอย่างจากตาราง 2.2 โดยกำหนดให้ค่า support มีค่าเป็น 50% ทำการสร้างสมาชิกใน 1-itemsets ได้ดังตาราง 2.3 และนำจำนวนข้อมูลในตาราง 2.4 เทียบกับค่า Minimum Support และทำการตัดค่าที่ต่ำกว่าค่า Minimum Support ได้ดังตาราง 3.5 ซึ่งจะเห็นว่าค่า Minimum Support ที่ตั้งไว้คือ 50% แต่ D มีค่าเพียง 25% จึงไม่เป็นสมาชิกใน Large 1-itemsets

ตารางที่ 2.2 ฐานข้อมูล

TID	Items
001	A C D
002	B C E
003	A B C E
004	B E

ตาราง 2.3 สร้างสมาชิก

1-itemsets
A
B
C
D
E

ตาราง 2.4 นับจำนวนสมาชิก

1-Itemsets	Count	S[%]
A	2	50%
B	3	75%
C	3	75%
D	1	25%
E	3	75%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 2.5 เลือกจำนวนสมาชิกที่มีจำนวนการเกิดในข้อมูลมากกว่าค่า Support

Large 1-Itemsets	Count	S[%]
A	2	50%
B	3	75%
C	3	75%
E	3	75%

ทำลักษณะเดียวกันไปเรื่อยๆเพื่อหาค่า Large 2-itemsets และ Large 3–Itemsets ต่อไปซึ่งจะได้ดังต่อไปนี้

ตาราง 2.6 สร้างสมาชิก

2-itemsets
A, B
A, C
A, E
B, C
B, E
C, E

ตาราง 2.7 นับจำนวนสมาชิก

2-Itemsets	Count	S[%]
A, B	1	25%
A, C	2	50%
A, E	1	25%
B, C	2	50%
B,E	3	75%
C, E	2	50%

ตาราง 2.8 เลือกจำนวนสมาชิกที่มีจำนวนการเกิดในข้อมูลมากกว่าค่า Support

Large 2-Itemsets	Count	S[%]
A, C	2	50%
B, C	2	20%
B, E	3	75%
C, E	2	50%

ตาราง 2.9 สร้างสมาชิก

3-itemsets
B, C, E

ตาราง 2.10 นับจำนวนสมาชิก

1-Itemsets	Count	S[%]
B, C, E	2	50%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 2.11 เลือกจำนวนสมาชิกที่มีจำนวนการเกิดในข้อมูลมากกว่าค่า Support

Large 1-Itemsets	Count	S[%]
B, C, E	2	50%

C_k = สมาชิกของ k-itemsets

L_k = สมาชิกของ large k-itemsets หรือ set ของ item ใน k-itemsets ที่มีการปรากฏข้อมูลในฐานข้อมูลมากกว่าค่า Support

และสามารถอธิบายวิธีทำงานของ Apriori Algorithm ในการหา large itemsets ได้ดังนี้

```
Result: = ∅;
k:=1;
C1 = set of all 1-itemsets;
While Ck ≠ ∅ do
        Create a counter for each itemset in Ck;
        Forall transactions in database do
                Increments the counters of itemsets in Ck which occur in the
transaction;
        Lk := All candidates in Ck which exceed the support threshold;
        Result := Result ∪ Lk;
        Ck+1 := all k+1 itemsets which have all of their k- item subsets in Lk
        k=k+1;
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

Multiple Level Association Rules

จาก Section ที่ผ่านเราได้ทราบเกี่ยวกับ Association Rules แล้ว จะเห็นว่าในการศึกษาเกี่ยวกับดาต้าไมนิ่งที่ผ่านมาจะเน้นไปถึงการไมนิ่งในลักษณะที่เป็นข้อมูลที่อยู่ในรูปแบบระดับชั้นเดียว ซึ่งทำให้การทำไมนิ่งยังมีข้อจำกัดสำหรับผู้ใช้งานที่ต้องการดูข้อมูลที่มีรูปแบบของความสัมพันธ์ในลักษณะที่เป็นระดับชั้น ดังนั้นจึงมีการนำ Multiple Level Association Rules เข้ามาใช้เพื่อทำการไมนิ่งความสัมพันธ์ที่มีลักษณะเป็นระดับชั้น ทำให้ผู้ใช้งานได้ข้อมูลที่ตรงกับความสนใจมากขึ้น และสามารถค้นหาความสัมพันธ์ในรูปแบบต่างๆได้มากขึ้น เช่นจากเดิมเราสามารถหาความสัมพันธ์ของผู้ที่ซื้อ นม และขนมปัง ด้วยกัน โดยการใช้หลักการ Association Rules ที่กล่าวมาแล้ว แต่หากเราต้องการทราบเพิ่มเติมว่า ผู้ที่ซื้อ นมยี่ห้อ Foremost และขนมปังพร้อมกันเป็นเช่นไร เราจะไม่สามารถทราบได้ แต่หลักการของ Multiple Level Association Rules จะทำให้เราทราบถึงความสัมพันธ์ดังกล่าวได้ ซึ่งหลักการของ Multiple Level Association Rules ก็มาจากหลักการพื้นฐานของ Association Rules ซึ่งเป็นระดับชั้นเดียวนั้นเอง

ในการค้นหาข้อมูลที่เป็นลักษณะของระดับชั้นนั้น มีวิธีการอยู่ 3 รูปแบบ คือ

1. **Progressive deepening**
   จะเป็นรูปแบบการทำงานที่เริ่มต้นจากระดับชั้นสูงลงไปยังระดับชั้นล่าง ในระดับชั้นสูงจะเป็นข้อมูลที่มีความเป็น Generalization มาก เช่น นม จะเป็น Generalization ของ นมยี่ห้อ Foremost เป็นต้น
2. **Progressive generalization**
   จะเป็นรูปแบบที่ตรงกันข้ามกับ Progressive deepening โดยจะเริ่มการทำงานจากระดับชั้นล่างขึ้นมา
3. **Interactive up-and-down**
   จะเป็นรูปแบบที่สามารถทำงานจากระดับชั้นล่างขึ้นบน หรือบนลงล่าง สลับกันไปหรืออย่างไรก็ได้ขึ้นอยู่กับการใช้งานของผู้ใช้งาน

ซึ่ง Multiple Level Association Rules จะใช้วิธีการแบบ Progressive deepening คือการทำงานจากระดับชั้นสูงลงมาระดับชั้นล่าง ซึ่งในรูปแบบของ Multiple Level Association Rules จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกี่ยวข้องกับการจัดกลุ่มของข้อมูลเป็นระดับชั้น ซึ่งในการจัดกลุ่มที่เหมาะสมจะต้องเกี่ยวข้องกับหลักการ Generalization และ Specialization ดังนี้

- ในระดับชั้นที่สูงขึ้นไปจะต้องมีลักษณะ Generalization ของระดับชั้นล่าง หรืออธิบายได้ว่าในระดับชั้นบนจะมีความเจาะจงของข้อมูลน้อยกว่าในระดับชั้นล่าง เช่น นมยี่ห้อ Foremost, นมยี่ห้อโชคชัย จะทำให้เป็น Generalization ก็คือนม ซึ่งนมจะครอบคลุมถึงนมทุกยี่ห้อนั้นเอง นมจะกล่าวแทน นมในยี่ห้อต่างๆได้ทั้งหมด
- ในระดับชั้นล่างจะต้องมีลักษณะ Specialization มาจากข้อมูลที่เป็นระดับชั้นบน ในระดับชั้นล่างจะมีความเจาะจงของข้อมูลที่มากกว่า เช่น นมยี่ห้อ Foremost ในระดับชั้นล่างจะประกอบไปด้วย นมไขมันต่ำ, นมแคลเซียมสูง เหล่านี้เป็นต้น ซึ่งข้อมูลในระดับชั้นล่างจะเฉพาะเจาะจงลงไปมากขึ้นนั้นเอง

หลักการไมนิ่งแบบ Association Rules ซึ่งเราจะทำการค้นหาความสัมพันธ์ที่ตรงกับความสนใจ หรือความสัมพันธ์ที่มีความน่าเชื่อถือของรูปแบบของการเกิดความสัมพันธ์นั้นๆ โดยอาศัยหลักการกำหนดค่า Minimum Support และ Minimum Confidence ดังนั้นจากการทำไมนิ่งในรูปแบบเดิมพบปัญหาดังต่อไปนี้ในการทำไมนิ่งในแบบ Multiple Level Association Rules

- ความเป็นไปได้ที่จะพอข้อมูลที่อยู่ในระดับชั้นสูงจะมีมากกว่าในระดับชั้นล่าง เนื่องจากของมูลในระดับชั้นสูงจะมีความครอบคลุมข้อมูลที่กว้างมากกว่า มีความเจาะจงน้อยกว่า การเกิดของข้อมูลจึงมีมากกว่า หากเราทำการระบุค่า Minimum Support และ Minimum Confidence เท่ากันในทุกระดับชั้น จะทำให้ความสัมพันธ์ในระดับชั้นล่างมีค่าต่ำกว่า Minimum Support และ Minimum Confidence ได้ง่าย ทำให้ผู้วิเคราะห์ไม่สามารถหาความสัมพันธ์ใดๆในระดับชั้นล่างได้ และหากผู้วิเคราะห์กำหนดค่า Minimum Support และ Minimum Confidence น้อยเกินไปจะทำให้ในระดับชั้นสูงเราจะได้ข้อมูลซึ่งไม่ตรงกับความต้องการ หรือข้อมูลที่ไม่ได้มีลักษณะเป็นรูปแบบชัดเจน ดังนั้นในการกำหนดค่าของ Minimum Support และ Minimum Confidence ในแต่ละระดับชั้นควรมีความแตกต่างกัน โดยระดับชั้นสูงควรมีค่ามากและลดลงมาในระดับชั้นล่างๆ
- ในการจัดกลุ่มขึ้นเป็นลำดับขั้นในระดับสูงซึ่งจะมีขอบเขตของข้อมูลที่กว้าง จะทำให้การทำไมนิ่งเกิดข้อมูลซึ่งไม่สามารถนำไปใช้ประโยชน์ได้ง่ายเช่น ข้อมูลที่เป็นลักษณะของรูปแบบทั่วไป ซึ่งเป็นที่รู้กันได้โดยไม่จำเป็นต้องทำการไมนิ่งข้อมูลออกมาก็ได้ เช่น การซื้อขนมปัง และนม ด้วยกัน หรือการซื้อยางรถ และใช้บริการของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ร้านซ่อมรถไปด้วยกัน เป็นลักษณะปกติซึ่งเกิดขึ้นอยู่แล้ว เพราะคนทั่วไปมักจะรับประทานขนมปังกับนม หรือเมื่อซื้อยางรถ ก็ต้องใช้บริการของร้านซ่อมรถเพื่อให้เปลี่ยนยางใหม่อยู่แล้ว ดังนั้นความสัมพันธ์ในระดับชั้นที่สูงมีโอกาสมากที่จะเกิดข้อมูลในลักษณะดังกล่าว ดังนั้นในการจัดกลุ่มเพื่อกำหนดข้อมูลในระดับชั้นสูงจะต้องมีการใช้มนุษย์เข้ามาช่วยในการพิจารณา หรือผู้เชี่ยวชาญในการวิเคราะห์ข้อมูลเพื่อทำให้การจัดกลุ่มได้ข้อมูลซึ่งมีประโยชน์เหมาะสมกับการนำมาใช้งานได้

3.1 นิยาม Multiple Level Association Rules

ในการนิยามการไมนิ่ง Multiple Level Association Rules โดยการกำหนดตัวแปรต่อไปนี้

T คือชุดข้อมูลทั้งหมด

T_i (Transaction) แทนชุดข้อมูลหนึ่งใน T ที่ระบุชุดข้อมูลด้วยตัวแปร i

I แทน Item ทั้งหมด

A_i เป็นสมาชิกของ I หรือ ของ Item

B_i เป็นสมาชิกของ I หรือ ของ Item

3.1.1 σ (A / S) คือการหาค่า Support ซึ่งได้มาจากจำนวนของ Transaction ใน S ที่เกิด A หารด้วยจำนวน Transaction ใน A, φ (A => B / S) คืออัตราส่วนของ σ (A $\wedge$ B / S) หารด้วย σ (A / S) หมายถึงความน่าจะเป็นที่จะเกิด B ร่วมกับ A

จากคำอธิบายข้างบนสามารถเขียนในรูปแบบเพื่อให้ง่ายในการตรวจสอบได้ว่า

1. $\sigma (A \wedge B / S) = \sigma (B \wedge A / S)$
2. $\sigma (A \wedge B / S) \leq \sigma (A / S)$

ข้อแรกจะเป็นขอกำหนดในรูปแบบทั่วไป และในแบบที่สองจะใช้กับรูปแบบหลากหลายที่เกิดขึ้นร่วมกัน

3.1.2 รูปแบบของ A จะเรียกว่า large ใน S ที่ระดับชั้นที่ l ถ้า จำนวนครั้งที่เกิด A ไม่น้อยกว่า ค่า Minimum support ที่ระดับชั้นที่ l (σ_l') และรูปแบบการเกิด A => B / S จะถูกเรียกว่า high ที่ระดับชั้น l เมื่อค่า Confidence ในระดับชั้นนี้มากกว่า minimum confidence ที่กำหนดไวในระดับชั้น l (φ_l')

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 กฎที่ว่า A => B / S จะถูกเรียกว่า Strong เมื่อ σ (A $\wedge$ B / S) Large ในระดับชั้นปัจจุบัน และ A => B / S high ในระดับชั้นปัจจุบัน และทุกระดับชั้นที่สูงกว่า A และ B จะต้อง Large ด้วยเช่นกัน

3.2 ขั้นตอนการทำ Multiple Level Association Rules

ตัวอย่างการเก็บข้อมูลการซื้อขายสินค้าของห้างสรรพสินค้าแห่งหนึ่ง ซึ่งประกอบไปด้วยข้อมูลรายการสินค้า, รายการขายสินค้า ดังนี้

ตารางที่ 3.1 ข้อมูลรายการสินค้า

bar_code	category	brand	producer	content_spec	size	Storage_period	price
17325	Milk	Foremost	Foremost Farm	2%	1 gallon	14 days	3.89
...	...	...	...	...	...	...	...

ตารางที่ 3.2 ข้อมูลการขายสินค้า

Transaction_id	bar_code_set
351428	{17325, 92108, 55349, 88157, 77493, 30816}
982510	{92458, 77451, 60392, 88144, 42316, 35672,29563}
...	{..., ...}

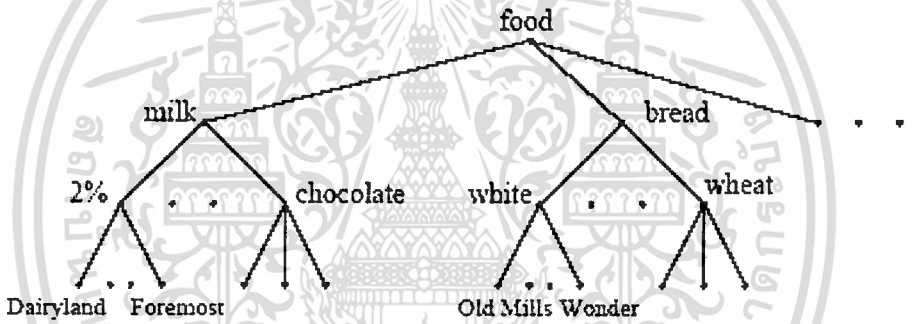
จากข้อมูลดังกล่าวเพื่อนำมาทำการไมนิ่งในแบบ Multiple Level Association Rules เริ่มจากการสร้างตารางซึ่งเกิดจากการจัดกลุ่มข้อมูลโดยเลือกระดับชั้นของข้อมูลที่นำมากำหนดลำดับในแต่ละระดับชั้น ซึ่งจะใช้ Category, content_spec, brand มาสร้างลำดับจากสูงไปยังลำดับล่าง ในตารางนี้จะสร้างรหัสของข้อมูลที่อยู่ในกลุ่มย่อยที่สุด (generalized_item_id) ว่าประกอบด้วยสินค้ารายการใดบ้าง ซึ่งเก็บไว้ใน bar_code_set ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 ข้อมูลที่ถูกสร้างใหม่จากการจัดกลุ่ม

Generalized_item_id	bar_code_set	category	content_spec	brand
112	{17325, 31414, 91265}	Milk	2%	Foremost
141	{29563, 77454, 89157}	Milk	Skim	Dairyland
171	{73295, 99184, 79520}	Milk	Chocolate	Dairyland
212	{88452, 35672, 98427, 31205}	Bread	Wheat	Wonder
...	{..., ...}	...	...	...
711	{32514, 78152}	Fruit_juice	Orange	Minute_maid

จากการจัดข้อมูลให้จะสามารถนำมาแสดงในรูปแบบต้นไมเพื่อให้เห็นถึงความสัมพันธ์ในแต่ละระดับชั้นได้ง่ายขึ้น ได้ดังภาพประกอบที่ 3.1



รูปที่ 3.1 แสดงระดับชั้นของการจัดกลุ่ม

จากตารางที่เราได้สร้าง generalized_item_id แล้วเราจะทำการนำค่า generalized_item_id มาแทน barcode ในตารางที่เก็บข้อมูลการขายสินค้า ซึ่งเรียกตารางนี้ว่า Encoded transaction ดังตัวอย่างของตารางใน 3.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 Encoded transaction Table

Transaction_id	Items
T_1	{111, 121, 211, 221}
T_2	{111, 211, 222, 323}
T_3	{112, 122, 221, 411}
T_4	{111, 121}
T_5	{111, 122, 211, 221, 413}
T_6	{211, 323, 524}
T_7	{323, 411, 524, 713}

เมื่อได้ Encode transaction Table มาแล้วจะเริ่มทำการหา Large Itemsets ใน Level ที่ 1 ดังตัวอย่างในตารางที่ 3.5 เมื่อหา Large itemsets ลำดับสุดท้ายได้แล้วจะจะทำการตัด Item ที่ไม่ได้อยู่ใน Large itemsets สุดท้ายของ Level นี้ออกไป จะได้ตารางใหม่คือ T[2] ซึ่งจะเห็นว่าจำนวน Transaction ใน T[2] จะน้อยกว่าในเริ่มแรก เพราะ T_7 จะถูกตัดออกไป เนื่องจาก T_7ประกอบไปด้วยสมาชิกที่ไม่ได้อยู่ใน L[1,2] ทั้งสิ้นจึงไม่มีสมาชิกเหลืออยู่จึงตัด T_7 ออกไป และนอกจากนี้จำนวนสมาชิกในแต่ละ Transaction ก็จะน้อยลงกว่าเดิม เพราะสมาชิกบางตัวไม่ได้เป็นสมาชิกใน L[1,2] นั้นเอง

ตารางที่ 3.5 การหา Large Itemsets ที่ Level 1

Level-1 minsup = 4

Level-1 large 1-itemsets: L[1,1]

Itemset	Support
{1 * *}	5
{2 * *}	5

Level-1 large 2-itemsets: L[1,2]

Itemset	Support
{1 * *}, {2 * *}	4

Filtered transaction table: T[2]

TID	Items
T_1	{111, 121, 211, 221}
T_2	{111, 211, 222}
T_3	{112, 122, 221}
T_4	{111, 121}
T_5	{111, 122, 211, 221}
T_6	{211}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในระดับชั้นที่ 2 จะใช้ T[2] แทนการใช้ T ทั้งหมดในการหา Large Itemsets ซึ่งจากตารางที่ 3.6 จะกำหนด Minimum support ไว้ที่ 3 และทำการนำจำนวน item จาก T[2] จนกระทั่งได้ Large 3-items

ตารางที่ 3.6 การหา Large Itemsets ที่ Level 2

Level-2 minsup = 3

Level-2 large 1-itemsets: L[2,1]

Itemset	Support
{11 *}	5
{12 *}	4
{21 *}	4
{22 *}	4

Level-2 large 3-itemsets: L[2,3]

Itemset	Support
{11 *}, {12 *}, {22 *}	3

Level-2 large 2-itemsets: L[2,2]

Itemset	Support
{11 *}, {12 *}	4
{11 *}, {21 *}	3
{11 *}, {22 *}	4
{12 *}, {22 *}	3
{21 *}, {22 *}	3

ในระดับชั้นที่ 3 ก็เช่นเดียวกับระดับชั้นที่ 2 จะใช้ T[2] แทนการใช้ T ทั้งหมดในการหา Large Itemsets ซึ่งจากตารางที่ 3.7 จะกำหนด Minimum support ไว้ที่ 3 และทำการนำจำนวน item จาก T[2] จนกระทั่งได้ Large 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 การหา Large Itemsets ที่ Level 3

Level-3 minsup = 3

Level-3 large 1-itemsets: L[3,1]

Itemset	Support
{111}	4
{211}	4
{221}	3

Level-3 large 2-itemsets: L[3,2]

Itemset	Support
{111}, {211}	3

ซึ่งจะเห็นว่าในการหา Large Itemsets ในระดับชั้นตั้งแต่ 2 ขึ้นไปจะใช้ T[2] เป็นตัวนับข้อมูล โดยจะไม่ใช้ Transaction ทั้งหมดในการนับ

3.3 อัลกอริทึม ML-T2

ML-T2 คืออัลกอริทึมที่ใช้ในการหา Large Itemsets ทั้งหมดในทุก Level โดยมีหลักการทำงานดังนี้

ข้อมูลที่ใช้คำนวณ

- ชุดของ Transaction ทั้งหมด
- ค่า Minimum Support ของแต่ละระดับชั้น

ผลลัพธ์ที่ได้

- Large Itemsets ทั้งหมดที่เกิดขึ้นในแต่ละระดับชั้น

วิธีการ

ทำการหา Large Itemsets จากระดับชั้นที่ 1 เมื่อได้แล้วทำการสร้าง T[2] จากการตัดข้อมูลที่ไม่ได้เป็นสมาชิกใน Large Itemsets ในระดับชั้นนี้ออกไป แล้วนำ T[2] ไปใช้ในการหา Lagre Itemsets ในระดับชั้นต่อไป ซึ่งอัลกอริทึม ML-T2 มีรายละเอียดดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

(1) for ($l := 1$; $\mathcal{L}[l,1] \neq \emptyset$ and $l < max_level$; l++) do {
(2) if $l = 1$ then {
(3) $\mathcal{L}[l,1] := get_large_1_itemsets(\mathcal{T}[1], l)$;
(4) $\mathcal{T}[2] := get_filtered_t_table(\mathcal{T}[1], \mathcal{L}[1,1])$;
(5) }
(6) else $\mathcal{L}[l,1] := get_large_1_itemsets(\mathcal{T}[2], l)$;
(7) for ($k := 2$; $\mathcal{L}[l,k-1] \neq \emptyset$; k++) do {
(8) $C_k := get_candidate_set(\mathcal{L}[l,k-1])$;
(9) forcach transaction $t \in \mathcal{T}[2]$ do {
(10) $C_t := get_subsets(C_k, t)$;
(11) forcach candidate $c \in C_t$ do c.support++;
(12) }
(13) $\mathcal{L}[l,k] := \{c \in C_k | c.support \geq minsup[l]\}$
(14) }
(15) $\mathcal{LL}[l] := \bigcup_k \mathcal{L}[l,k]$;
(16) } □

3.4 ดาต้าไมนิ่งในโครงสร้างแบบผสม

ในการทำไมนิ่งที่ได้กล่าวมาแล้วจะเห็นว่าการไมนิ่งจะทำให้เราทราบความสัมพันธ์ของข้อมูลที่อยู่ในระดับชั้นเดียวกันเช่น 2% Milk => Wheat Bread เป็นความสัมพันธ์ระหว่าง content_spect หรือ Foremost Milk => Wonder Bread เป็นความสัมพันธ์ระหว่ายี่ห้อของสินค้า ซึ่งหากต้องการทราบความสัมพันธ์ของ 2% Milk => Wonder Bread จะไม่สามารถทราบได้เนื่องจาก content_spec อยู่ในระดับชั้นล่างกว่ายี่ห้อนั้นเอง ดังนั้นการไมนิ่งเพื่อหาความสัมพันธ์ลักษณะนี้ จะเป็นลักษณะผสมของต้นไม้เหมือนการมองต้นไม้ที่มีการเรียงระดับชั้นต่างกันพร้อมกัน โดยต้นหนึ่งเห็นการเรียงระดับชั้นจาก category, content_spect ส่วนอีกต้นหนึ่งเรียงระดับชั้นเป็น category, brand ในการทำไมนิ่งแบบนี้ในหา Large itemsets ที่ระดับชั้นที่ 2 จะทำการเลือก Items ที่อยู่ในต้นไม้ทั้งสองแบบมา เช่น {11 *}, {2 * 1} ดังตารางที่ 3.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.8 การหาทำไมนิ่งในโครงสร้างแบบผสม

Level-2 minsup = 3

Level-2 large 1-itemsets: L[2,1]

Itemset	Support
{11 *}	5
{12 *}	4
{1 * 1}	4
{21 *}	4
{22 *}	4
{2 * 1}	4

Level-2 large 2-itemsets: L[2,2]

Itemset	Support
{11 *}, {12 *}	4
{11 *}, {21 *}	3
{11 *}, {22 *}	4
{11 *}, {2 * 1}	4
{12 *}, {22 *}	3
{12 *}, {2 * 1}	4
{21 *}, {22 *}	3
{21 *}, {1 * 1}	3

Level-2 large 3-itemsets: L[2,3]

Itemset	Support
{11 *}, {12 *}, {22 *}	3
{11 *}, {12 *}, {2 * 1}	3

ผลการทำไมนิ่งจะทำให้เราหาความสัมพันธ์ของ 2% Milk => Wonder Bread ได้ แต่การทำไมนิ่งในแบบนี้ทำให้เราได้ความสัมพันธ์ในรูปแบบ ของ 2% Milk => Foremost Milk ซึ่งไม่มีความหมายใดๆ และนำไปใช้งานในการวิเคราะห์ไม่ได้

3.5 ดาต้าไมนิ่งระหว่างระดับชั้น

ในการมองหารูปแบบความสัมพันธ์ซึ่งอยู่ในระดับชั้นต่างกันจะไม่สามารถหาได้จากรูปแบบไมนิ่งในก่อนหน้านี้ เช่น ความสัมพันธ์ในรูปแบบ 2% milk => bread หรือ milk => 2% milk เพราะเกิดจากความสัมพันธ์ในระดับชั้นที่แตกต่างกันของข้อมูล ดังนั้นในการค้นหาความสัมพันธ์ที่เกิดจากระดับชั้นที่ต่างกันในการหา large 2-itemsets จะเป็นการสร้างสมาชิกที่นอกเหนือจากการนำ Large 1-itemsets ในระดับชั้นที่ 2 มาสร้างสมาชิกทั้งหมดของ itemsets แล้วนำมาเทียบกับ Minimum support ในระดับชั้นที่ 2 เพื่อให้ได้ Large 2 itemsets แต่ยังรวมถึงสมาชิกที่เกิดจาก Large 1-itemsets ในระดับชั้นที่ 1 กับสมาชิก Large 1-itemsets ในระดับชั้นที่ 2เป็นสมาชิกร่วมกันใน itemsets และมีค่ามากว่า Minimum support ของระดับชั้นที่ 2 ด้วย ดังตัวอย่างในตารางที่ 3.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.9 การทำไมนิ่งระหว่างระดับชั้นใน Level ที่ 2

Level-2 minsup = 3

Level-2 large 1-itemsets: L[2,1]

Itemset	Support
{11 *}	5
{12 *}	4
{21 *}	4
{22 *}	4

Level-2 large 2-itemsets: L[2,2]

Itemset	Support
{11 *}, {12 *}	4
{11 *}, {21 *}	3
{11 *}, {22 *}	4
{12 *}, {22 *}	3
{21 *}, {22 *}	3
{11 *}, {2 * *}	4
{12 *}, {2 * *}	3
{21 *}, {1 * *}	3
{22 *}, {1 * *}	4

Level-2 large 3-itemsets: L[2,3]

Itemset	Support
{11 *}, {12 *}, {22 *}	3
{21 *}, {22 *}, {1* *}	3

และการหา Large itemsets ในระดับชั้นที่ 3 ก็ทำเช่นเดียวกับในระดับชั้นที่ 2 คือนำ Large 1-itemsets ของระดับชั้นที่ 1 และ 2 มาสร้าง Large 2-itemsets ของระดับชั้นที่ 3 ด้วย แต่หากสังเกตดูเห็นว่าในการหา Large 2-itemsets ของระดับชั้นที่ 2 และที่ 3 แม้จะนำสมาชิกมาจาก Large itemsets ของระดับชั้นที่ต่ำกว่ามาใช้ก็ตาม แต่จะไม่สร้างสมาชิกที่เกิดจากระดับชั้นที่ต่ำกว่าทั้งหมด แต่จะต้องมี Large 1-itemsets ของระดับชั้นนั้นๆ เป็นสมาชิกอยู่ด้วยอย่างน้อย 1 ตัวเสมอ ตามตัวอย่างจากตารางที่ 3.10

ตารางที่ 3.10 การทำไมนิ่งระหว่างระดับชั้นใน Level ที่ 3

Level-3 minsup = 3

Level-3 large 1-itemsets: L[3,1]

Itemset	Support
{111}	4
{211}	4
{221}	3

Level-3 large 2-itemsets: L[3,2]

Itemset	Support
{111}, {211}	3
{111}, {21 *}	3
{111}, {22 *}	3
{111}, {2 * *}	4
{11 *}, {211}	3
{1* *}, {211}	3

Level-3 large 3-itemsets: L[3,3]

Itemset	Support
{111}, {21 *}, {22 *}	3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

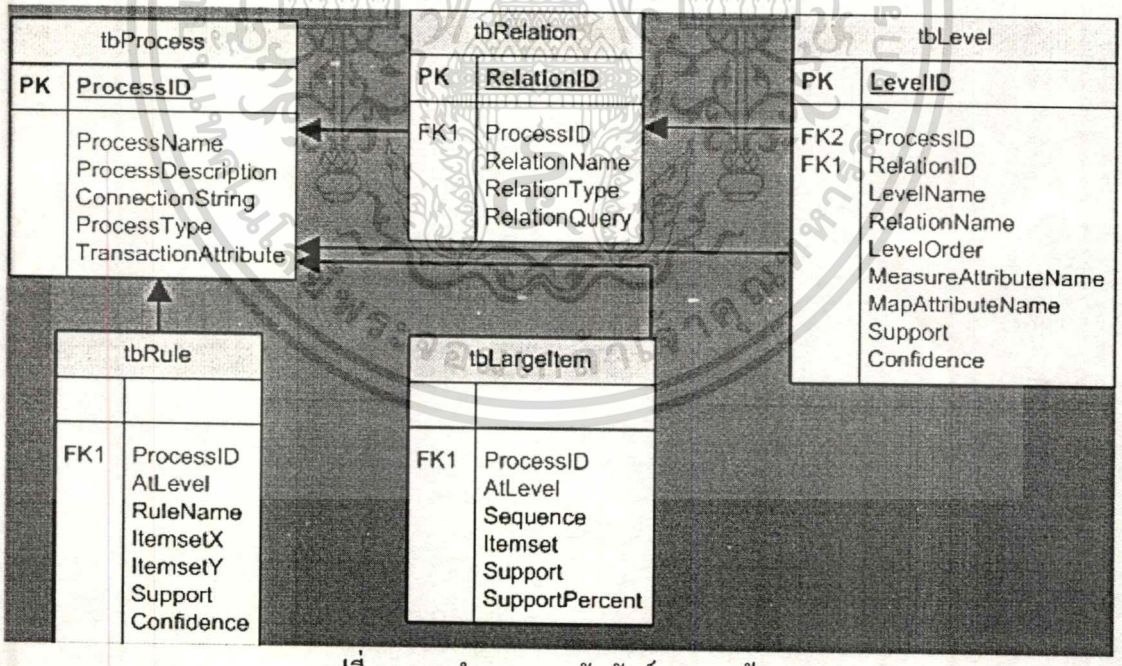
การวิเคราะห์และออกแบบระบบ

ระบบการค้นหาความสัมพันธ์ในแบบ Multiple Level Association Rules การทำงานของระบบมีดังนี้

1. สามารถนำเข้าข้อมูลจากระบบอื่นๆได้ โดยทำการจับคู่กับโครงสร้างของข้อมูลที่มีอยู่
2. ระบบสามารถวิเคราะห์ความสัมพันธ์แบบ Multiple Level Association Rules

4.1 ฐานข้อมูลในระบบ

โครงสร้างของข้อมูลที่ใช้ประกอบไปด้วย Table ต่างๆดังนี้



รูปที่ 4.1 แบบจำลองความสัมพันธ์ของฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ข้อมูลรายละเอียดของงาน

Table name : tbProcess

Description : จะเก็บงานทั้งหมดที่ทำการวิเคราะห์ เพื่อใช้อธิบายรายละเอียดของงานที่ทำการไมนิ่ง ซึ่งจะประกอบไปด้วยรายละเอียดเกี่ยวกับชนิดของแหล่งข้อมูลที่นำมาทำการไมนิ่ง ตลอดจนการกำหนดค่าต่างๆที่นำมาใช้ในการทำไมนิ่ง

Attribute Name	Contents	Type	Size	Key	Foreign Key Referenced Table
ProcessID	รหัสของงาน	nvarchar	50	PK	
ProcessName	ชื่อของงานที่ทำการไมนิ่ง	nvarchar	100		
ProcessDescription	รายละเอียดของงานที่ทำการไมนิ่ง	nvarchar	1000		
ConnectionString	Connection String ที่ใช้ในการติดต่อกับแหล่งของข้อมูลที่จะการไมนิ่ง	nvarchar	1000		
ProcessType	ชนิดของแหล่งข้อมูลที่ทำการไมนิ่งซึ่งประกอบด้วย 3 ประเภท คือ 1.ฐานข้อมูล SQL 2.ข้อมูลจากเอกสารที่เป็น XML 3.ข้อมูลจากเอกสารที่เป็น CSV (Comma Delimited)	nvarchar	50		
TransactionAttribute	อ้างอิงถึง Attribute ซึ่งเก็บรหัสของข้อมูลที่ระบุถึงข้อมูลที่เกิดขึ้นในครั้งเดียวกัน	nvarchar	50		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ความสัมพันธ์และการเรียกใช้ข้อมูลในการไมนิ่ง

Table name : tbRelation

Description : จัดเก็บข้อมูลอ้างอิงถึงความสัมพันธ์ของข้อมูล ซึ่งจะเป็นในลักษณะของการอ้างถึงตารางเพียงหนึ่งตาราง หรือหลายตารางในรูปแบบของการใช้ Query String ที่ใช้ในการเรียกข้อมูลความสัมพันธ์นั้นมาใช้งาน หรือการอ้างถึงเอกสารที่เป็นแหล่งของข้อมูลที่นำมาใช้ในการทำไมนิ่ง เช่น เอกสาร XML, เอกสาร CSV เป็นต้น ตลอดจนรายละเอียดในการนำมาใช้งานในการทำไมนิ่ง

Attribute Name	Contents	Type	Size	Key	Foreign Key Referenced Table
RelationID	รหัสของความสัมพันธ์ของข้อมูล	nvarchar	50	PK	
ProcessID	รหัสของงาน	nvarchar	50	FK	tbProcess
RelationName	ชื่อความสัมพันธ์ของข้อมูล	nvarchar	100		
RelationType	ประเภทของความสัมพันธ์ ประกอบด้วยความสัมพันธ์ 2 ประเภท คือ 1.ความสัมพันธ์ที่ใช้ในการอ้างถึงแหล่งข้อมูลหลักที่ใช้ทำการไมนิ่ง หรือข้อมูลซึ่งเป็น Transaction นั้นเอง 2.ข้อมูลที่มีเป็นการจัดหมวดหมู่ของข้อมูล จะเชื่อมโยงกับความสัมพันธ์กับข้อมูลในแบบแรก	nvarchar	50		
RelationQuery	Query String ที่ใช้ในการเรียกข้อมูลในฐานข้อมูลแบบ SQL ซึ่งถ้าเป็นลักษณะงานที่เป็นแบบใช้เอกสาร XML และ CSV จะทำการเก็บตำแหน่งที่จัดเก็บของเอกสารนั้นๆ	nvarchar	2000		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ความสัมพันธ์และการเรียกใช้ข้อมูลในการไมนิ่ง

Table name : tbLevel Description : เก็บข้อมูลของการจัดระดับชั้นในการทำไมนิ่งแบบหลายระดับชั้น ซึ่งจะจัดเก็บเกี่ยวกับรายละเอียดของแต่ละระดับชั้น และการเชื่อมโยงของข้อมูลแต่ระดับชั้นกับข้อมูลหลัก					
Attribute Name	Contents	Type	Size	Key	Foreign Key Referenced Table
LevelID	รหัสของระดับชั้น	nvarchar	50	PK	
ProcessID	รหัสของงาน	nvarchar	50	FK	tbProcess
RelationID	รหัสของความสัมพันธ์ของข้อมูล	nvarchar	50	FK	tbRelation
LevelName	ชื่อของระดับชั้น	nvarchar	50		
LevelOrder	ระดับชั้นที่	int	4		
MeasureAttributeName	ชื่อของ Attribute ในข้อมูลหลักที่ใช้สร้างความสัมพันธ์กับชุดข้อมูลที่จัดเก็บระดับชั้นนี้	nvarchar	50		
MapAttributeName	ชื่อของ Attribute ในข้อมูลที่เก็บระดับชั้น ที่ใช้สร้างความสัมพันธ์กับข้อมูลหลัก	nvarchar	50		
Support	Minimum Support ที่ใช้ในการทำไมนิ่งหา Large Itemsets ในระดับชั้นนี้	decimal	7		
Confidence	ค่า Minimum Confidence	decimal	7		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 ข้อมูลของ Large Itemsets

Table name : tbLargeItem

Description : จัดเก็บรายละเอียดของข้อมูลที่เป็น Large Itemsets ของงานนั้น และรายละเอียดของ Itemset นั้นว่ามีจำนวนครั้งที่เกิดขึ้นมากน้อยแค่ไหน และเป็น Large Itemsets ในระดับชั้นใด หรือเป็น Large Itemsets ในลำดับที่เท่าไร

Attribute Name	Contents	Type	Size	Key	Foreign Key Referenced Table
ProcessID	รหัสของงาน	nvarchar	50	FK	tbProcess
AtLevel	ระดับชั้นที่ทำการหา Large Itemsets	int	4		
Sequence	Large Itemsets ลำดับที่ทำการหา	int	4		
Itemset	ชื่อของ Itemset ที่มีค่าเกินกว่าค่าของ Minimum Support ซึ่งหากน้อยกว่าค่า Minimum Support จะไม่เป็น Large Itemsets (ระบบจะจัดเก็บเฉพาะข้อมูลที่มากกว่าค่า Minimum Support ของ Level นั้นๆเท่านั้น)	nvarchar	50		
Support	จำนวนครั้งของ Itemset ดังกล่าวที่เกิดขึ้นในชุดข้อมูล	Number	9		
SupportPercent	ความเปอร์เซ็นต์ของความน่าจะเป็นที่เกิด Itemset นี้ใน Tranasaction	Number	9		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ข้อมูลของกฎที่มีความน่าสนใจ

Table name : tbRule

Description : จัดเก็บข้อมูลเกี่ยวกับกฎที่มีความน่าสนใจ หรือมีค่าของ Confidence สูงกว่าค่า Minimum Confidence ที่ได้กำหนดไว้ โดยกฎจะอยู่ในรูปแบบของ X => Y หรือการเกิด Item X และจะเกิด Item Y ในข้อมูลที่โอกาสเกิดขึ้นได้มากน้อยเพียงใด ซึ่งทั้งหมดจะถูกจัดเก็บในตารางข้อมูลนี้

Attribute Name	Contents	Type	Size	Key	Foreign Key Referenced Table
ProcessID	รหัสของงาน	nvarchar	50	FK	tbProcess
AtLevel	ระดับชั้นที่ทำการหากฎนั้นๆ	int	4		
RuleName	ชื่อของกฎในรูป X => Y	nvarchar	1000		
ItemsetX	ชื่อของ Item X จากรูปแบบกฎ X => Y	nvarchar	50		
ItemsetY	ชื่อของ Item Y จากรูปแบบกฎ X => Y	nvarchar	50		
Confidence	ค่า Confidence ของกฎนี้	decimal	7		
Support	ค่าความน่าจะเป็นที่จะเกิด Item X และ Item Y	decimal	7		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การพัฒนาระบบงาน และหน้าจอการทำงาน

5.1 สภาพแวดล้อมการพัฒนาระบบ

สภาพแวดล้อมประกอบไปด้วย 2 ส่วนหลักคือ ด้านฮาร์ดแวร์ และด้าน ซอฟต์แวร์

5.1.1 ฮาร์ดแวร์ ประกอบด้วย

- เครื่องคอมพิวเตอร์ Pentium 4 2.0 MHz.
- RAM 512 MB
- ฮาร์ดดิสก์ ไม่น้อยกว่า 1 G
- Mouse, Keyboard
- CD-ROM

5.1.2 ซอฟต์แวร์ ประกอบด้วย

- Window XP
- Visual Studio .Net 2003 (Visual Basic)
- Crystal Report
- MSDN Library
- MS SQL Server

5.2 การจัดเตรียมสภาพแวดล้อมในการพัฒนา

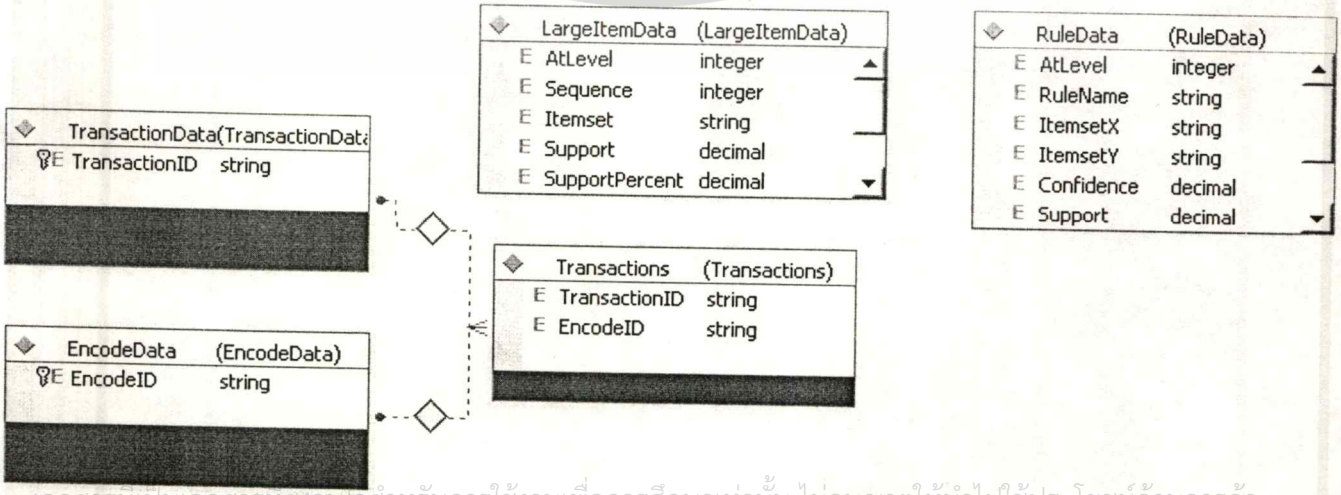
ในการพัฒนาระบบการจัดเตรียมสภาพแวดล้อมให้พร้อมสำหรับการพัฒนา โดยการติดตั้ง Window XP หลังจากนั้นจึงติดตั้ง Visual Studio .Net 2003 สำหรับทำการพัฒนาโปรแกรม นอกจากนี้เพื่อให้สะดวกในการพัฒนาการติดตั้ง MSDN เพื่อช่วยค้นหาข้อมูลต่างๆในการพัฒนา

ในการใช้งานฐานข้อมูลซึ่งจะเก็บข้อมูลของระบบเองจะใช้ MS Access 2000 ในการจัดเก็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 ขั้นตอนการทำงานในการไมนิ่ง ขั้นตอนในส่วนของการประมวลผลซึ่งใช้ตามอัลกอริทึม MLT2 มีขั้นตอนในการทำงานดังนี้

- **สร้างตาราง EncodeData** ซึ่งมีโครงสร้างดังตารางที่ 5.1 เพื่อเก็บรหัสอ้างอิงใหม่ตามการจัดกลุ่ม
- **Transform ข้อมูล** Transaction ลงในตาราง TransactionData ซึ่งมีโครงสร้างดังตารางที่ 5.2 และ Transactions ซึ่งมีโครงสร้างดังตารางที่ 5.3 เพื่อให้ข้อมูลอยู่ในรูปแบบที่พร้อมในการทำไมนิ่ง
- **สร้างตาราง T2** ซึ่งมาจากการหา Large Itemsets ที่ 1 ของ Level ที่ 1 และนำข้อมูลที่ได้ทำดังนี้ เพื่อให้ได้ตาราง T2
  - ตัด Item ที่ไม่อยู่ใน Large Itemsets ออกจาก Transaction
  - ตัดข้อมูล Transaction ที่ไม่มี Item ออก
- **หา Large Itemsets อื่นๆ** ใน Level 1 และใน Level ต่อไปโดยใช้ T2 ในการหาแทนตาราง Transaction จนกระทั่งสิ้นสุด และนำผลลัพท์ที่ได้เก็บไว้ในตาราง LargeItemData ซึ่งมีโครงสร้างดังตารางที่ 5.4
- **ทำการหา** Rule จากข้อมูลใน LaregItemData และนำผลที่ได้เก็บไว้ในตาราง RuleData ซึ่งมีโครงสร้างดังตารางที่ 5.5
- **จัดเก็บผลลัพท์เข้าฐานข้อมูล** นำข้อมูลจากตาราง LargeItemData และ RuleData จัดเก็บในฐานข้อมูลที่ตาราง tbLargeItem และ tbRule



รูปที่ 5.1 โครงสร้างข้อมูลในระหว่างทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในระหว่างขั้นตอนการทำไมนิ่งจะมีการสร้างตารางชั่วคราวขึ้นใช้งานโดยมีโครงสร้างดังรูปประกอบที่ 5.1 ซึ่งอธิบายรายละเอียดได้ดังนี้

ตารางที่ 5.1 EncodeData

Table name : EncodeData **Description** เพื่อเก็บรหัสอ้างอิงใหม่ตามการจัดกลุ่ม		
Attribute Name	**Contents**	**Type**
EncodeID	รหัสของ Item ที่ใช้อ้างอิงการจัดกลุ่มตามระดับชั้น	String

ตารางที่ 5.2 TransactionData

Table name : TransactionData **Description** : ข้อมูลแสดงรหัสชุดของข้อมูล		
Attribute Name	**Contents**	**Type**
TransactionID	รหัสชุดของข้อมูล	String

ตารางที่ 5.3 Transactions

Table name : Transactions **Description** : ข้อมูลแสดงรหัสชุดของข้อมูลในแต่ละรายการ		
Attribute Name	**Contents**	**Type**
TransactionID	รหัสชุดของข้อมูล จะสัมพันธ์กับชุดข้อมูลในตาราง TransactionData	String
EncodeID	รหัสของ Item ที่ใช้อ้างอิงการจัดกลุ่มตามระดับชั้น	String

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.4 LargeItemData

Table name : LargeItemData **Description** : รายละเอียดของข้อมูลที่เป็น Large Itemsets ของงานนั้น และรายละเอียดของ Itemset นั้นว่ามีจำนวนครั้งที่เกิดขึ้นมากน้อยแค่ไหน และเป็น Large Itemsets ในระดับชั้นใด หรือเป็น Large Itemsets ในลำดับที่เท่าไร		
Attribute Name	**Contents**	**Type**
AtLevel	ระดับชั้นที่ทำการหา Large Itemsets	integer
Sequence	Large Itemsets ลำดับที่ทำการหา	integer
Itemset	ชื่อของ Itemset ที่มีค่าเกินกว่าค่าของ Minimum Support ซึ่งหากน้อยกว่าค่า Minimum Support จะไม่เป็น Large Itemsets (ระบบจะจัดเก็บเฉพาะข้อมูลที่มากกว่าค่า Minimum Support ของ Level นั้นๆเท่านั้น)	String
Support	จำนวนครั้งของ Itemset ดังกล่าวที่เกิดขึ้นในชุดข้อมูล	Decimal
SupportPercent	เปอร์เซ็นต์ของความน่าจะเป็นที่จะเกิด Itemset นี้ใน Transaction	Decimal

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

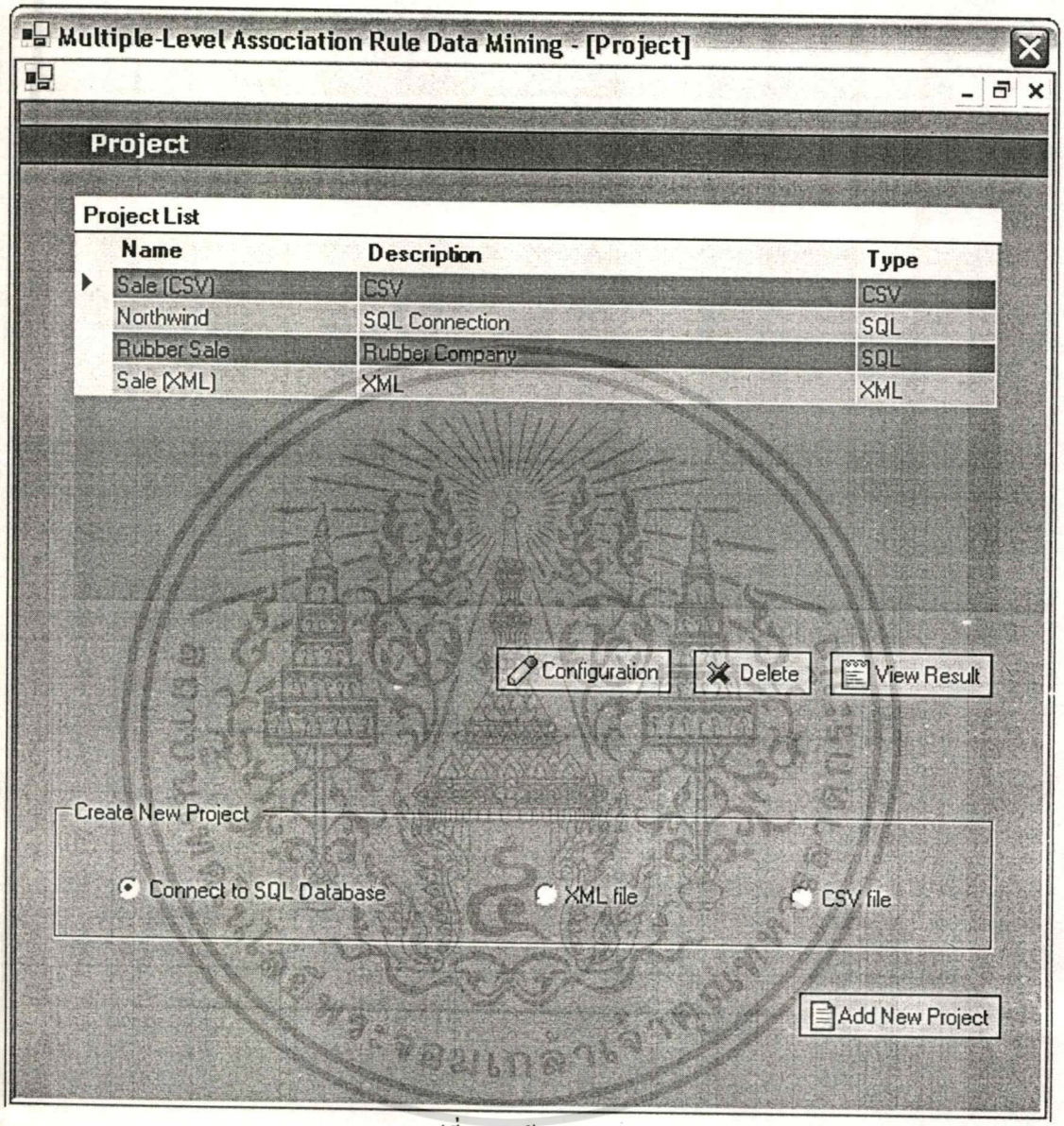
ตารางที่ 5.5 RuleData

Table name : RuleData **Description** :ข้อมูลเกี่ยวกับกฎที่มีความน่าสนใจ หรือมีค่าของ Confidence สูงกว่าค่า Minimum Confidence ที่ได้กำหนดไว้ โดยกฎจะอยู่ในรูปแบบของ X => Y หรือการเกิด Item X และจะเกิด Item Y ในข้อมูลที่โอกาสเกิดขึ้นได้มากน้อยเพียงใด ซึ่งทั้งหมดจะถูกจัดเก็บในตารางข้อมูลนี้		
Attribute Name	**Contents**	**Type**
AtLevel	ระดับชั้นที่ทำการหากฎนั้นๆ	integer
RuleName	ชื่อของกฎในรูป X => Y	String
ItemX	ชื่อของ Item X จากรูปแบบกฎ X => Y	String
ItemY	ชื่อของ Item Y จากรูปแบบกฎ X => Y	String
Confidence	ค่า Confidence ของกฎนี้	Decimal
Support	เปอร์เซ็นต์ความน่าจะเป็นที่จะเกิด Item X และ Item Y ใน Transaction	Decimal

5.4. การพัฒนาระบบและหน้าจอการทำงาน จากการวิเคราะห์และออกแบบระบบงานจะแบ่งส่วนของหน้าจอการทำงานต่างๆได้ดังนี้

5.4.1 หน้าหลักของระบบ ซึ่งจะแสดงรายละเอียดเกี่ยวกับงานที่จะทำการวิเคราะห์ และการกำหนดข้อมูลเริ่มต้น แบ่งออกเป็น 2 ส่วน ดังรูปที่ 5.2 คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 หน้าหลักของระบบ

- **Project List** คือส่วนของการแสดงงานที่ได้กำหนดรายละเอียดของงานไว้แล้ว หรืออาจทำการไมนิ่ง และทราบผลลัพท์แล้ว รายละเอียดของข้อมูลที่แสดงในตารางเกี่ยวกับงานมีดังนี้
  - **Name** แสดงชื่อของงานที่ได้ทำการบันทึกไว้
  - **Description** แสดงรายละเอียดของงานที่ผู้ใช้งานสามารถบันทึกเพิ่มเติมเพื่ออธิบายเกี่ยวกับลักษณะของงานได้
  - **Type** แสดงประเภทของงาน ซึ่งมี 3 ประเภท คือ SQL, XML, CSV เพื่อบ่งบอกที่มาของฐานข้อมูลที่นำมาวิเคราะห์ในครั้งนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

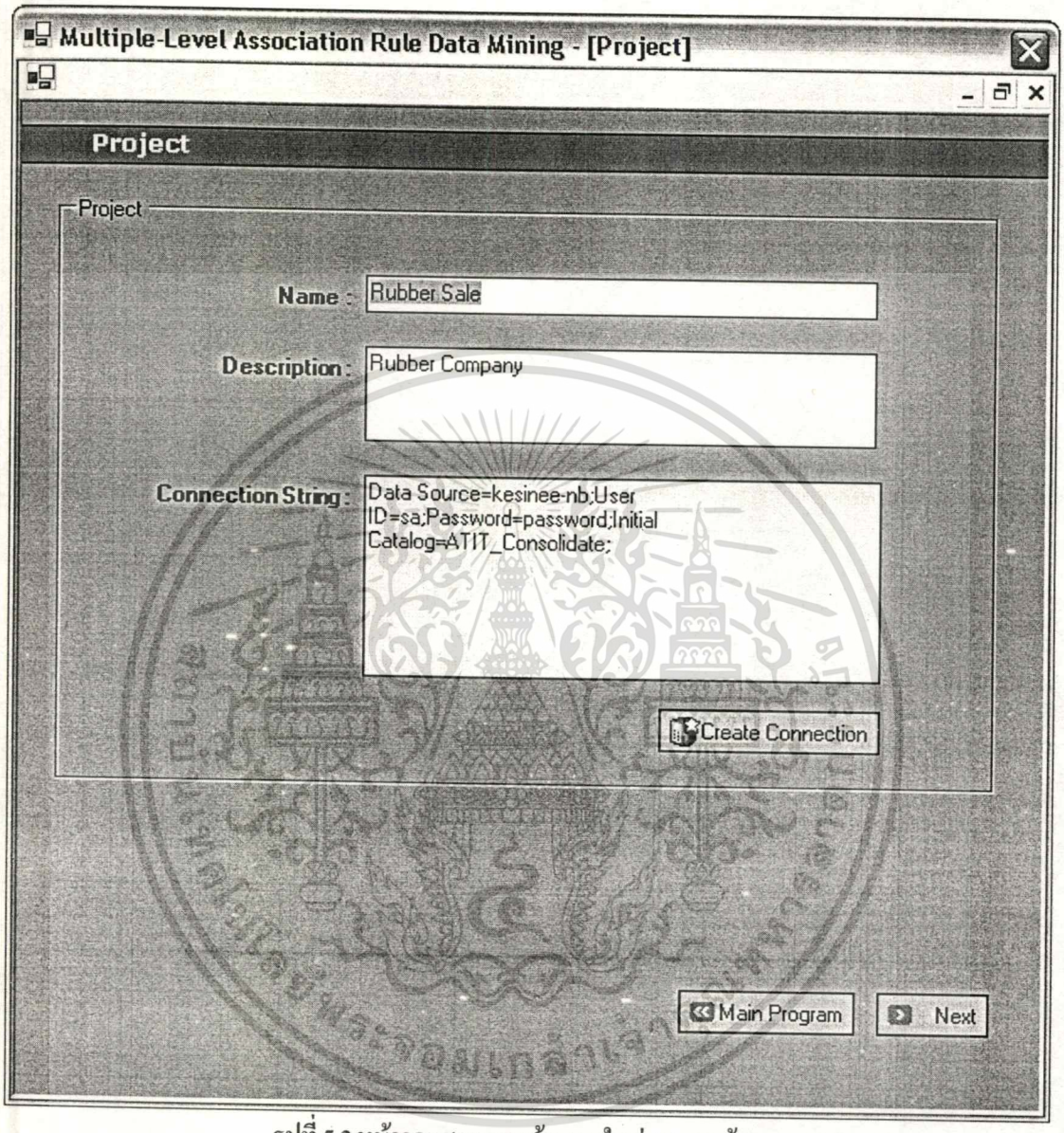
ผู้ใช้งานสามารถเปิดงานที่ต้องการโดยการเลือกที่ชื่องานนั้น และคลิ๊กที่ปุ่มต่างๆ เพื่อทำงานดังนี้

- ○ **ปุ่ม Configuration** เพื่อทำการปรับแต่งข้อมูลที่เคยได้บันทึกไว้ ตลอดจนทำการไมนิ่งเพื่อหาผลลัพธ์ใหม่จากรายละเอียดที่ได้แก้ไขไปใหม่
- ○ **ปุ่ม Delete** เพื่อทำการลบงาน และผลการทำไมนิ่งออกจากระบบ
- ○ **ปุ่ม View Result** จะทำการแสดงผลของการทำไมนิ่งของงานที่กำลังเลือกอยู่ โดยระบบจะดึงผลการไมนิ่งจากการทำไมนิ่งในครั้งสุดท้ายของงานชิ้นนั้นขึ้นมาแสดง

- **Create New Project** คือส่วนสำหรับสร้างงานใหม่ ซึ่งผู้ใช้งานจะต้องทำการระบุชนิดของงานที่ต้องการสร้างก่อน ซึ่งมีด้วยกัน 3 ประเภท คือ SQL, XML, CSV ตามแหล่งที่มาของข้อมูล และหลังจากนั้นจึงทำการคลิ๊กที่ปุ่ม Add New Project ระบบจะนำเข้าสู่ขั้นตอนในการสร้างงานใหม่ต่อไป

5.4.2 การสร้างงานใหม่จากฐานข้อมูล SQL หากผู้ใช้งานต้องการสร้างงานใหม่โดยใช้แหล่งที่มาของข้อมูลจากฐานข้อมูลของ SQL หลังจากผู้ใช้งานคลิ๊กปุ่ม Create New Project แล้ว ระบบจะแสดงหน้าจอให้ผู้ใช้งานกำหนดรายละเอียดเกี่ยวกับงาน ตามรูปประกอบที่ 5.3 ซึ่งมีรายละเอียดดังนี้

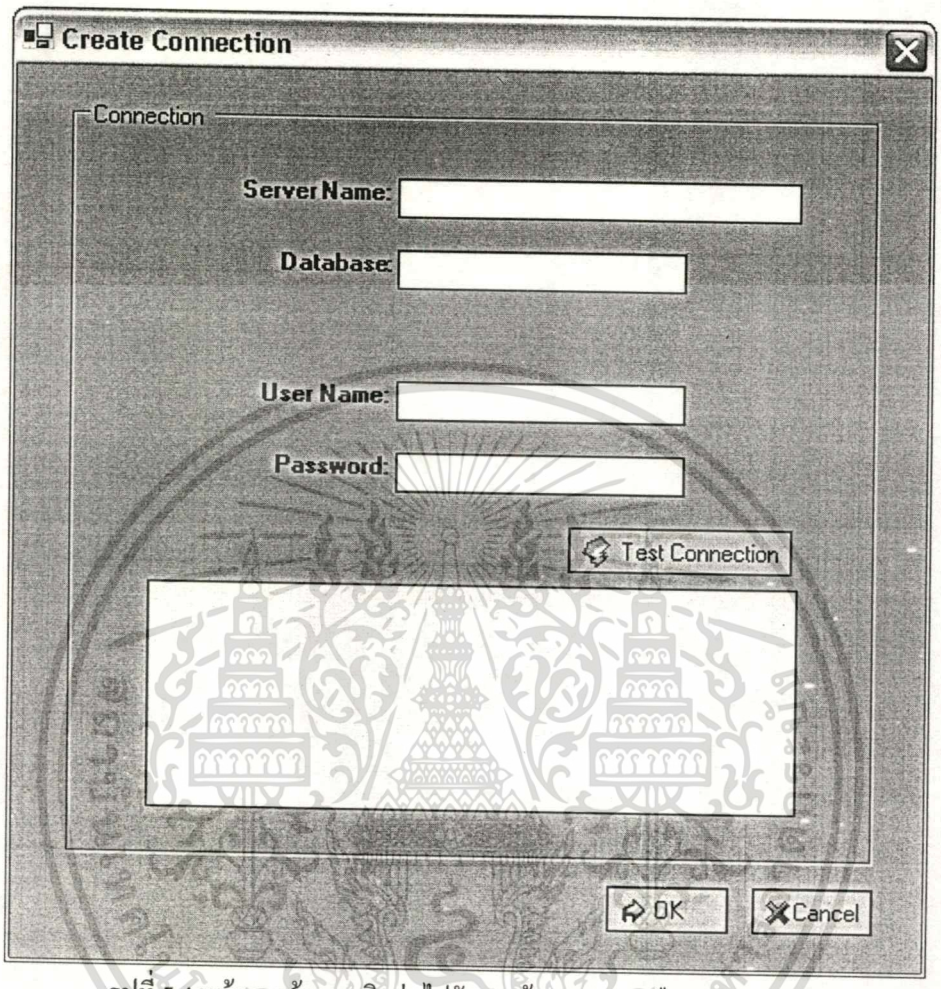
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.3 หน้าจอแสดงการสร้างงานใหม่จากฐานข้อมูล SQL

- **Name** ชื่อของงานที่ต้องการระบบ
- **Description** รายละเอียดของงาน เพื่อช่วยอธิบายรายละเอียดของงานที่จะทำการไมนิ่ง
- **Connection String** คือส่วนของการกำหนดส่วนเชื่อมต่อกับฐานข้อมูล SQL โดยผู้ใช้งานสามารถกำหนดโดยการคลิ๊กที่ปุ่ม Create Connection ซึ่งจะปรากฏหน้าจอดังรูปประกอบที่ 5.4 ซึ่งมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



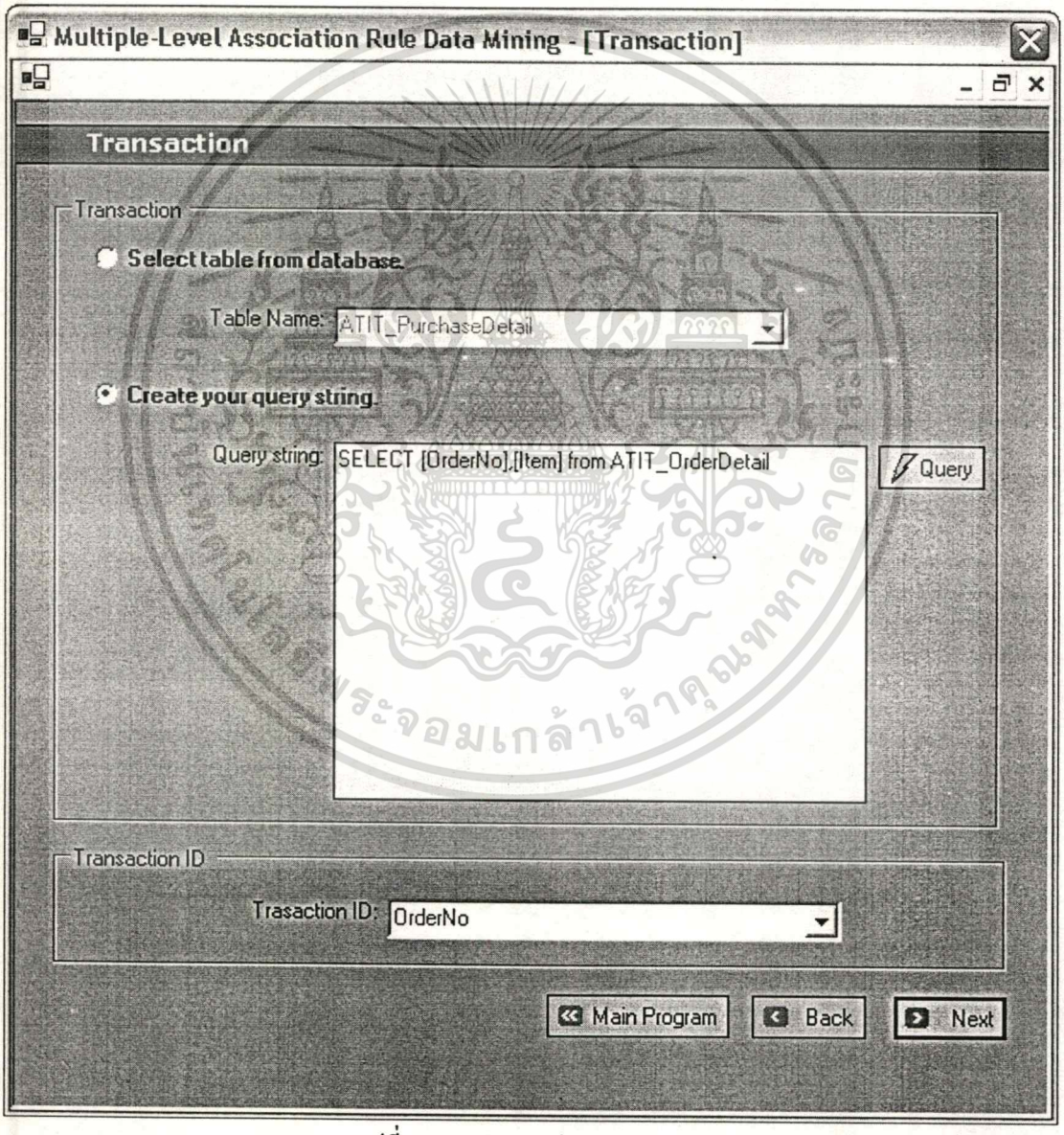
รูปที่ 5.4 หน้าจอสร้างการติดต่อไปยังฐานข้อมูล SQL หรือ Connection String

- **Server Name** ชื่อ Server ที่ข้อมูลเก็บอยู่
- **Database Name** ชื่อของฐานข้อมูลที่ต้องการนำมาใช้ในการทำไมนิ่ง
- **User Name** ชื่อของผู้ใช้งานที่มีสิทธิเข้าใช้งานในฐานข้อมูลดังกล่าว
- **Password** รหัสผ่านที่ใช้สำหรับผู้ใช้งานที่จะเข้าไปยังฐานข้อมูลได้
- **ปุ่ม Test Connection** ใช้เพื่อทำการทดสอบการใช้งานว่าสามารถติดต่อไปยังฐานข้อมูลดังกล่าวได้หรือไม่

- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป
- **ปุ่ม Next** เมื่อผู้ใช้งานทำการกำหนดรายละเอียดครบถ้วนแล้ว ก็จะทำการคลิ๊กที่ปุ่มนี้เพื่อเข้าสู่ขั้นตอนในลำดับต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

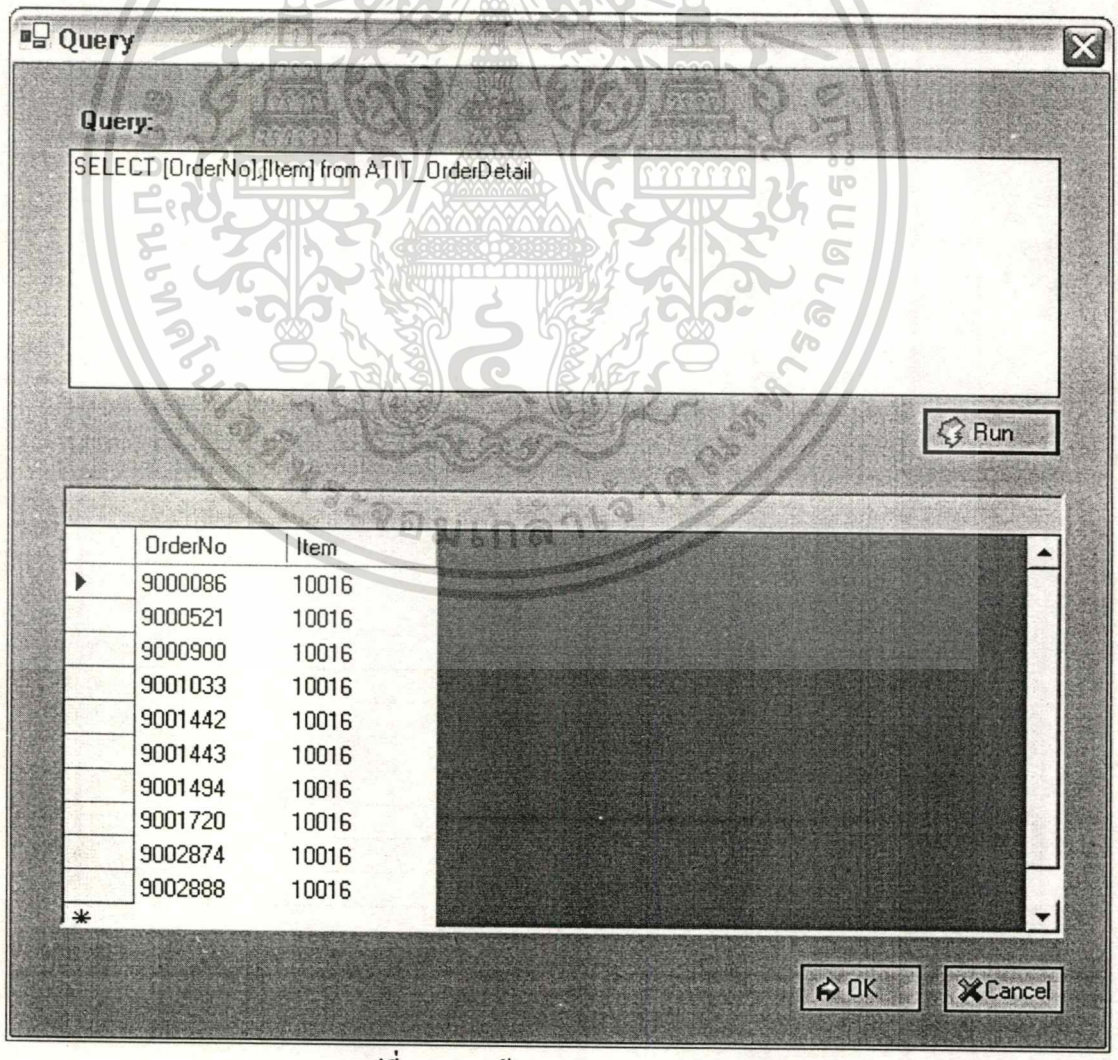
5.4.3 การกำหนดส่วนของ Transaction เมื่อผู้ใช้งานผ่านในส่วนของการกำหนดการติดต่อกับฐานข้อมูลแล้วจะเข้าสู้หน้าจอในส่วนของ Transaction นั้นคือผู้ใช้งานจะต้องระบุรายละเอียดของตารางที่จะนำข้อมูลมาวิเคราะห์โดยตารางที่จะนำมาเป็น Transaction คือข้อมูลหลักที่จะใช้นั้นเอง เช่นถ้าเป็นการขายสินค้าคือข้อมูลในส่วนของบันทึกการขายสินค้า เป็นต้น ดังรูปประกอบที่ 5.5 ในหน้าจอนี้จะมีรายละเอียดดังนี้



รูปที่ 5.5 การกำหนดส่วนของ Transaction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **Select table from database** ถ้าผู้ใช้มีข้อมูลที่สมบูรณ์เก็บไว้ให้ตารางใดตารางหนึ่ง ระบบจะแสดงรายการตารางจากในฐานข้อมูลทั้งหมดมาให้ ผู้ใช้จะต้องเลือกว่าจะเลือกจากตารางที่มีอยู่แล้ว และทำการเลือกชื่อตารางในรายการที่แสดงไว้ให้
- **Transaction ID** ระบบข้อมูลซึ่งใช้เก็บรหัสของชุดข้อมูล
- **Create your query string** หากผู้ใช้ต้องการกำหนดรายละเอียดของข้อมูลซึ่งมาจากการผนวกตารางมากกว่าหนึ่งตารางไว้ด้วยกัน ผู้ใช้สามารถเลือกในการกำหนดแบบ Create query string และหลังจากนั้นผู้ใช้งานต้องคลิ๊กที่ปุ่ม Query เพื่อทำการสร้าง Query ที่ต้องการ จะปรากฏหน้าจอเพื่อทำการสร้าง Query ดังรูปประกอบที่ 5.6 ซึ่งมีรายละเอียดการทำงานดังนี้



รูปที่ 5.6 การสร้าง Query String

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Query คือส่วนที่ผู้ใช้งานกำหนด Query String ตามรูปแบบของ SQL ลงไป แล หลังจากนั้นคลิ๊กยังปุ่ม Run เพื่อทำการประมวลผล และตรวจสอบความถูกต้องของ Query ที่กำหนดเมื่อ Query ที่กำหนดถูกต้อง จึงคลิ๊กปุ่ม OK เพื่อกลับเข้าสู้หน้าจอ Transaction ต่อไป

- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป
- **ปุ่ม Back** เมื่อผู้ใช้งานต้องการกลับไปยังหน้าจอก่อนหน้านี้
- **ปุ่ม Next** เมื่อผู้ใช้งานทำการกำหนดรายละเอียดครบถ้วนแล้ว ก็จะทำการคลิ๊กที่ปุ่มนี้เพื่อเข้าสู้ขั้นตอนในลำดับต่อไป

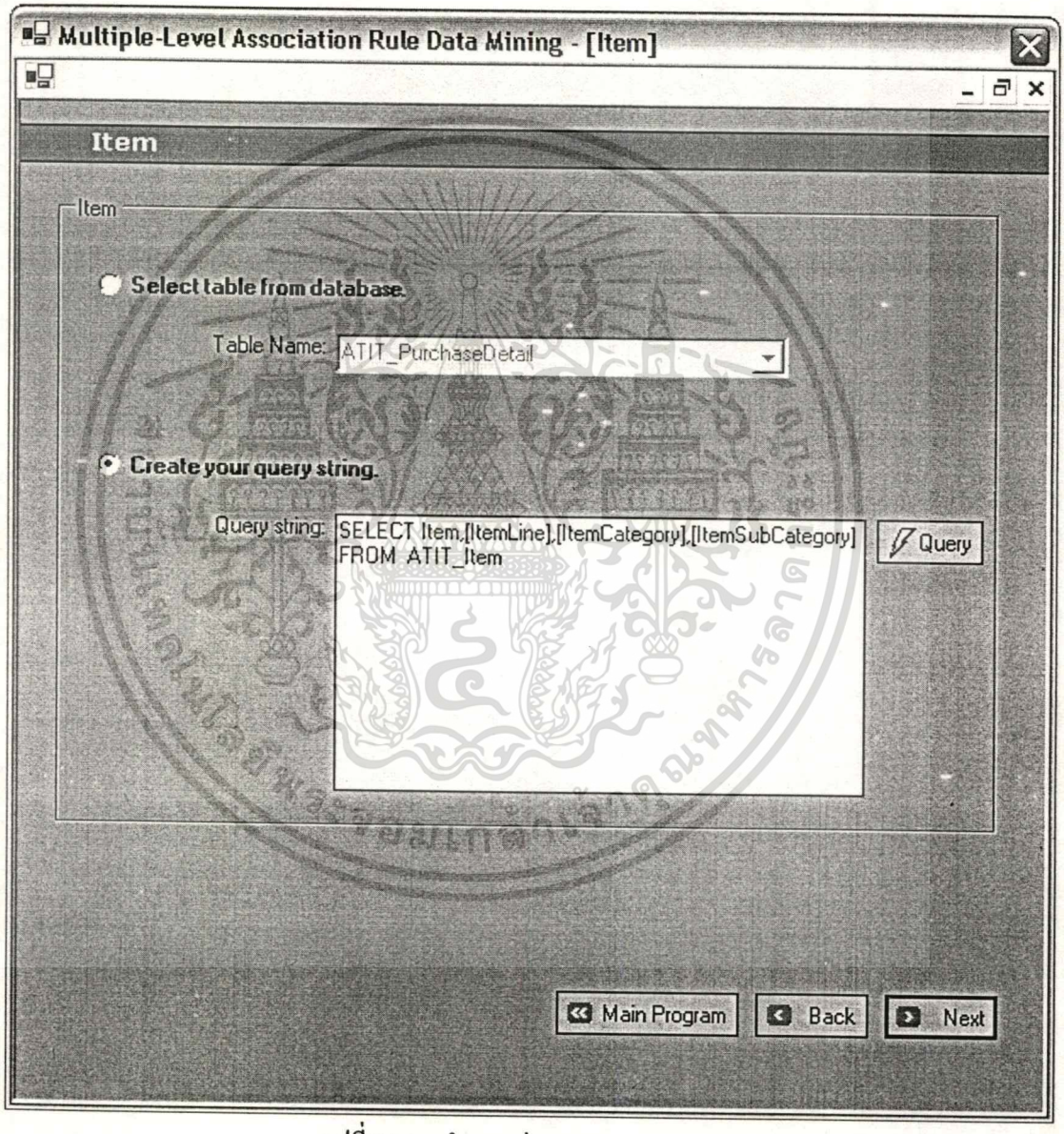
5.4.4 การกำหนดส่วนของ Item เมื่อผู้ใช้งานผ่านในส่วนของการกำหนด Transaction แล้ว ผู้ใช้งานจะต้องทำการระบุส่วนของ Item ที่นำมาใช้ในการทำไมนิ่งแบบ Multiple-Level ว่าข้อมูลที่ระบุ Item มาจากตารางใด โดยจะเป็นข้อมูลที่เก็บรายละเอียดของ Item ที่สัมพันธ์กับตาราง Transaction และมีรายละเอียดการจัดหมวดหมู่ของ Item ที่สามารถนำไปใช้กำหนดระดับชั้นของหมวดหมู่ต่อไปได้ ดังรูปประกอบที่ 5.7 ในหน้าจอนี้จะมีรายละเอียดดังนี้

- **Select table from database** ถ้าผู้ใช้มีข้อมูลที่สมบูรณ์เก็บไว้ให้ตารางใดตารางหนึ่ง ระบบจะแสดงรายการตารางจากในฐานข้อมูลทั้งหมดมาให้ ผู้ใช้จะต้องเลือกว่าจะเลือกจากตารางที่มีอยู่แล้ว และทำการเลือกชื่อตารางในรายการที่แสดงไว้ให้
- **Create your query string** หากผู้ใช้ต้องการกำหนดรายละเอียดของข้อมูลซึ่งมาจากการผนวกตารางมากกว่าหนึ่งตารางไว้ด้วยกัน ผู้ใช้สามารถเลือกในการกำหนดแบบ Create query string และหลังจากนั้นผู้ใช้งานต้องคลิ๊กที่ปุ่ม Query เพื่อทำการสร้าง Query ที่ต้องการ
- Query คือส่วนที่ผู้ใช้งานกำหนด Query String ตามรูปแบบของ SQL ลงไป และหลังจากนั้นคลิ๊กยังปุ่ม Run เพื่อทำการประมวลผล และตรวจสอบความถูกต้องของ Query ที่กำหนดเมื่อ Query ที่กำหนดถูกต้อง จึงคลิ๊กปุ่ม OK เพื่อกลับเข้าสู้หน้าจอ Item ต่อไป

- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

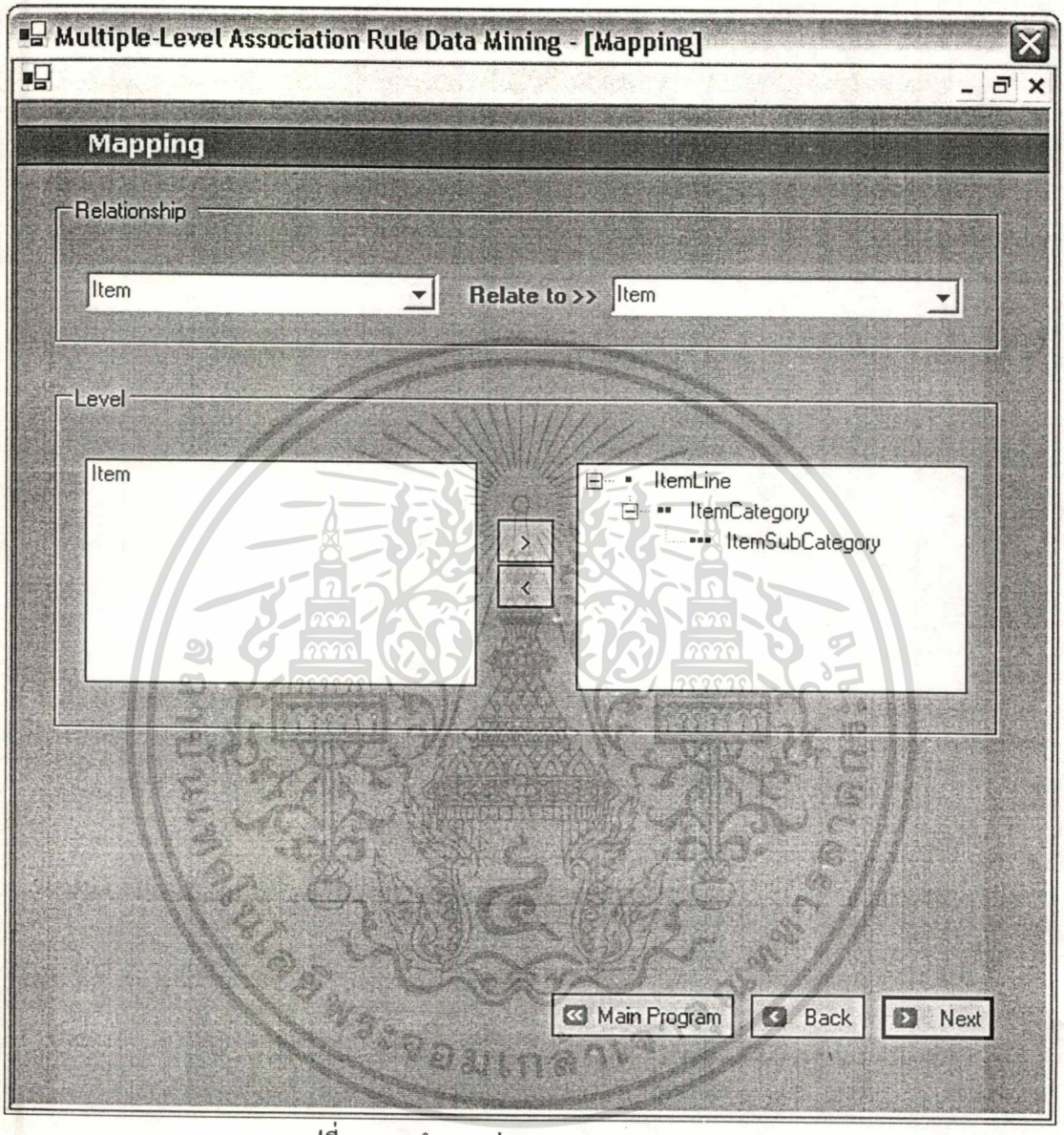
- ปุ่ม **Back** เมื่อผู้ใช้งานต้องการกลับไปยังหน้าจอก่อนหน้านี้
- ปุ่ม **Next** เมื่อผู้ใช้งานทำการกำหนดรายละเอียดครบถ้วนแล้ว ก็จะทำการคลิ๊กที่ปุ่มนี้เพื่อเข้าสู่ขั้นตอนในลำดับต่อไป



รูปที่ 5.7 การกำหนดส่วนของ Item

5.4.5 การกำหนดส่วนของ mapping และ Level เมื่อผู้ใช้งานผ่านในส่วนของการกำหนด Item แล้ว ผู้ใช้งานจะต้องทำการระบุส่วนความสัมพันธ์ระหว่าง Transaction และ Item โดยกำหนดรายละเอียดในการเชื่อมโยงทั้งสองตาราง และการกำหนด Level ให้กับการทำไมนิ่ง ดังรูปประกอบที่ 5.8 ในหน้าจอนี้จะมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.8 การกำหนดส่วนของ Mapping และ Level

- **Relationship** คือส่วนที่ผู้ใช้งานระบุ Field ข้อมูลที่ใช้สร้างความสัมพันธ์ระหว่างข้อมูล Transaction และ Item
- **Level** คือส่วนที่ผู้ใช้งานระบุชื่อของ Field ข้อมูลที่เป็นการจัดหมวดหมู่ โดยผู้ใช้งานจะระบุหมวดหมู่ที่มีขอบเขตใหญ่สุดไว้บนสุด และหมวดหมู่ที่มีขอบเขตน้อยลงในระดับชั้นรองลงมา ตัวอย่างการกำหนดหมวดหมู่ เช่น ข้อมูลสินค้า โดยกำหนดระดับชั้นที่ 1 คือ กลุ่มของสินค้า, ระดับชั้นที่ 2 คือ คุณสมบัติของสินค้า (ขนาด, สี, กลิ่น, รส) , ระดับชั้นที่ 3 คือตราสินค้า เป็นต้น วิธีการในการกำหนดดังนี้
  - ผู้ใช้งานเลือกรายการทางด้านขวามือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

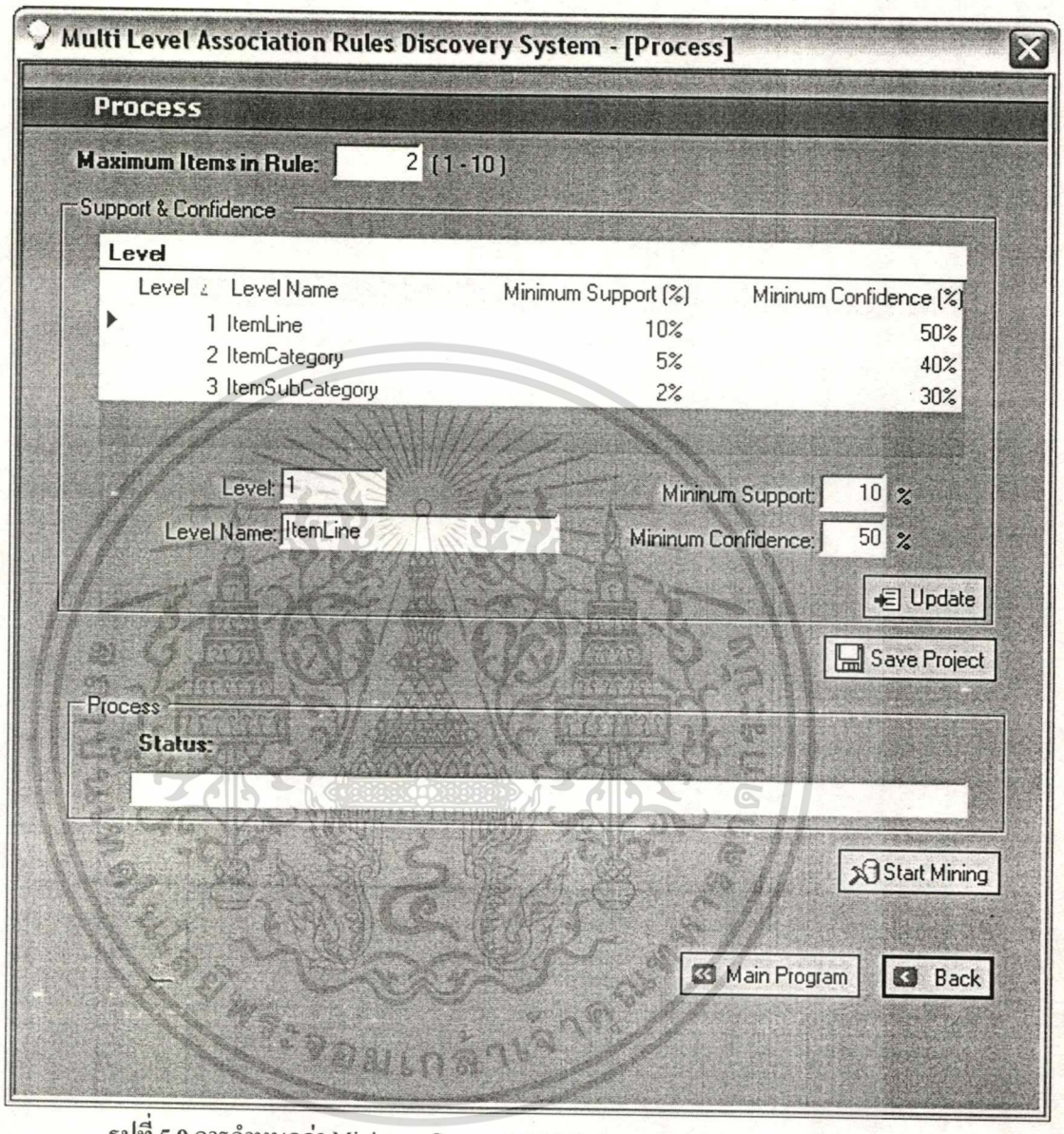
- คลิ๊กปุ่มซึ่งอยู่บนเพื่อทำการจัดระดับชั้น
- หากต้องการลบก็ทำการคลิ๊กปุ่มล่างเพื่อย้ายกลับ
- ข้อมูลจะเรียงลำดับที่เลือกก่อนไว้เป็นระดับชั้นบนสุด

- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป
- **ปุ่ม Back** เมื่อผู้ใช้งานต้องการกลับไปยังหน้าจอก่อนหน้านี้
- **ปุ่ม Next** เมื่อผู้ใช้งานทำการกำหนดรายละเอียดครบถ้วนแล้ว ก็จะทำการคลิ๊กที่ปุ่มนี้เพื่อเข้าสู้ขั้นตอนในลำดับต่อไป

5.4.6 การกำหนดค่า Minimum Support และ Minimum Confidence เมื่อผู้ใช้งานผ่านในส่วนของการกำหนด Level แล้ว จะเข้าสู่ส่วนของการกำหนดค่า Minimum Support และ Minimum Confidence และเริ่มทำการไมนิ่งเพื่อหาผลลัพท์ ดังรูปประกอบที่ 5.9 ในหน้าจอนี้จะมีรายละเอียดดังนี้

- **Max Items in Rule** เป็นการกำหนดจำนวน Item สูงสุดที่ให้แสดงในกฎ
- **Support & Confidence** คือส่วนที่ให้ผู้ใช้งานกำหนดค่า Minimum Support และ Minimum Confidence โดยระบบจะระบุค่าเริ่มต้นให้กับ Minimum Support ในแต่ล่ะ Level คือ Level 1 คือ 50%, Level 2 คือ 10%, Level 3 คือ 4%, Level 4 และ Level ถัดไป คือ 2% และค่าเริ่มต้นของ Minimum Confidence ที่ 50% ในทุก Level โดยผู้ใช้งานสามารถเปลี่ยนแปลงได้ โดยเลือก Level ที่ต้องการเปลี่ยนแปลงในตาราง Level ที่แสดงอยู่ทางขวามือ และคลิ๊กปุ่ม Update เพื่อทำการเปลี่ยนแปลง
- **ปุ่ม Save Project** เมื่อทำการเปลี่ยนแปลงเรียบร้อยแล้ว และต้องการทำการไมนิ่ง ให้ผู้ใช้งานคลิ๊กปุ่ม Save Project เพื่อทำการบันทึก
- **ปุ่ม Start Mining** เมื่อผู้ใช้ต้องการหาผลลัพธ์จากการไมนิ่ง ให้คลิ๊กที่ปุ่มนี้ ระบบจะทำการไมนิ่งหาผลลัพธ์ และทำการแสดงผลให้ผู้ใช้งานทราบต่อไป
- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป
- **ปุ่ม Back** เมื่อผู้ใช้งานต้องการกลับไปยังหน้าจอก่อนหน้านี้

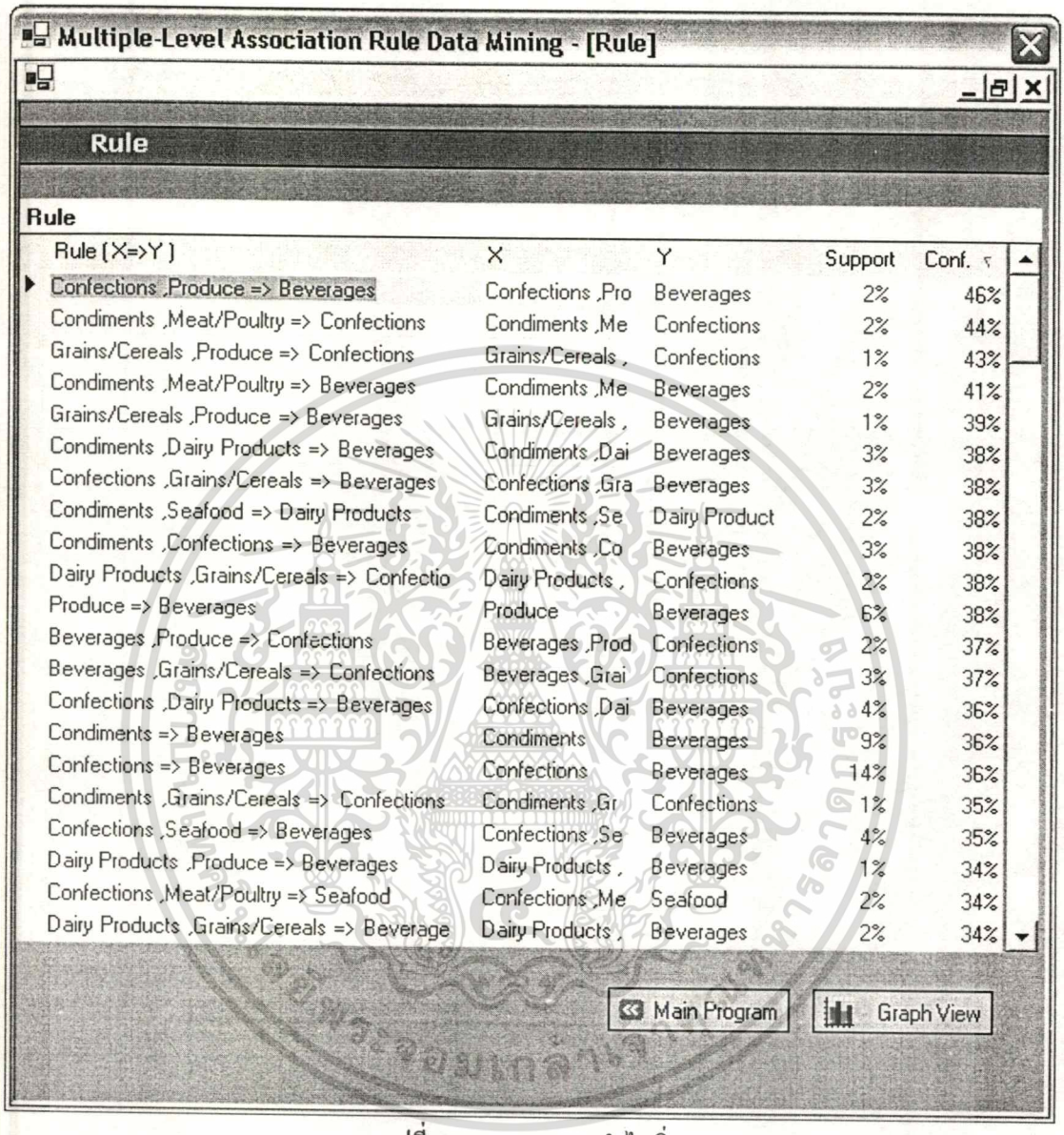
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.9 การกำหนดค่า Minimum Support และ Minimum Confidence

5.4.7 แสดงผลการทำไมนิ่ง เมื่อทำการไมนิ่งแล้วระบบจะแสดงผลลัพท์หรือกฎที่ถูกพิจารณาว่าอยู่ในเกณฑ์ที่ผู้ใช้งานสนใจโดยเทียบจาก Minimum Confidence ดังรูปประกอบที่ 5.10 โดยข้อมูลต่างๆที่แสดงมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



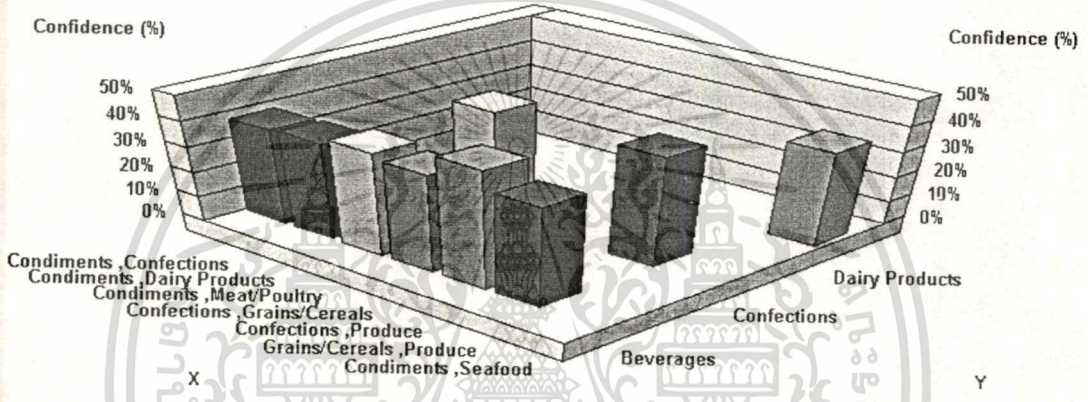
Rule (X=>Y)	X	Y	Support	Conf.
Confections ,Produce => Beverages	Confections ,Pro	Beverages	2%	46%
Condiments ,Meat/Poultry => Confections	Condiments ,Me	Confections	2%	44%
Grains/Cereals ,Produce => Confections	Grains/Cereals ,	Confections	1%	43%
Condiments ,Meat/Poultry => Beverages	Condiments ,Me	Beverages	2%	41%
Grains/Cereals ,Produce => Beverages	Grains/Cereals ,	Beverages	1%	39%
Condiments ,Dairy Products => Beverages	Condiments ,Dai	Beverages	3%	38%
Confections ,Grains/Cereals => Beverages	Confections ,Gra	Beverages	3%	38%
Condiments ,Seafood => Dairy Products	Condiments ,Se	Dairy Product	2%	38%
Condiments ,Confections => Beverages	Condiments ,Co	Beverages	3%	38%
Dairy Products ,Grains/Cereals => Confectio	Dairy Products ,	Confections	2%	38%
Produce => Beverages	Produce	Beverages	6%	38%
Beverages ,Produce => Confections	Beverages ,Prod	Confections	2%	37%
Beverages ,Grains/Cereals => Confections	Beverages ,Grai	Confections	3%	37%
Confections ,Dairy Products => Beverages	Confections ,Dai	Beverages	4%	36%
Condiments => Beverages	Condiments	Beverages	9%	36%
Confections => Beverages	Confections	Beverages	14%	36%
Condiments ,Grains/Cereals => Confections	Condiments ,Gr	Confections	1%	35%
Confections ,Seafood => Beverages	Confections ,Se	Beverages	4%	35%
Dairy Products ,Produce => Beverages	Dairy Products ,	Beverages	1%	34%
Confections ,Meat/Poultry => Seafood	Confections ,Me	Seafood	2%	34%
Dairy Products ,Grains/Cereals => Beverage	Dairy Products ,	Beverages	2%	34%

รูปที่ 5.10 แสดงผลการทำไมนิ่ง

- **Rule (X=>Y)** แสดงผลในรูปแบบของความสัมพันธ์ของ Association Rule ในรูปแบบของ Item X ที่มีผลต่อ Item Y เช่น การซื้อสินค้า X แล้วมีโอกาสในการซื้อสินค้า Y เป็นต้น
- **X** คือ Item X ในกฎเช่นการซื้อขนมปังมีผลในการซื้อนม X จะแทนขนมปัง
- **Y** คือ Item Y ในกฎเช่นการซื้อขนมปังมีผลในการซื้อนม Y จะแทนนม
- **Support** คือ ค่า Support หรือความน่าจะเป็นของการเกิด Item X ร่วมกับ Item Y
- **Conf.** คือ ค่า Confidence หรือค่าความมั่นใจของกฎนี้ ซึ่งหากมีค่าสูงหมายถึงกฎนี้มีความน่าเชื่อถือมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **ปุ่ม Main Program** เพื่อให้ผู้ใช้งานกลับไปยังหน้าแรกของระบบ เมื่อผู้ใช้งานไม่ต้องการจะทำส่วนของการสร้างงานใหม่ต่อไป
- **ปุ่ม Graph View** จะแสดงข้อมูลในรูปแบบกราฟ ระหว่างความสัมพันธ์ของ Item X และ Item Y ของกฎทั้งหมด ดังรูปประกอบที่ 5.11 และ รูปประกอบที่ 5.12



รูปที่ 5.11 Graph View

Item X \ Item Y	Beverages	Confections	Dairy Products
Condiments ,Confections	38%	0%	0%
Condiments ,Dairy Products	38%	0%	0%
Condiments ,Meat/Poultry	41%	44%	0%
Condiments ,Seafood	0%	0%	38%
Confections ,Grains/Cereals	38%	0%	0%
Confections ,Produce	46%	0%	0%
Grains/Cereals ,Produce	39%	43%	0%

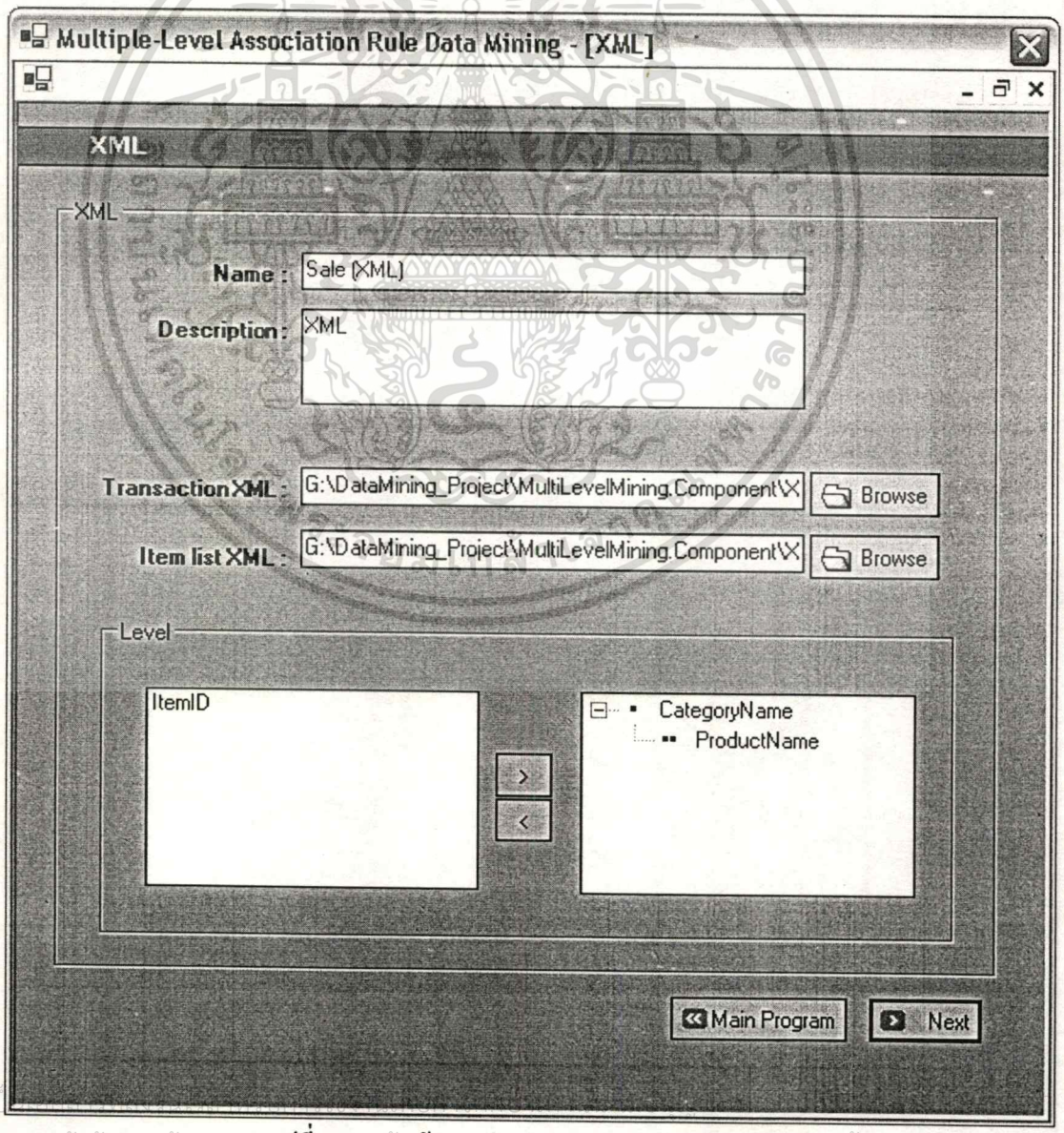
รูปที่ 5.12 ข้อมูลประกอบกราฟ

5.4.8 การสร้างงานใหม่จากเอกสาร XML หากผู้ใช้งานต้องการสร้างงานใหม่โดยใช้แหล่งที่มาจากเอกสาร XML หลังจากผู้ใช้งานคลิ๊กปุ่ม Create New Project จากหน้าจอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักแล้ว ระบบจะแสดงหน้าจอให้ผู้ใช้งานกำหนดรายละเอียดเกี่ยวกับงาน ตามรูปประกอบที่ 5.13 ซึ่งมีรายละเอียดดังนี้

- **Name** ชื่อของงานที่ต้องการระบบ
- **Description** รายละเอียดของงาน เพื่อช่วยอธิบายรายละเอียดของงานที่จะทำการไมนิ่ง
- **Transaction XML** คือส่วนของเอกสาร XML ที่ต้องการใช้ในการทำไมนิ่ง โดยผู้ใช้งานคลิ๊กที่ปุ่ม Browse เพื่อทำการเลือกไฟล์เอกสารหลักที่ต้องการทำไมนิ่ง โดยเอกสารที่ถูกเลือกจะถูกตรวจสอบรูปแบบเอกสารที่ระบบยอมรับหรือไม่ โดยมี Schema ของเอกสาร ดังตารางประกอบที่ 5.6



รูปที่ 5.13 การสร้างงานใหม่จากเอกสาร XML

เอกสารนี้เป็น
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ทำซ้ำ ดัดแปลง ใช้ประโยชน์ หรือเผยแพร่ เนื้อหาของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.6 โครงสร้างของเอกสาร XML ที่ใช้เป็น Transaction

```
<?xml version="1.0" encoding="utf-16"?>
<xs:schema id="Transaction" xmlns="" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:msdata="urn:schemas-microsoft-com:xml-msdata">
 <xs:element name="Transaction" >
  <xs:complexType>
   <xs:choice maxOccurs="unbounded">
    <xs:element name="Transactions">
     <xs:complexType>
      <xs:attribute name="TransactionID" type="xs:string" />
      <xs:attribute name="ItemID" type="xs:string" />
     </xs:complexType>
    </xs:element>
   </xs:choice>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

- **Item List XML** คือส่วนของเอกสาร XML ที่ใช้ในการเชื่อมกับเอกสารในส่วนของ Transaction และนำมาจัดระดับชั้นเพื่อทำการไมนิ่งในแบบ Multiple Level ต่อไป ซึ่งจะไม่ได้กำหนดรูปแบบเอกสารที่แน่นอน แต่จะทำการเชื่องโยงข้อมูลใน Attribute แรกกับ TransactionID ของเอกสาร XML ซึ่งเป็น Transaction
- **ตัวเอย่างเอกสาร XML ในส่วนของ Transaction** ดังตารางที่ 5.7
- **ตัวเอย่างเอกสาร XML ในส่วนของ Item list** ดังตารางที่ 5.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.7 ตัวอย่างของเอกสาร XML ที่ใช้เป็น Transaction

```
<?xml version="1.0" encoding="utf-8" ?>
<Transaction>
        <Transactions TransactionID="10248" ItemID="11"/>
        <Transactions TransactionID="10248" ItemID="42"/>
        <Transactions TransactionID="10252" ItemID="60"/>
        <Transactions TransactionID="10253" ItemID="31"/>
        <Transactions TransactionID="10253" ItemID="39"/>
        <Transactions TransactionID="10253" ItemID="49"/>
        <Transactions TransactionID="10254" ItemID="24"/>
        <Transactions TransactionID="10254" ItemID="55"/>
        <Transactions TransactionID="10254" ItemID="74"/>
        <Transactions TransactionID="10255" ItemID="2"/>
        <Transactions TransactionID="10255" ItemID="16"/>
        <Transactions TransactionID="10255" ItemID="36"/>
        <Transactions TransactionID="10255" ItemID="59"/>
        <Transactions TransactionID="10256" ItemID="53"/>
        <Transactions TransactionID="10256" ItemID="77"/>
        <Transactions TransactionID="10257" ItemID="27"/>
        <Transactions TransactionID="10257" ItemID="39"/>
        <Transactions TransactionID="10257" ItemID="77"/>
        <Transactions TransactionID="10258" ItemID="2"/>
        <Transactions TransactionID="10260" ItemID="57"/>
        .
        .
        .
</Transaction>
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.8 ตัวอย่างของเอกสาร XML ที่ใช้เป็น Item List

```
<?xml version="1.0" encoding="utf-8" ?>
<ItemData>
  <Item ItemID="1" ProductName="Chai" CategoryName="Beverages" />
  <Item ItemID="2" ProductName="Chang" CategoryName="Beverages" />
  <Item ItemID="3" ProductName="Aniseed Syrup" CategoryName="Condiments" />
  <Item ItemID="4" ProductName="Chef Anton's Cajun Seasoning" CategoryName="Condiments" />
  <Item ItemID="5" ProductName="Chef Anton's Gumbo Mix" CategoryName="Condiments" />
  <Item ItemID="7" ProductName="Uncle Bob's Organic Dried Pears" CategoryName="Produce" />
  <Item ItemID="8" ProductName="Northwoods Cranberry Sauce" CategoryName="Condiments" />
  <Item ItemID="9" ProductName="Mishi Kobe Niku" CategoryName="Meat/Poultry" />
  <Item ItemID="10" ProductName="Ikura" CategoryName="Seafood" />
  .
  .
  .
</ItemData>
```

5.4.9 การสร้างงานใหม่จากเอกสาร CSV หากผู้ใช้งานต้องการสร้างงานใหม่โดยใช้แหล่งที่มาจากเอกสาร CSV หลังจากผู้ใช้งานคลิ๊กปุ่ม Create New Project จากหน้าจอหลักแล้ว ระบบจะแสดงหน้าจอให้ผู้ใช้งานกำหนดรายละเอียดเกี่ยวกับงาน ตามรูปประกอบที่ 5.14 ซึ่งมีรายละเอียดดังนี้

- **Name** ชื่อของงานที่ต้องการระบบ
- **Description** รายละเอียดของงาน เพื่อช่วยอธิบายรายละเอียดของงานที่จะทำการไมนิ่ง
- **Transaction CSV** คือส่วนของเอกสาร CSV ที่ต้องการใช้ในการทำไมนิ่ง โดยผู้ใช้งานคลิ๊กที่ปุ่ม Browse เพื่อทำการเลือกไฟล์เอกสารหลักที่ต้องการทำไมนิ่ง รูปแบบของเอการ CSV ต้องมีลักษณะดังตารางที่ 5.9

ตารางที่ 5.9 รูปแบบเอกสาร CSV สำหรับ Transaction

TransactionID	ItemID

- **Item List CSV** คือส่วนของเอกสาร CSV ที่ใช้ในการเชื่อมกับเอกสารในส่วนของ Transaction และนำมาจัดระดับชั้นเพื่อทำการไมนิ่งในแบบ Multiple Level ต่อไป ซึ่ง

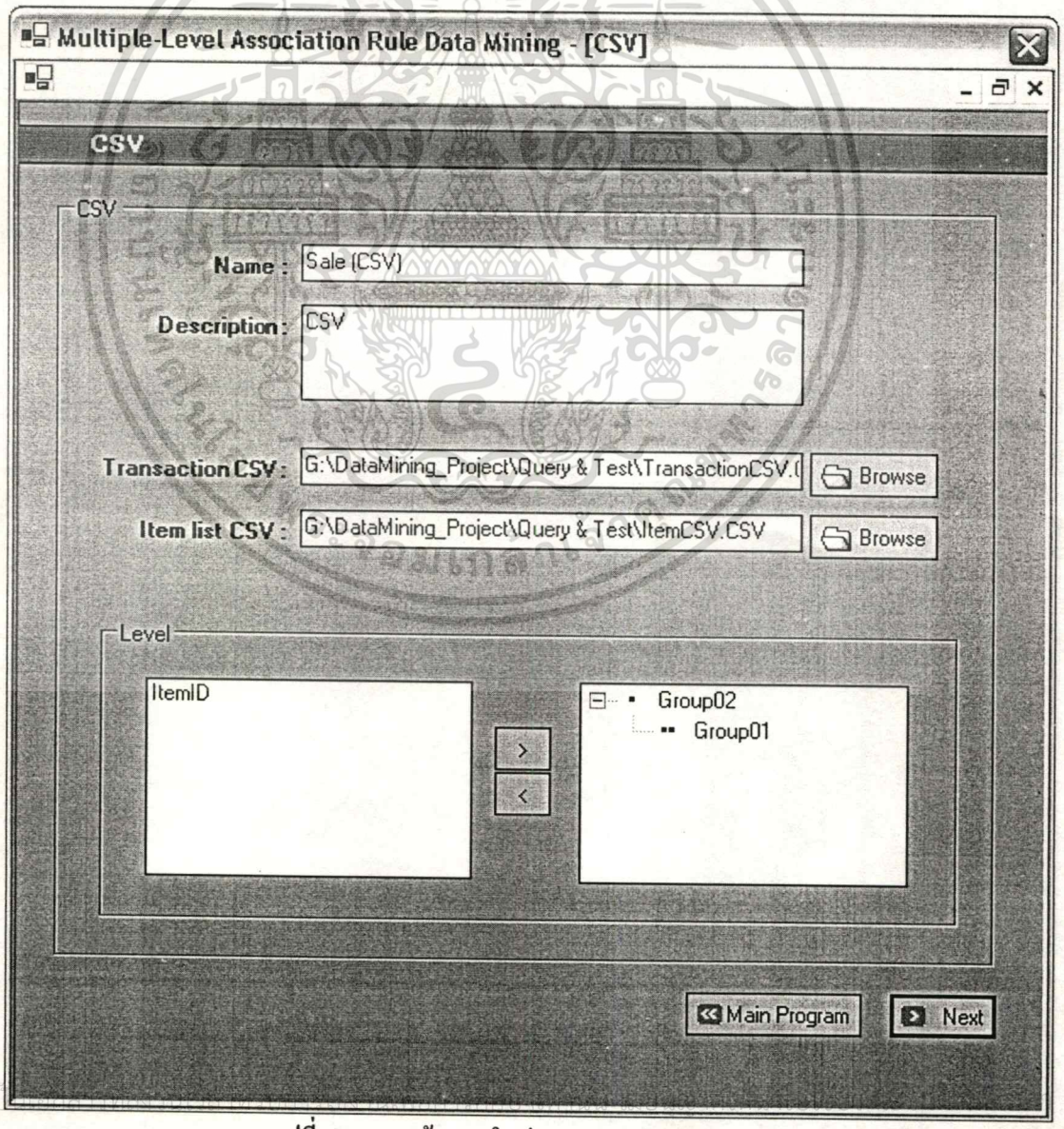
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะไม่ได้กำหนดรูปแบบเอกสารที่แน่นอน แต่จะทำการเชื่องโยงข้อมูลในลำดับแรกกับ ข้อมูลในลำดับที่สองของ Transaction ดังตารางที่ 5.10

ตารางที่ 5.10 รูปแบบเอกสาร CSV สำหรับ Item List

ItemID	Level 1	Level 2	...

- ตัวเอย่างเอกสาร CSV ในส่วนของ **Transaction** ดังตารางที่ 5.11
- ตัวเอย่างเอกสาร CSV ในส่วนของ **Item list** ดังตารางที่ 5.12



รูปที่ 5.14 การสร้างงานใหม่จากเอกสาร CSV

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.11 ตัวอย่างของเอกสาร CSV ที่ใช้เป็น Transaction

```
10285,1
10294,1
10317,1
10348,1
10354,1
10370,1
10406,1
10413,1
10477,1
10522,1
10526,1
10576,1
10590,1
10609,1
10611,1
10628,1
10646,1
10689,1
10691,1
10700,1
10729,1
10752,1
10838,1
10847,1
10918,1
.
.
.
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.12 ตัวอย่างของเอกสาร CSV ที่ใช้เป็น Item List

```
1,Chai,Beverages
2,Chang,Beverages
3,Aniseed Syrup,Condiments
4,Chef Anton's Cajun Seasoning,Condiments
5,Chef Anton's Gumbo Mix,Condiments
6,Grandma's Boysenberry Spread,Condiments
7,Uncle Bob's Organic Dried Pears,Produce
8,Northwoods Cranberry Sauce,Condiments
9,Mishi Kobe Niku,Meat/Poultry
10,Ikura,Seafood
11,Queso Cabrales,Dairy Products
12,Queso Manchego La Pastora,Dairy Products
13,Konbu,Seafood
14,Tofu,Produce
15,Genen Shouyu,Condiments
16,Pavlova,Confections
17,Alice Mutton,Meat/Poultry
18,Carnarvon Tigers,Seafood
19,Teatime Chocolate Biscuits,Confections
20,Sir Rodney's Marmalade,Confections
21,Sir Rodney's Scones,Confections
22,Gustaf's Knäckebröd,Grains/Cereals
23,Tunnbröd,Grains/Cereals
27,Schoggi Schokolade,Confections
28,Rössle Sauerkraut,Produce
29,Thüringer Rostbratwurst,Meat/Poultry
.
.
.
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปการพัฒนาระบบ

6.1 ผลการวิเคราะห์และออกแบบระบบการทำดาต้าไมนิ่งแบบ Multiple Level Association Rule

การพัฒนาระบบสำหรับการทำดาต้าไมนิ่งในแบบ Multiple Level Association Rule ช่วยเพิ่มขอบเขตในการทำดาต้าไมนิ่งจากเดิมที่มีเพียง Level เดียว หรือการทำไมนิ่งแบบ Association Rule ที่ไม่ได้สนใจในเรื่องการจัดกลุ่มของข้อมูลเป็นระดับชั้น ทำให้ผู้ใช้งานเกิดข้อจำกัดในการวิเคราะห์ข้อมูล ดังนั้นระบบจึงนำการไมนิ่งแบบ Multiple Level Association Rule เข้ามาช่วยเพื่อให้สามารถทำการไมนิ่งได้ทั้ง 2 ลักษณะ และมีการรองรับการทำงานร่วมกับฐานข้อมูลต่างๆ โดยเฉพาะฐานข้อมูลซึ่งมีการใช้กันมากเช่น ฐานข้อมูล SQL และสนับสนุนการใช้งานร่วมกับเอกสาร XML และ CSV ทำให้ผู้ใช้งานสามารถนำเข้าข้อมูลจากระบบอื่นๆนอกเหนือจากฐานข้อมูลแบบ SQL เพียงอย่างเดียว ซึ่งการนำเข้าจากเอกสาร CSV มีการใช้กันโดยทั่วไปในหลายๆระบบ และเพื่อรองรับรูปแบบของเอกสารใหม่ๆที่เริ่มมีการใช้งานมากขึ้นในปัจจุบัน ผู้ใช้งานที่มีข้อมูลในรูปแบบของเอกสาร XML สามารถนำมาใช้ในระบบได้เช่นกัน ซึ่งเพิ่มความยืดหยุ่นสำหรับผู้ใช้งานมากยิ่งขึ้น

6.2 ประโยชน์ที่คาดว่าจะได้รับ

- ผู้ใช้งานสามารถทำการวิเคราะห์ข้อมูลจากการทำดาต้าไมนิ่งในแบบ Association Rule ได้จากฐานข้อมูลในระบบต่างๆที่ผู้ใช้จัดเก็บข้อมูลอยู่ โดยไม่จำเป็นต้องทำการส่งออกข้อมูลหรือเปลี่ยนแปลงรูปแบบ ทำได้การทำดาต้าไมนิ่งได้สะดวกมากขึ้น
- สามารถวิเคราะห์ข้อมูลจากการดาต้าไมนิ่งได้ในข้อมูลที่มีการจัดกลุ่มแบบระดับชั้นได้ตามการทำไมนิ่งแบบ Multiple Level Association Rule
- เป็นแนวทางในการนำข้อมูลที่มีอยู่ในระบบมาใช้งานให้เกิดประสิทธิภาพในด้านการวิเคราะห์และนำผลลัพธ์เพื่อใช้ในการปรับปรุงเปลี่ยนแปลง หรือสร้างสิ่งใหม่ๆซึ่งก่อให้เกิดประโยชน์แก่องค์กรต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.3 แนวทางการพัฒนาระบบเพิ่มเติม

- พัฒนาเพิ่มเติมในส่วนของการทำไมนิ่งแบบ Multiple Level Association Rule ในลักษณะต่าง เช่นแบบโครงสร้างระดับชั้นแบบผสม หรือแบบต่างระดับชั้น เพื่อเพิ่มการใช้งานมากยิ่งขึ้น
- เพิ่มประสิทธิภาพในการทำงานเพื่อให้การหาผลลัพธ์ทำได้รวดเร็วยิ่งขึ้น โดยการวิเคราะห์จากอัลกอริทึมอื่นๆที่เกี่ยวข้องกับ Multiple Level Association Rule ในการพัฒนาเปรียบเทียบต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

เรย์โนลด์, แมทธิว. 2545. **คัมภีร์การใช้ Visual Basic .Net ฉบับสมบูรณ์.** กรุงเทพ: ซีเอ็ดยูเคชั่น.

J.Han and Y. Fu. 1995. **Discovery of Multiple-Level Association Rules from Large Database.** Washington D.C.: ACM-SIGMOD

R. Agrawal, T. Imielinski, and A. Swami. 1993. **Mining association rules between sets of item in large database.** Washington D.C.: ACM-SIGMOD

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล	น.ส. เกศินี บุญช่วย
ประวัติการศึกษา	ระดับชั้นประถมศึกษา โรงเรียนหาดใหญ่อำนวยวิทย์
	ระดับชั้นมัธยมศึกษา โรงเรียนหาดใหญ่วิทยาลัย
	ระดับอุดมศึกษา มหาวิทยาลัยสงขลานครินทร์ (บธ.บ)
ประวัติการทำงาน	บริษัท พรีเมียซิสเต็ม เอ็นจิเนียริ่ง จำกัด (2543 - 2545)
	บริษัท ภูมิซอต์ฟ จำกัด (2546 - ปัจจุบัน)
ที่อยู่ปัจจุบัน	19/1 ถ.รามคำแหง ซ.รามคำแหง 12 10240
ที่ทำงาน	บริษัท ภูมิซอต์ฟ จำกัด
E-mail	kasinee_b@yahoo.com, bkasinee@hotmail.com

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้