

**FINDING SUBJECTIVE INTERESTINGNESS PATTERNS
BY USING META-RULES GUIDED MINING
BASED ON ONTOLOGY**



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2006**

**ISBN 974-04-7365-2
COPYRIGHT OF MAHIDOL UNIVERSITY**

Thesis
Entitled

**FINDING SUBJECTIVE INTERESTINGNESS PATTERNS
BY USING META-RULES GUIDED MINING
BASED ON ONTOLOGY**



Preecha Tangworakithaworn
.....
Mr. Preecha Tangworakithaworn
Candidate

Jarernsri L. Mitranont
.....
Assoc. Prof. Jarernsri L. Mitranont, Ph.D.
Major-Advisor

Thanwadee Sunetnanta
.....
Asst. Prof. Thanwadee Sunetnanta, Ph.D.
Co-Advisor

Kash
.....
Prof. M.R. Jisnuson Svasti, Ph.D.
Dean
Faculty of Graduate Studies

Supachai Tangwongsan
.....
Assoc. Prof. Supachai Tangwongsan, Ph.D.
Chair
Master of Science Programme in
Computer Science
Faculty of Science

Thesis
Entitled

**FINDING SUBJECTIVE INTERESTINGNESS PATTERNS
BY USING META-RULES GUIDED MINING
BASED ON ONTOLOGY**

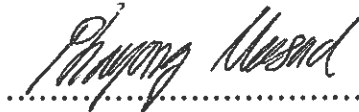
was submitted to the Faculty of Graduate Studies, Mahidol University
For the degree of Master of Science (Computer Science)

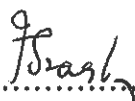
on
May 17, 2006


.....
Mr. Preecha Tangworakitthaworn
Candidate


.....
Assoc. Prof. Jarernsri L. Mitpanont, Ph.D.
Chair


.....
Asst. Prof. Thanwadee Sunetnanta, Ph.D.
Member


.....
Asst. Prof. Phayung Meesad, Ph.D.
Member


.....
Prof. M.R. Jisnuson Svasti, Ph.D.
Dean
Faculty of Graduate Studies
Mahidol University


.....
Prof. Amaret Bhumiratana, Ph.D.
Dean
Faculty of Science
Mahidol University

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to Assoc. Prof. Jarensri L. Mitranont, my major advisor, for her continuous help, encouragement, and support, during my study. She always offers the time for frequent discussions and consultations despite her busy schedule. With her creative thinking and insightful guidance make our discussions fruitful and interesting. The production of this thesis would not have been successful without her kind supervision.

I would also like to thank Asst. Prof. Thanwadee Sunetnanta, for her guidance, valuable advice, encouragement and constructive comments, which are very beneficial for this thesis.

I am grateful to Asst. Prof. Phyung Meesad sitting as the external examiner for his constructive feedback and generous recommendations. Moreover, his invaluable help in patient reading through my thesis draft as well as correcting the writing are very appreciated.

I would also like to take this opportunity to thank to all the people in the Mahidol University Computing Center (MUCC) for their generous help and kind supports.

My thanks also go to many fellow graduate students who made my days at MUCC enjoyable: Chart, Waan, Nam, Kai, and Chin, especially, Chart or Somchart Fugkeaw for his special consultation and revision of the thesis manuscript.

I am very grateful to my mom, who always energizes me with her love, understanding, and support. I am also indebted to my brothers and sisters for their everlasting understanding and encouragement.

For their contribution to this thesis, I dedicate the thesis to my dad who sadly passed away in 1997. In memory of his love, you will be greatly missed, greatly admired, and forever.

Preecha Tangworakitthaworn

FINDING SUBJECTIVE INTERESTINGNESS PATTERNS BY USING META-RULES GUIDED MINING BASED ON ONTOLOGY

PREECHA TANGWORAKITTHAWORN 4437187 SCCS/M

M.Sc. (COMPUTER SCIENCE)

THESIS ADVISORS: JARERNSRI L. MITRANONT, Ph.D.,
THANWADEE T. SUNETNANTA, Ph.D.

ABSTRACT

This research proposes a novel framework to construct the Meta-Rules Guided Mining by applying the ontology to describe the Conceptual Model of the objects in the domain of interest. The proposed framework is an “Extended Version” of traditional data mining in which the focus is on the post analysis of the data mining. The proposed framework consists of two main parts. First, the specialized rules discovery part is responsible for discovering the patterns of data from the prepared database based on the Apriori Algorithm. Second, the generalized rules discovery part is used to discover the “Generalized Rules” defined as the meta-rules patterns which are generated based on the ontology. In order to represent the categorical concepts and relation types of the ontology, the Formal Concept Analysis (FCA) theory was adopted to construct the lattice.

The prototype system called FCA2MR (Formal Concept Analysis to Meta-Rules) was developed as an easy-to-use and step-by-step process generation tool. It was developed to test and verify the idea of the proposed approach. The FCA2MR provides a flexible way to specify the pre-threshold measurement including minimum support and minimum confidence values to discover interestingness patterns. In addition, the post-threshold measurement called minimum precision values was used to determine the strength of the generated meta-rules patterns. In the experiment, three benchmark datasets from the UCI machine learning repository were used to test the prototype and three active datasets were collected and used to evaluate the interestingness results. The system was evaluated in three parts. First, the functional comparison of the ontology generators was conducted to compare the ontology features of three ontology generators, namely, FCA2MR, Protégé2000 and ConImp. Second, the interestingness evaluation showed comparable results between the FCA2MR and the Magnum Opus program. Finally, the user participation evaluation was carried out by using a number of questionnaires to check whether the ranking results reflect the subjective interestingness of the users. The experimental results indicated that this approach is feasible and effective in generating reasonable solutions for subjective interestingness data mining comparable to commercial tools. Moreover, the prototype system gives high accuracy and the interestingness results are relevant to the user’s needs.

KEY WORDS: DATA MINING/ INTERESTINGNESS/
SUBJECTIVE INTERESTINGNESS/
META-RULES GUIDED MINING/ ONTOLOGY/
FORMAL CONCEPT ANALYSIS (FCA)

172 P. ISBN 974-04-7365-2

การค้นหารูปแบบที่มีความน่าสนใจตามอรรถวิสัยด้วยเทคนิคเมต้า-รูล์ ไกด์ ไมนิ่งบนพื้นฐานออนโทโลยี
(FINDING SUBJECTIVE INTERESTINGNESS PATTERNS BY USING META-RULES GUIDED MINING BASED ON ONTOLOGY)

ปรีชา ตั้งวรกิจถาวร 4437187 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : เจริญศรี มิตรภานนท์ Ph.D., ธันวดี สุเนตนันท์ Ph.D.

บทคัดย่อ

งานวิจัยนี้ได้นำเสนอแนวทางการออกแบบสถาปัตยกรรมรูปแบบใหม่ในการสร้างเมต้า-รูล์ ไกด์ ไมนิ่งด้วยการประยุกต์ใช้ออนโทโลยี เพื่ออธิบายถึง "โมเดลเชิงแนวคิด" ของวัตถุในโดเมนที่สนใจ หลักการของงานวิจัยที่นำเสนอนี้ เป็นส่วนขยายต่อยอดของระบบการสืบค้นข้อมูลแบบพื้นฐาน โดยมุ่งเน้นในส่วนของการวิเคราะห์ขั้นตอนสุดท้ายของระบบ งานวิจัยนี้ประกอบด้วยการทำงาน 2 ส่วนหลัก ได้แก่ 1. ส่วนของการสืบค้นรูปแบบลักษณะเฉพาะ ทำหน้าที่ในการสืบค้นรูปแบบที่เกิดขึ้นได้ทั้งหมดของข้อมูลด้วยอัลกอริทึม Apriori และ 2. ส่วนของการสืบค้นรูปแบบลักษณะทั่วไป ซึ่งทำหน้าที่ในการสืบค้น "รูปแบบลักษณะทั่วไป" ของชุดข้อมูลซึ่งเรียกว่า "เมต้า-รูล์" เราได้ทำการออกแบบกระบวนการในการสร้าง เมต้า-รูล์ บนพื้นฐานของออนโทโลยี โดยการประยุกต์เทคนิคที่เรียกว่า Formal Concept Analysis (FCA) มาใช้ในการสร้างออนโทโลยีและใช้อธิบายถึงคอนเซ็ปต์และความสัมพันธ์ของออนโทโลยีในรูปแบบแลททิซ

ในการทดลองและตรวจสอบความสมเหตุสมผลของงานวิจัย เราได้ออกแบบและสร้างระบบจำลอง FCA2MR (Formal Concept Analysis to Meta-Rules) เพื่อใช้เป็นเครื่องมือที่สะดวกต่อการใช้งานและอธิบายการทำงานแต่ละขั้นตอนที่ชัดเจน ซึ่งระบบจำลองนี้ มีความคล่องตัวในการเลือกกำหนดค่าตรวจวัดคุณภาพเบื้องต้น ได้แก่ Minimum Support และ Minimum Confidence เพื่อใช้ในการสืบค้นรูปแบบลักษณะเฉพาะ ในขณะที่เดียวกันสามารถกำหนดค่าตรวจวัดคุณภาพขั้นสุดท้าย ได้แก่ Minimum Precision เพื่อใช้ในการสร้างรูปแบบลักษณะทั่วไป โดยในการทดสอบระบบ เราได้ใช้ข้อมูลมาตรฐาน 3 ชุดจากคลังข้อมูลเพื่อการเรียนรู้ด้วยเครื่องของ UCI และข้อมูลปัจจุบัน 3 ชุดได้จากการรวบรวมจากระบบงานจริง สำหรับใช้ในการวิเคราะห์ระบบ ซึ่งเราได้ทำการวิเคราะห์ใน 3 ส่วน ได้แก่ 1. วิเคราะห์การสร้างออนโทโลยี โดยการเปรียบเทียบลักษณะของออนโทโลยีที่สร้างได้จากเครื่องมือสร้างออนโทโลยี 3 ตัว ได้แก่ FCA2MR, Protégé2000 และ ConImp 2. วิเคราะห์รูปแบบที่มีความน่าสนใจ โดยการเปรียบเทียบผลลัพธ์ระหว่าง FCA2MR และ Magnum Opus และ 3. วิเคราะห์ความต้องการของผู้ใช้ โดยการสร้างแบบสอบถามเพื่อตรวจสอบผลลัพธ์ที่มีอิทธิพลต่อรูปแบบที่มีความน่าสนใจในเชิงอรรถวิสัยต่อผู้ใช้งาน โดยผลลัพธ์จากการทดลองชี้ให้เห็นว่าแนวคิดที่เราได้นำเสนอนั้น สามารถให้ผลลัพธ์ที่มีประสิทธิภาพ ประสิทธิผลและสมเหตุสมผลสำหรับการสืบค้นรูปแบบที่มีความน่าสนใจในเชิงอรรถวิสัย โดยสามารถเปรียบเทียบผลลัพธ์ได้เท่ากับเครื่องมือเพื่อการค้า นอกจากนี้แล้วระบบจำลองที่เราได้พัฒนาขึ้น ยังให้ความถูกต้องแม่นยำที่สูงและให้ผลลัพธ์ที่มีความน่าสนใจ ที่สอดคล้องกับความต้องการ ของผู้ใช้ได้เป็นอย่างดี

172 หน้า. ISBN 974-04-7365-2

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1. INTRODUCTION	
1.1 Introduction	1
1.2 The Organization of the Thesis	4
CHAPTER 2. LITERATURE REVIEW	
2.1 Overview of Knowledge Discovery in Database (KDD)	5
2.1.1 The Process Life Cycle of KDD	6
2.2 Overview of Data Mining (DM)	7
2.2.1 Data Mining Methods	7
2.2.2 Association Rules Mining	7
2.3 Interestingness Approach of Data Mining	9
2.3.1 Objective Interestingness Patterns	9
2.3.2 Subjective Interestingness Patterns	10
2.3.3 Meta-Rules Guided Mining	11
2.4 Ontology	12
2.4.1 Conceptualization	12
2.4.2 Type of Ontology	12
2.4.3 Ontology, Information Science and AI	16
2.4.4 Ontology Based Data Mining	17
2.4.5 Formal Concept Analysis (FCA)	17
2.5 Related Works	18
CHAPTER 3. PROBLEM STATEMENT	
3.1 Problem Statement	20
3.2 Research Objective	21
3.3 Research Contribution	21
3.4 Scope of the Research	21

CONTENTS (CONT.)

	Page
CHAPTER 4. A UNIFIED FRAMEWORK OF META-RULES GUIDED MINING BASED ON ONTOLOGY	
4.1 The Fundamental Concepts of a Unified Framework of Meta-Rules Guided Mining Based on Ontology	23
4.2 The Proposed Framework of the Meta-Rules Guided Mining Based on Ontology	24
4.3 THE FIRST COMPONENT: Association Rules Mining Module	28
4.4 THE SECOND COMPONENT: Ontology Construction Module	28
4.4.1 The Ontology Construction	29
4.5 THE THIRD COMPONENT: Meta-Rules Guided Mining Construction Module	45
4.5.1 Rules Implication on Lattice	45
4.5.2 The Process of Meta-Rules Guided Mining Construction	47
4.6 THE FOURTH COMPONENT: Interestingness Patterns Matching Module	55
4.6.1 The Process of Interestingness Patterns Matching	55
CHAPTER 5. DESIGN AND IMPLEMENTATION	
5.1 The Implementation of A Unified Framework of Meta-Rules Guided Mining Based on Ontology	58
5.2 System Design	59
5.2.1 Structure Chart	59
5.2.2 Data Flow Diagrams (DFD)	60
5.2.3 Entity Relations Diagrams (ER)	67
5.2.4 Database Designed Relations	68
5.2.5 System Defined Relations	70
5.3 Ontology Components Design	72
5.4 FCA2MR System Functions	75
5.4.1 Specialization Rules Mining Function	75
5.4.2 Generalization Rules Mining Function	77

CONTENTS (CONT.)

	Page
CHAPTER 6. EXPERIMENT AND RESULT	
6.1 Scope of the Experiment	89
6.2 Experimental Framework	89
6.2.1 Experimental Setting	91
6.2.2 Experimental Methods	94
6.3 Experimental Results	96
6.3.1 The Data Mining Testing Results	96
6.3.2 The Ontology Generator Testing Results	97
6.3.3 The Interestingness Testing Results	107
6.4 FCA2MR Functional Evaluation	127
6.4.1 The Functional Comparison of Three Ontology Generators	127
6.4.2 Interestingness Evaluation	129
6.4.3 User Participation Evaluation	133
CHAPTER 7. DISCUSSION AND CONCLUSION	
7.1 Discussion and Conclusion	162
7.2 Future Works	165
REFERENCES	166
BIOGRAPHY	172

LIST OF TABLES

	Page
Table 4-1: An Example of the PowerSET of Objects and Attributes	34
Table 4-2: An Example of the Common Attributes (A') Extractor	35
Table 4-3: An Example of the Common Objects (O') Extractor	36
Table 4-4: An Example of the Class Matching Process	37
Table 4-5: An Example of the PowerSET of Formal Concept C	41
Table 4-6: An Example of the Order Pair Generation Process	42
Table 4-7: An Example of the Associated Concept Checking	42
Table 6-1: The Details of Datasets	91
Table 6-2: The Details of Data Model	92
Table 6-3: Summary Results of Three Datasets in Specialization Module	96
Table 6-4: The Summary Results of Ontology Generation Process of Each Data Model	97
Table 6-5: The Summary Results of the Interestingness Testing	107
Table 6-6: The Detail Results (Car_C_1 Domain) of the Interestingness Testing	109
Table 6-7: The Detail Results (Car_C_2 Domain) of the Interestingness Testing	111
Table 6-8: The Detail Results (Car_C_3 Domain) of the Interestingness Testing	113
Table 6-9: The Detail Results (Nursery_N_1 Domain) of the Interestingness Testing	115
Table 6-10: The Detail Results (Nursery_N_2 Domain) of the Interestingness Testing	117
Table 6-11: The Detail Results (Nursery_N_3 Domain) of the Interestingness Testing	119
Table 6-12: The Detail Results (Adult_A_1 Domain) of the Interestingness Testing	121
Table 6-13: The Detail Results (Adult_A_2 Domain) of the Interestingness Testing	123
Table 6-14: The Detail Results (Adult_A_3 Domain) of the Interestingness Testing	125
Table 6-15: The Functional Comparison of Three Ontology Generators	127
Table 6-16: The Summary Results of the Interestingness Comparisons with Magnum Opus	130
Table 6-17: The Summary Results of the User Participation Evaluation of the Library_Q1 Domain	138

LIST OF TABLES (CONT.)

	Page
Table 6-18: The Summary Results of the User Participation Evaluation of the Library_Q2 Domain	141
Table 6-19: The Summary Results of the User Participation Evaluation of the Student_Q1 Domain	147
Table 6-20: The Summary Results of the User Participation Evaluation of the Student_Q2 Domain	150
Table 6-21: The Summary Results of the User Participation Evaluation of the ShortCourse_Q1 Domain	156
Table 6-22: The Summary Results of the User Participation Evaluation of the ShortCourse_Q2 Domain	158

LIST OF FIGURES

	Page
Figure 2-1: An Overview of the Steps of the KDD Process	6
Figure 2-2: Kinds of Ontology, according to their level of dependence on a particular task or point of view.	14
Figure 4-1: Taxonomy of the Meta-Rules Guided Mining	24
Figure 4-2: An Overview Framework of the Meta-Rules Guided Mining Based on Ontology	25
Figure 4-3: An Example of the Itemsets, Specialized and Generalized Rules	27
Figure 4-4: The Cross-Table of the Formal Context K	30
Figure 4-5: An Example of the Formal Context K	31
Figure 4-6: An Example of the Formal Concept C	39
Figure 4-7: The Result of the Formal Lattice Generation Process of the Example 4-2	43
Figure 4-8: The Schematic View of the Ontology Construction Process	44
Figure 4-9: An Example of Rules Implication on Lattice	46
Figure 4-10: An Example of the Basket Transaction Database	48
Figure 4-11: An Example of the Formal Lattice and Its Precision Value	48
Figure 4-12: An Example of the Method to Calculate the Precision of Each Pair of Concept	49
Figure 4-13: A General View of Ontology	51
Figure 4-14: The Schematic View of the Meta-Rules Construction Process	54
Figure 4-15: The Process of Interestingness Patterns Matching	55
Figure 5-1: The Structure Chart of A Unified Framework of Meta-Rules Guided Mining Based on Ontology	59
Figure 5-2: A Data Flow Diagram: The Context Diagram	60
Figure 5-3: A Data Flow Diagram Level 1	61
Figure 5-4: A Data Flow Diagram Level 2 – Process 1: Specialized Rules Discovery	62
Figure 5-5: A Data Flow Diagram Level 2 – Process 2: Generalized Rules Discovery	63
Figure 5-6: A Data Flow Diagram Level 3 – Process 2.1: Formal Concept Analysis	64

LIST OF FIGURES (CONT.)

	Page
Figure 5-7: A Data Flow Diagram Level 3 – Process 2.2: Meta-Rules Guided Generation	65
Figure 5-8: A Data Flow Diagram Level 3 – Process 2.3: Interestingness Pattern Matching	66
Figure 5-9: An Entity Relationship Diagram (E-R)	67
Figure 5-10: Frequent ItemSets Results of Adult DataSET (Min_Supp=5)	76
Figure 5-11: The Association Rules Results of Adult DataSET (Min_Conf=5)	77
Figure 5-12: The Object Initialization Screen	78
Figure 5-13: The Attribute Initialization Screen	79
Figure 5-14: The Cross Table of the Formal Context Generation Screen	80
Figure 5-15: The Generated Formal Context	80
Figure 5-16: The Formal Concept Generation screen	81
Figure 5-17: The Formal Lattice Generation Screen	82
Figure 5-18: Ontology Browser: General View with Formal Concept and Relation Properties	83
Figure 5-19: Ontology Browser: The Extent View with Formal Concept and Relation Properties	84
Figure 5-20: Ontology Browser: The Intent View with Individual Properties	85
Figure 5-21: The Meta-Rules Result Screen (Min_P=5.00 %)	86
Figure 5-22: The Interestingness Patterns Result Screen	87
Figure 5-23: The Interestingness Browser Screen	88
Figure 6-1: An Experimental Framework	89
Figure 6-2: Ontology of the Car_C_1 Domain	98
Figure 6-3: Ontology of the Car_C_2 Domain	99
Figure 6-4: Ontology of the Car_C_3 Domain	100
Figure 6-5: Ontology of the Nursery_N_1 Domain	101

LIST OF FIGURES (CONT.)

	Page
Figure 6-6: Ontology of the Nursery_N_2 Domain	102
Figure 6-7: Ontology of the Nursery_N_3 Domain	103
Figure 6-8: Ontology of the Adult_A_1 Domain	104
Figure 6-9: Ontology of the Adult_A_2 Domain	105
Figure 6-10: Ontology of the Adult_A_3 Domain	106
Figure 6-11: Comparative Interestingness/Discarded Results of Car_C_1 Domain	110
Figure 6-12: Comparative Interestingness/Discarded Results of Car_C_2 Domain	112
Figure 6-13: Comparative Interestingness/Discarded Results of Car_C_3 Domain	114
Figure 6-14: Comparative Interestingness/Discarded Results of Nursery_N_1 Domain	116
Figure 6-15: Comparative Interestingness/Discarded Results of Nursery_N_2 Domain	118
Figure 6-16: Comparative Interestingness/Discarded Results of Nursery_N_3 Domain	120
Figure 6-17: Comparative Interestingness/Discarded Results of Adult_A_1 Domain	122
Figure 6-18: Comparative Interestingness/Discarded Results of Adult_A_2 Domain	124
Figure 6-19: Comparative Interestingness/Discarded Results of Adult_A_3 Domain	126

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, it is practical and easy to discover a huge number of patterns in a large set of databases because many research works have improved the capability of the discovery process in Data Mining (DM). However, many of the discovered patterns are actually useless or uninteresting to users. Furthermore, due to the massive number of discovered patterns, it is difficult for a user to interpret and identify such patterns that are interesting to him/her.

The interestingness issue of DM has increasingly received an attention in such a kind of research area since it not only serves the users for determining the usefulness and understanding of the discovered patterns, but also improves the efficiency and effectiveness of the DM in real-life usage.

In practice, the traditional DM generally discovers the useful and understandable patterns by using the measurements, such as Support and Confidence used as the threshold values in the discovery process [4]. The result of the discovery process, called discovered patterns, usually is enormous due to the large size of databases.

The effective quantitative measure for evaluating these discovered patterns is highly needed. In many cases, it is possible to define a measurement of degree of certainty by using the statistical measurements [4, 10]. Moreover, according to the user's view, which is one of crucial factor for determining the final result, many useful patterns are used subjectively. Notion such as novelty and understandability are much more subjective. Hence, the qualitative measures are utilized concurrently during the post-processing of Knowledge Discovery in Database (KDD) to measure the subjective criteria.

Over the past, the KDD community has developed the discovery framework which aims at discovery the useful or *Interestingness Patterns* [35]. The discovered patterns are classified into two types: Objective Interestingness Patterns and Subjective Interestingness Patterns [11, 35, 37].

First, “*Objective Interestingness Patterns*” is the discovery process, which matches the prepared data together with the objective measurement to purify the result of the DM process. The objective measurements depend only on the structure of a patterns (or data itself) and the underlying data use in the discovery process. This approach is named as the “*Data-Oriented paradigm*”.

Second, “*Subjective Interestingness Patterns*” play a crucial role on the user-criterions focusing. In the last decade, most researchers proposed the methodology to improve the discovery of subjective patterns that are close to user’s need based on main criteria [11, 17, 27, 35, 37]. Those also depend on the class of users who examine the patterns, based on their experience or background knowledge. This approach is named as the “*User-Oriented paradigm*”.

The subjective measurements are classified into two main categories, namely, *Actionability* and *Unexpectedness*. A pattern is called an “*Actionability*” interestingness pattern if the user can do something on it to his/her advances. However, a pattern is called an “*Unexpectedness*” interestingness pattern if it is surprising to the user.

The research questions are motivated by the real-life usage of DM applications that aim to serve the user’s need as the main focus. Our research idea is inspired by the need of DM application development targeting on the user-based needs. We propose and develop the DM Framework to discover the interestingness patterns of DM, especially the Subjective Interestingness Patterns in terms of an actionability category.

We proposed a novel framework to construct the Meta-Rules Guided Mining by applying the Ontology. The proposed framework consists of two main parts. First, the specialized rules discovery part, which serves for the crucial task of data mining to discover the discovered patterns of data from prepared database. In contrast to the traditional data mining application, this part is generated once for discovering all the patterns that might be seen in the database. Second, the generalized rules discovery

part is used to discover the “Generalized Rules” defined as the meta-rules patterns. This part is called an “*Extended Version*” of the traditional data mining to discover the interestingness patterns. We design the mechanism to generate the meta-rules patterns based on ontology. According to the formal concept analysis (FCA) theory, we adopt the FCA technique to construct the ontology and represent in the graphical representation style.

For the implementation, we developed the prototype system called FCA2MR (Formal Concept Analysis to Meta-Rules) to serve as an easy-to-use and step-by-step process generation tool. The prototype system was developed to test and verify the idea of our proposed approach.

In the experiment, three benchmark datasets from the UCI machine learning repository were used to test the prototype system and three active datasets from Mahidol University were used to evaluate the interestingness results. We set up the FCA2MR functional evaluation in three parts. First, the functional comparison of the ontology generator is conducted to compare the ontology features of three ontology generators, namely, FCA2MR, Protégé2000 and ConImp. Second, the interestingness evaluation shows the comparison results between the FCA2MR and the Magnum Opus program. Finally, the user participation evaluation is conducted to carry out a number of questionnaires to check whether the ranking results do reflect people’s intuitions of subjective interestingness.

The test and evaluation of experimental results indicate that our approach is feasible to implement and serve as a reasonable solution for subjective interestingness data mining tool comparable to the commercial tools and acceptable from the users.

In summary, the main contributions of our proposed approach consist of four issues: First, the proposed framework is a generic framework that is applicable for various application domains. Second, our proposed approach stimulates the using of new technology in the data mining research area. The ontology is the emerging research trend in AI and data mining. In addition, the Formal Concept Analysis (FCA) is the mathematical founded technique which can apply in the data mining also. These techniques are found to be applicable and efficient for knowledge discovery in database

(KDD). Moreover, the meta-rules guided mining is the recent technique to produce the interestingness of data mining. Third, the prototype system provides the user friendly interface that allows user to analyze the data mining process as the step-by-step fashion. Finally, the benefits of the outstanding features of the FCA technique can serve the process of ontology construction automatically.

1.2 The organization of the Thesis

This thesis is organized in seven chapters as follows:

Chapter 1: Introduction

This chapter introduces the overview of this thesis.

Chapter 2: Literature Review

This chapter provides the necessary background concepts of the data mining, the interestingness approach of data mining, and the ontology. In addition, the related works to this thesis are discussed.

Chapter 3: Problem Statement

This chapter describes the problems of the existing approaches. Moreover, the objectives and contributions of the work are summarized.

Chapter 4: A Unified Framework of Meta-Rules Guided Mining Based on Ontology

This chapter introduces the proposed framework to construct the Meta-Rules Guided Mining by using the Ontology. It describes all components constituting the framework. Then, the mechanisms used in the designed process are scrutinized in depth.

Chapter 5: Design and Implementation

This chapter explains the design and implementation of the FCA2MR prototype system.

Chapter 6: Experiment and Result

This chapter summarizes the details of the experimental and its results.

Chapter 7: Discussion and Conclusion

This chapter discusses and concludes the overall research solution. Also, the interesting future works are recommended.

CHAPTER 2

LITERATURE REVIEW

In this chapter, the concepts of knowledge discovery in database (KDD), data mining, interestingness approach of data mining, meta-rules guided mining, and ontology are introduced. Then many related works for our research are reviewed.

The knowledge discovery in database or KDD is introduced in section 2.1. In section 2.2 and 2.3, we introduce the KDD framework and data mining, respectively. Then, interestingness approach is illustrated in section 2.3, together with the approach of meta-rules guided mining in section 2.4. We review the ontology in section 2.5. Finally, the related works are also discussed in the last section of this chapter.

2.1 Overview of Knowledge Discovery in Database (KDD)

The knowledge discovery in database (KDD) is known as a sub-field of artificial intelligence. The term “KDD” also describes the whole extraction process of knowledge from the data. It involves the identification of potentially useful and understandable patterns in the data. The knowledge is discovered by means of “*Patterns or Rules*”. The patterns show the association between data that closely related in the way of statistical definition. Fayyad et al. [10] defined the meaning of KDD as the following:

“KDD is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data”

The “*data*” is a set of facts and “*patterns*” that are described as an expression used to represent a model applicable to that data. Hence, the extracted patterns are designated to fit a model to data, or to find the structures from data [10].

2.1.1 The Process Life Cycle of KDD

Most researchers recommend that the process of KDD can produce as the interactive process. Fayyad et al. [10] described the whole process of KDD that consists of five stages: data selection, data preprocessing, data transformation, data mining, and interpretation/evaluation, respectively.

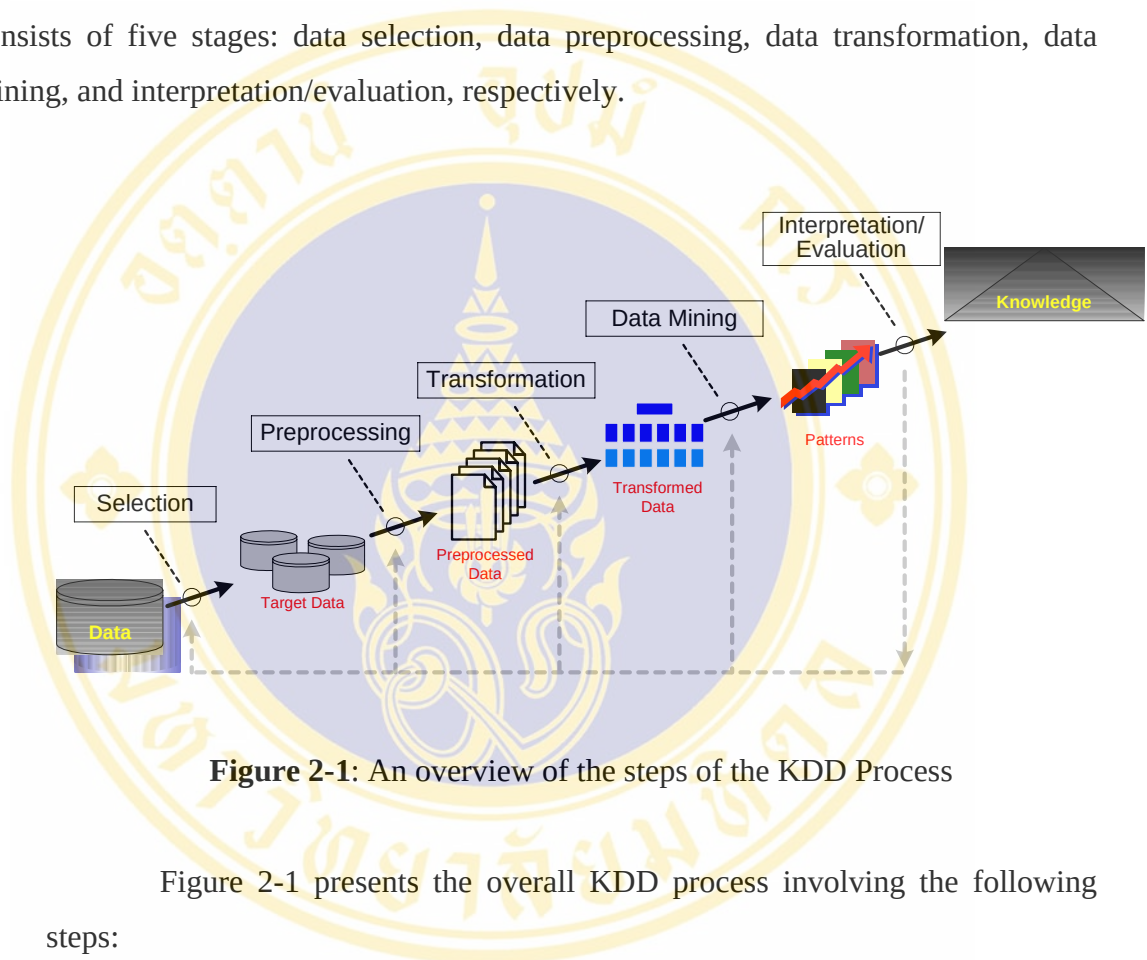


Figure 2-1: An overview of the steps of the KDD Process

Figure 2-1 presents the overall KDD process involving the following steps:

- Data Selection: to create a target data set from large database.
- Data Preprocessing: to prepare the cleaned data by reducing noise over the data. This step consists of several sub processes, such as, removing noise, handling the missing data fields, determining the corrected domain of data, and so forth.
- Data Transformation: to reduce and project the useful data set that suitable for the main criteria.
- Data Mining: to search for patterns of interest in a particular representational form (e.g., decision tree, association rules, etc.).

- **Data Interpretation and Evaluation:** includes interpreting the discovered patterns, and analyzing the patterns automatically and semiautomatically to identify the truly interesting/useful patterns for the user.

2.2 Overview of Data Mining (DM)

Data Mining (DM) is strongly noted to be the real discovery process of the KDD. The DM is a pivotal step in the KDD that aims at the discovery of new patterns from the prepared data. Fayyad et al. [10] defined the meaning of DM as follow:

“Data Mining is a step in the KDD Process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data.”

2.2.1 Data Mining Methods

The DM involves repeated iterative application of particular data mining methods. Most DM methods are based on tried and tested techniques from machine learning, pattern recognition, and statistical techniques. The statistic methods are commonly used and play an essential role in the traditional DM. Three typical statistical techniques are:

- (1) Classification used to learn a function for classifying data items into one of several predefined classes,
- (2) Clustering used to identify a finite set of categories or clusters to describe the data, and
- (3) Regression used to learn a function, which maps a data item to a real-valued prediction variable and the discovery of functional relationships between variables.

2.2.2 Association Rules Mining

Association Rules Mining is one of the most popular DM Techniques. It is particularly useful for discovering the relationship among data in huge databases.

Association Rules Mining is used to discover association of patterns amongst items in terms of statistical definition. The statistics play an important role to discover the patterns. The most summarized values of instances in items are primarily obtained the patterns from its data.

Nanavati et al. [27] defined the formal model of association rules mining as the following:

“Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of literals, called items. D is the set of transactions where each transaction T is a set of items such that $T \subseteq I$. We say that a transaction T contains A , a set of items (itemset) in I , if $A \subseteq T$. An Association Rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$.”

Furthermore, an association rule is an implication $A \Rightarrow B$, where A and B are itemsets. The intuitive meaning of each rule can be determined by the transactions that containing item A always have item B . The rule $A \Rightarrow B$ holds with confidence value and support value to represent the strength of the rules [10].

Statistically, the “Support Value” describes the percentage of transactions that contain all of the items. The support value for a rule $A \Rightarrow B$ is the support for the set of items $LHS \cup RHS$ (LHS, referred to as the Premise of the Rules, which stands for “Left Hand Side” and RHS, referred to as the consequence of the Rules, which stands for “Right Hand Side”). That means the LHS is occurred together with the RHS at the same time. The formula used for calculation is as follows:

$$\begin{aligned}
 S &= \frac{\text{N Transactions involving all items in LHS and RHS of this Rule}}{\text{N total number of transactions}} \\
 &= \frac{\text{Count}(A \cup B)}{\text{Count}(N)} \times 100
 \end{aligned}$$

For the “Confidence Value”, it is defined as the percentage of transactions that contain A also contain B . It indicates the degree of correlation between LHS and RHS. The formula used for calculation is as follows:

$$C = \frac{\text{N transactions involving all items in LHS and RHS of rule}}{\text{N transactions involving all items in LHS}}$$

$$= \frac{\text{Count}(A \cup B) \times 100}{\text{Count}(A)}$$

$$\text{Or } \frac{\text{Support}(A \Rightarrow B) \times 100}{\text{Support}(A)}$$

The confidence denotes the strength of the implication, while the Support indicates the frequencies of the occurring patterns in the database. Normally, the traditional DM. process aims at using these two measurements as the objective criteria.

2.3 Interestingness Approach of Data Mining

Normally, it is very easy to discover a huge number of patterns in a large set of database. Most of these discovered patterns are actually useless or uninteresting to the user. However, it is difficult for a user to interpret and identify the numerous patterns that are interesting to him/her.

The discovered patterns are given into two ways of interesting: objective interestingness patterns and subjective interestingness patterns [21, 22, 33, 34].

2.3.1 Objective Interestingness Patterns

One approach to define the interestingness of a pattern uses the objective terms. The discovery process matches the prepared data together with the objective measurement to refine the result of the DM. The objective measurement depends only on the structure of a pattern (or data itself) and the underlying data used in the discovery process. This approach is known as the “*Data-Oriented paradigm*”.

Zhengxin Chen [52] described two important measures of traditional objective measurement, namely, support and confidence (referred to session 2.2.2 for more details). Traditional DM process accepts the Min_Supp. (Minimal Support Value) and Min_Conf. (Minimal Confidence Value) as the primarily user-criterion from the users.

The DM actually discovers the patterns that exceed these two measurements. The discovered patterns are also called the objective interestingness patterns.

Moreover, Klemettinen et al. [18] observed that the objective measurements of interestingness are not sufficient in many DM applications. Ability of this approach to discover a large number of strong rules is objectively interesting but it is difficult for users to interpret and identify the results in real usage. Thus, the latter approach shed some light on the practical usage. It will be discussed in the next section.

2.3.2 Subjective Interestingness Patterns

The latter approach is the subjective interestingness patterns which plays a crucial role on the user-criterion focusing. In the last decade, most approaches [18, 21, 22, 33, 34] proposed methodology to discover the subjective patterns that close to user's needs as the major factor. The subjective interestingness patterns also depend on the class of users who examine the patterns. This approach is named as the “*User-Oriented paradigm*”.

The subjective interestingness patterns deal with profit-driven actions required by business decision-making or users [47]. A user might find a pattern interesting due to various reasons—some of which may be difficult to articulate. In some case, a particular behavior in a domain might be interesting. The same behavior exhibited in another domain may not be interesting. Thus, interestingness is a domain and user dependent.

The subjective measurements are classified into two main categories: [33, 34]

- *Actionability*: a pattern is interesting if the user can do something on it to his/her advances. The actionability measure is an important subjective measurement of interestingness because users mostly interested in the discovered patterns that allow them to do their jobs better by taking some action in response to the new knowledge.
- *Unexpectedness*: a pattern is interesting if it is surprising to the user. According to the user-expectations, the new knowledge is surprising if and only if those knowledge contradicted their expectations.

Sigal Sahar [30] proposed a technique to discover the subjective interestingness patterns by eliminating the large families of discovered patterns that are not interesting, while the proposed approach is required very low-intensity user interaction. However, this work does not focus on two approaches of subjective interestingness: Unexpectedness and Actionability.

In [34], the unexpected patterns are defined in terms of a belief system that the user has. Interestingness of a pattern is expressed in terms of how it affects the belief system.

2.3.3 Meta-Rules Guided Mining

The meta-rules guided mining is an interactive and iterative approach in which user can probe the data under analysis by specifying hypotheses in the form of meta-rules or pattern-templates [14]. *The main reasons of interactive and iterative of meta-rules are to gather the user-hypotheses, as the expected criteria that are represented in the form of meta-rules.* This approach attempts to confirm the hypotheses by matching the data with the given meta-rules.

The meta-rule is a set of rules used to find generalized rules that are interesting to users. It is a guidance of finding generalized level of association rules. Kamber et al. defined the meaning of the meta-rules as follows [14, 15]:

“A Meta-Rule is a rule template of the form

$$P_1 \wedge P_2 \wedge \dots \wedge P_m \Rightarrow Q_1 \wedge Q_2 \wedge \dots \wedge Q_n$$

where P_i ($i = 1, 2, \dots, m$) and Q_j ($j = 1, 2, \dots, n$) are either instantiated predicates or predicate variables, and $p = m + n$ is the number of predicates in the Meta-Rules. A rule, R , compiles with a Meta-Rule, M_R , if and only if it can be unified with M_R .”

Shen and Leng [31] described the Meta-Rules as the “Second-Order Logical Expression, SOL” in a language L that describe a type of pattern to be discovered. However, the naming of their work is defined as the “Metapattern”.

Klemettinen et al. [18] used the term “*Rule Template*” as the same meaning of Meta-Rules for defining the structure of interesting rules. Moreover, the proposed technique use the Rule Templates to describe the form of interesting rules, and also to specify which rules are not interesting.

The Meta-Rules Guided Mining approach makes the discovery process more efficient by using the Meta-Rules to constrain the rule search space. The Meta-Rules Guided Mining makes the search focused on desired rule patterns that close to user’s needs. It is a new approach which is promising for integrating with the data mining to discover the interesting patterns.

2.4 Ontology

Ontology is a science, as a branch of philosophy that states the being or existence, which describes the structure, relation, property, event and process. It covers all area of the truth or reality. This section describes conceptualization, types of ontology, and the related principles such as information science, artificial intelligence and ontology based data mining.

2.4.1 Conceptualization

Ontology, a challenging topic in Artificial Intelligence, has been widely studied in order to describe the “*Conceptual Model*” of the things in the interesting domain. Ontology conceptualizes all concepts for describing the relationships of interesting things and emphasizes on categorized things by giving the fine definitive and exhaustive classification of the existence in the domain of interest.

2.4.2 Type of Ontology [37]

Several different types of ontology have been proposed. The state of the art concerning ontology can be described by considering the following aspects: ontology theoretical foundations, methodologies for constructing ontology, ontology representation, and environments for ontology development process. In the following, the types of ontology are discussed.

2.4.2.1 SOWA Ontology

J. F. Sowa [37] classified the type of ontology into two categories:

1. Informal Ontology

The informal ontology describes the “Catalog of types”, which is defined or undefined in the Natural Language format.

2. Formal Ontology

The formal ontology states that the concept and relation types are categorized or organized in the subtype-supertype view such as type-subtype relationship or taxonomy. This view can be classified as follows:

- **Terminological Ontology:** An ontology whose categories need not be fully specified by axioms and definitions. An example of a terminological ontology is WordNet, whose categories are partially specified by relations such as subtype-supertype or part whole, which determine the relative positions of the concepts with respect to one another but do not completely define them. (<http://www.cogsci.princeton.edu/~wn/>).
- **Prototype-based Ontology:** It is a kind of Terminological Ontology but utilize the typical instances or prototype for matching to derive the new subtype.
- **Axiomatized Ontology:** The division of subtype is mainly based on the axiom of each concept and the relation inside the ontology itself.
- **Mixed Ontology:** It is an ontology constructed from the combination of one or more ontology types listed above.

2.4.2.2 GUARINO Ontology

Guarino [12] described the different kinds of Ontology according to their level of generality, as shown in Figure 2.2.

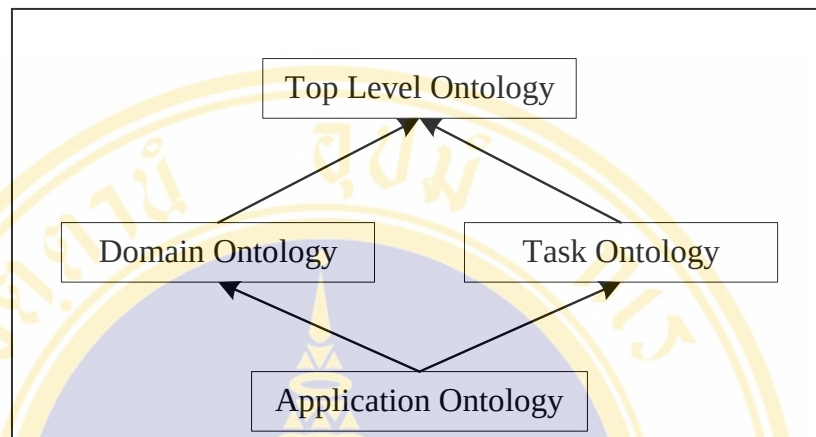


Figure 2-2: Kinds of Ontology, according to their level of dependence on a particular task or point of view. Thick arrows represent specialization relationships.

1. Top-Level Ontology

Top-Level Ontology describes very general concepts of a real world domain like space, time, matter, object, event, action, etc., which are independent of a particular problem or domain. It seems therefore reasonable, at least in a theory, to have a unified top-level ontology for large communities of users.

2. Domain Ontology and Task Ontology

Domain ontology describes the vocabulary related to a generic domain (like medicine or automobiles), while Task ontology explains the generic task or activity (like diagnosing or selling), by specializing the terms introduced in the top-level ontology.

3. Application Level Ontology

Application ontology describes concepts depending on both a particular domain and task. The concepts often correspond to roles played by domain entities while performing a certain activity, like replaceable units or spare components.

2.4.2.3 JURISCICA Ontology

Juriscica et al. [12] classified the concepts used for knowledge representation into four broad ontological categories:

1. Static Ontology

Static Ontology describes static aspects of the world, i.e., what things exist, their attributes and relationships. This ontology is not trivial. For certain applications it is useful to distinguish between different modes of existence for entities, including physical existence.

2. Dynamic Ontology

Dynamic Ontology describes changing aspects of the world. Typical primitive concepts include state, state transition, and process.

Temporal information is needed when describing dynamic worlds. A temporal ontology can be based on time points and associated relations. An event can be represented as a single time point or two time points. Relations such as before or after can be used to relate individual points.

Causality is a concept that is closely related to time in ontology. Causality imposes existence constraints on events. For instance, if event A causes event B and A has been observed then B can be expected as well, possibly with some time delay.

3. Intentional Ontology

Intentional Ontology encompasses the world of motivations, intents, goals, beliefs, alternatives, choices, etc. An intentional ontology allows alternate realities to be expressed and reasoned about. The subject of agents having beliefs and goals and being capable of carrying out actions has been studied extensively.

4. Social Ontology

Social Ontology covers social settings, organizational structures, or shifting networks of alliances and interdependencies. Traditionally,

social ontology has been characterized in terms of concepts such as actor, position, role, authority, commitment, etc. Social ontology is also of interest in distributed artificial intelligence. Some of the concepts have been formalized using specialized logic.

2.4.3 Ontology, Information Science, and Artificial Intelligence (AI)

In Information Science and Artificial Intelligence (AI), ontology is identified as the existence in the explicit formal specification. The theory states that we can represent all of existence or the knowledge in the interesting domain by identifying the declarative language in order to describe set of object or called “*Universe of discourse*”.

To reduce the complication of interesting domain, some simplification can be introduced by limiting the set of objects under the consideration. This limitation is called the universe of discourse or the *domain of individuals*.

Ontology is not only a technique but also includes all concepts for describing the relation of interesting things in the domain or universe of discourse, by focusing on the set of representation terms. Thus, the set of representation terms is similar to dictionary or glossary but more details and structures available for processing various contents [39]. Furthermore, terms are determined by the relation of entities in the universe of discourse; the main components are such as Components Concept (Class), Relation, Instance, Axioms, organized in a hierarchical structure, which are easy to understand and to develop in the computer system.

Particularly in the area of AI, ontology is emphasized to categorize things by giving the fine definitive and exhaustive classification of the existence. The Competency Questions [11] or Ontological Commitments [35] is set to limit scope of the interesting domain, for example;

- What classes of entities are needed for a complete description and explanation of all the going-on in the universe?
- What classes of entities are needed to give an account of what makes true all truths?

The answer should also be dedicated enough to describe all kinds of entities in the interesting domain.

2.4.4 Ontology based Data Mining

DM aims at discovering the associations amongst items in a large set of data. The knowledge or discovered patterns are then modeled into knowledge base for further decision support or question-answering in order to provide common “*Semantic*” platform [43].

The most effective way to build a “*Semantic-Based*” for knowledge discovery is “*Ontological Modeling*”. Ontology arises on how to represent a list of terms and relationship among the concepts in the domain of interest.

Ontology based Data Mining is the new era in the area of DM. This approach has been used to utilize the semantic in term of Semantic-Based for further decision on discovered patterns. Ontology is closely linked with the rules mining method, especially the Association Rules Mining (as described in section 2.2.2). Furthermore, Ontology also provides an important source of background knowledge or hypotheses that can be accepted from the experts or users, and then conceptualize them in terms of concepts of the interesting domain.

2.4.5 Formal Concept Analysis (FCA)

The FCA technique explicitly represents information about the concepts of the ontology [13]. The FCA is a mathematically founded technique, which is based on the “*Lattice Theory*”. The FCA has been extensively used for class hierarchy design and management.

A concept of the FCA is considered as a “*Unit of Thought*” constituted of two parts, namely, “*Extension*” and “*Intension*”. The extension is represented in terms of “*Objects*” in the domain of interest, while the intension is represented in terms of “*Attributes*”.

The FCA technique starts with binary relations, which is called the “*Formal Context*”. The formal context conceptualizes the relationship between objects and

attributes. Every finite binary relation has a finite set of concepts. The set of concepts form the “*Formal Concept*” of the ontology, and then the formal concepts are organized in term of class hierarchy and form the “*Formal Lattice*” of the ontology.

This implies that the FCA guarantees the “Lossless” of information, because every binary relationship between certain objects always has the relationship with certain attributes. Furthermore, every binary relation, a lattice can be constructed.

Theoretically, when the concept lattice is completely formulated, it can always be converted to the original relation, as the same way of the “Fourier Analysis” [36].

2.5. Related Works

The research issue of Interestingness of discovered patterns has been discussed in general by Silberschatz et al. [34]. They also identified the subjective measure of interestingness, especially concentrate on the unexpected measure. The proposed method is based on the beliefs, which are distinguished between hard and soft belief. The definition of hard belief is defined as the “constraints” that the belief cannot be changed when new evidences are occurred, while the soft belief means that the new evidence can induce the change of belief. The approach of Silberschatz focused on representing the degree of belief in form of Bayesian terms and determining the patterns intuitively. The more the patterns “affect” the belief system, the more interesting they are.

Bing Liu et al. [21] proposed a technique of subjective interestingness, called the user-expectation method. In this technique, the user is first asked to provide his/her expected patterns according to his/her part of knowledge or intuitive feelings. Given these expectations, the system uses a fuzzy matching technique to match the discovered patterns against the user’s expectations and ranks the discovered patterns according to the matching results. The method is efficient and allows the user to identify interesting patterns incrementally and interactively. However, Bing Liu’s work presents the user-expectations in the same way of Silberschatz’s work, which is to capture the background knowledge from the user.

Klemettinen et al. [18] proposed a simple formalism of “Rule Template” to easily describe the structure of interesting and uninteresting classes of patterns. The

rule templates are divided into two categories: “Inclusive Template”, and “Restrictive Template”. To be interesting, a rule has to match an inclusive template. However, if a rule matches a restrictive template, it is considered uninteresting. Moreover, the prototype system supports only the small rule sets and a remaining problem is a redundancy.

Kamber et al. [14, 15] presented the other approach conforming to the Interestingness of discovered patterns by construction the Meta-Rules Guided Mining by using Data Cubes, which are applied with the multi-dimensional association rule mining to explore a “data cube” structure. The proposed methodology is based on a given Meta-Rules Guided containing p predicates (p -D), to compare mining on n -dimensional (n -D) cube structure ($p \leq n$, where p is the number of predicates; n is the number of dimension cube). The proposed algorithms examine two cube-based algorithms for mining on a data cube, namely, multi-D-slicing algorithm, and n -D cube search algorithm. After an appropriate n -D cube is available, two “abridged construction” algorithms are designed to integrate with cube computation, namely, abridged n -D cube construction, and abridged multi- p -D cube construction.

The main idea of Kamber’s work partially inspire our research in applying the concept of Meta-Rules Guided with the Interestingness issue of DM.

CHAPTER 3

PROBLEM STATEMENT

This chapter firstly explains the problem related to the knowledge discovery, particularly in the interestingness of data discovered from data mining application. The described problem inspires the production of this research. Hereafter, the research objectives, research contributions, and scope of the research are formally presented.

3.1 Problem Statement

A large Amount of discovered patterns from the data mining process is not always be applicable to the users. Some of them are actionable, however, most of them are discarded. The discovered patterns are close to user's needs if they are interesting.

Due to a huge volume of database, the data mining process produces a large amount of discovered patterns as the results. The important point is when a number of discovered patterns is large, post analysis becomes difficult.

A number of discovered patterns and the difficulty of the post analysis are the main problem of our research. We emphasizes on how to obtain the interestingness of the large number of discovered patterns.

In the last decade, the interestingness of discovered patterns are on going research in the data mining area. Many researchers propose methodologies to identify the interestingness patterns by focusing on two ways: Objective Measure, and Subjective Measure [34]. For the objective measure, the statistical measurement are proposed, such as, Lift [19], Influence and Conditional Influence [6], etc. On the other hand, the subjective measure deals with the proposed framework, which aims at the user's needs as the main methodology.

In our research, we emphasize on how to generate the subjective interestingness patterns and select the interesting ones from them. Many researchers propose the method of capturing the domain knowledge from users [18, 23], however, they aim at gathering information from users. *The main problem of this research aim at the*

process of discovery the useful patterns from the data rather than the data owner. Therefore, our focus is based on how to provide the methods and tools for processing, selecting, ordering, and grouping of rules.

3.2 Research Objectives

1. To study the KDD, DM, interestingness approach, the Meta-Rules Guided Mining and Ontology.
2. To propose a novel model using the Ontology to construct the Meta-Rules Guided Mining to discover the subjective interestingness patterns of data mining.
3. To design a data mining framework and develop a prototype software for discovery the interestingness patterns conforming to our proposed model.
4. To evaluate the capabilities of the proposed model to discover the subjective interestingness patterns of the data mining process.

3.3 Research Contribution

The main contributions of this research are as follows:

1. We proposed an idea of the novel data mining framework that aims to address the minimality of the discarded patterns and the interestingness issues of the data mining.
2. The Meta-Rules Guided Mining methodology will guide the discovery process of data mining and refine the discarded patterns. The result of the process will improve the efficiency and effectiveness of the data mining.
3. Our methodology adapts the ontology to construct the Meta-Rules Guided Mining in terms of the generalized rules, which are used to prune the specialized rules.

3.4 Scope of the Research

Our research covers the following tasks:

1. We emphasize on the subjective interestingness issue of DM, especially in terms of the “Actionability” category. This implies that the proposed approach is used to support the user to purify the discovered patterns of the DM, in the way that user’s

needs as a main criterion. The proposed framework is an “*Extended-Version*” of the traditional DM of which the post analysis of DM is the main focus.

2. The proposed framework is based on the Ontology of the background knowledge of each user. The “*Individual Ontology*” is defined to support the personality of the user.
3. For the implementation part, the prototype FCA2MR is developed and verified to support our proposed idea.



CHAPTER 4

A UNIFIED FRAMEWORK OF META-RULES GUIDED MINING BASED ON ONTOLOGY

4.1 The Fundamental Concept of a Unified Framework of Meta-Rules Guided Mining Based on Ontology

We propose a novel idea to construct the Meta-Rules Guided Mining by applying Ontology, a challenging technology in artificial intelligence (AI) and data mining area. The ontology has been widely studied in order to describe the “Conceptual Model” of things in the interesting domain. The ontology conceptualizes all concepts for describing the relationships of interesting things and emphasizes on categorizing things by given the fine definitive and exhaustive classification of the existence in a particular domain.

In our research, the ontology is reduced to a special sub-form: “*Taxonomy*” of generality relations among the concepts, which are defined in our approach as the level of generalized rules, as depicted in Figure 4-1.

The taxonomy is a hierarchical structure of the concepts that is described in the form of “*Direct Acyclic Graph (DAG)*”.

Our goal is to suggest and generate new ontology concepts based on the set of objects occurring in a database. These objects are defined as the specialized rules level of the taxonomy.

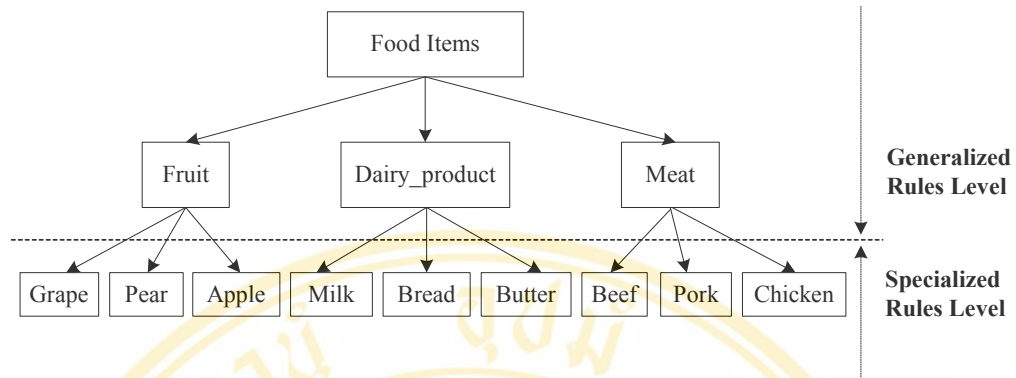


Figure 4-1: Taxonomy of the Meta-Rules Guided Mining

Figure 4-1 presents an example of the taxonomy of the Meta-Rules Guided Mining. The root is labeled the Food Items concept that contains three sub concepts, namely, *Fruit*, *Dairy_product*, and *Meat* concept. These concepts are generalized by the objects that occur in datasets of basket transactions: *Grape*, *Pear*, *Apple*, *Milk*, *Bread*, *Butter*, *Beef*, *Pork*, and *Chicken*, respectively.

Our proposed idea primarily accumulates the concepts of the taxonomy to construct the Meta-Rules Guided Mining. When the completed taxonomy is generated, the process of the meta-rules construction will be performed.

4.2 The Proposed Framework of the Meta-Rules Guided Mining Based on Ontology

The proposed framework consists of two parts. First, we refer to the theory of the “*First Order Logic (FOL)*” to define the “*Specialized Rules*” via the specialized rules discovery part. This part is basically designed for the DM process in order to generate the discovered patterns and then store them in the discovered-patterns repository. In contrast of the traditional DM, this part is generated once for discovering all the patterns that might be seen in the database.

Second, we refer to the theory of the “*Second Order Logic (SOL)*” [31, 37], in terms of the meta-rules, which is used to propose the idea to discover the “*Generalized Rules*” via the generalized rules discovery part. This part is called the “*Extended Version*” of the DM to discover the interestingness patterns.

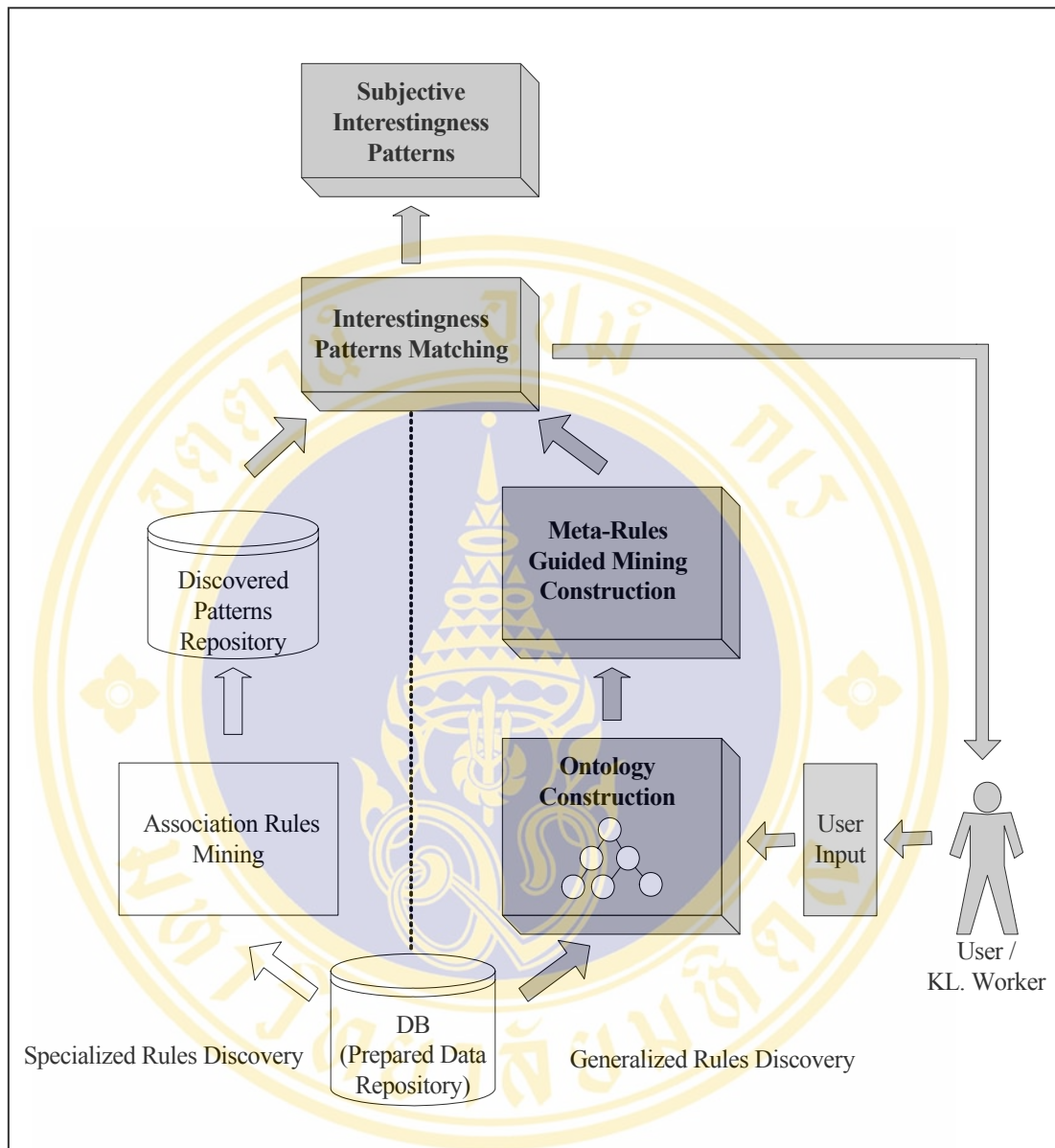


Figure 4-2: An Overview of the Meta-Rules Guided Mining Based on Ontology Framework

Figure 4-2 illustrates an overall Framework of the Meta-Rules Guided Mining methodology based on ontology. On the left-hand-side, it is the specialized rules discovery part consisting of three modules. The first module, *Prepared-Data Repository module*, is used to store the data prompted for using in data mining process. The second module, *Association Rules Mining*, performs the crucial task of data mining by using an Apriori algorithm to discover the patterns of data from database (referred to

session 4.3 for more details). After that, the discovered patterns are stored in the *Discovered Pattern Repository module*.

On the right-hand-side, the generalized rules discovery part is outlined in the highlighted block diagram which consists of five modules, namely, *User Input Module*, *Ontology Construction module*, *Meta-Rules Guided Mining Construction module*, *Interestingness Patterns Matching module*, and *Subjective Interestingness Patterns module*.

1. *Ontology Construction Module* plays the important role for our proposed model. This module uses the concept of ontology to construct the “*Taxonomy*” of the concepts. *In our research, the ontology is reduced to a special subform: the taxonomy of generality relations among the concepts.* The generality relationships are represented as the taxonomy in the form of a graph. The user-expected patterns are gathered from the user or knowledge worker summarized in terms of ontology concepts.
2. The *Meta-Rules Guided Mining Construction Module* is used to generate the meta-rules from the prepared ontology concepts. This module traverses through the taxonomy to generate the meta-rules the same fashion of association rules mining.
3. The *Interestingness Patterns Matching Module* is responsible for comparing the discovered-patterns with the meta-rules (the results from the previous module) to discover the interestingness patterns. The results of this module are thus defined as the subjective interestingness patterns and stored in the last module--the subjective interestingness patterns module.

In the last step of the model -- interestingness patterns matching module, if the results are not sufficiently interesting to the user, it needs iterative process to re-construct the ontology concepts. The overall process is interactive and iterative until the results are determined as the interesting patterns.

Technically, our proposed model is based on the iterative and interactive process in the generalized rules discovery part. It implies that the proposed model is a contrast idea of the traditional DM framework, which aims to interact with the users

while the process is in the DM task (as described earlier in the specialized rules discovery part).

For a clearer view point of the Meta-Rules Guided Mining, Figure 4-3 shows the examples of two types of rules generated from the taxonomy in Figure 4-1.

Itemsets	I = {Grape, Pear, Apple, Milk, Bread, Butter, Beef, Pork, Chicken}
Specialized Rules	Discovered-Rule#1 = Bread $\Rightarrow$ Butter #2 = Apple $\Rightarrow$ Milk
Generalized Rules	Meta-Rule#1 = Dairy_Product $\Rightarrow$ Dairy_Product #2 = Fruit $\Rightarrow$ Dairy_Product

Figure 4-3: An example of the Itemsets, Specialized and Generalized Rules

Suppose the relation schema, *I*, holds for the itemset of *Grape, Pear, Apple, Milk, Bread, Butter, Beef, Pork, and Chicken*. The Association Rules Mining module discover the patterns of specialized rules, such as

- *Discovered-Rule#1: Bread $\Rightarrow$ Butter*
- *Discovered-Rule#2: Apple $\Rightarrow$ Milk*

After that, we can generate the Meta-Rules, such as

- *Meta-Rule#1: Dairy_Product $\Rightarrow$ Dairy_Product*
- *Meta-Rule#2: Fruit $\Rightarrow$ Dairy_Product.*

In each concept of the taxonomy, we define the class hierarchy as follows:

{Grape, Pear, Apple}	$\subset$ Fruit	$\subset$ Food Items
{Milk, Bread, Butter}	$\subset$ Dairy_Product	$\subset$ Food Items
{Beef, Pork, Chicken}	$\subset$ Meat	$\subset$ Food Items

Our methodology will guide the process of interestingness patterns discovery to bind the target patterns that occur in each path of taxonomy. It means that if we can organize the objects in the hierarchical structure or taxonomy, then we will be able to conceptualize the relationships of the objects in which each object is instantiated by

each other. Based on the user's criterions, if we can traverse through the taxonomy in the same direction of user's needs, the discovery process will generate the interestingness patterns as the result.

4.3 THE FIRST COMPONENT: Association Rules Mining Module

The first component describes an implementation of the specialized rules discovery process via the Association Rules Mining module by using the Apriori algorithm. The Apriori algorithm is the best known algorithm for association rule induction [7, 38]. This algorithm works in two steps:

- First, the “*frequent itemsets*” are determined. There are sets of items that have at least the given minimum support (Min_Supp).
- Second, the association rules are generated from the frequent itemsets found in the first step by determining with the given minimum confidence (Min_Conf). Usually the first step is the more important, because it accounts for the greater part of the processing time.

To implement the Association Rule Mining module, the share library of the Apriori implementation by Borgelt et al. [3] is adopted. This share library has been referenced and recognized among the theoretical works and commercial tools of data mining, such as the Cmentine , a one core product of SPSS.

4.4 THE SECOND COMPONENT: Ontology Construction Module

The lattice for taxonomy is derived from the attributes of the database by “*Formal Concept Analysis (FCA)*” technique. New generalized concepts can be derived from the prepared-data repository module, and the data for constructing FCA lattice are the same as those used in the association rules mining module. Furthermore, FCA technique is also applied to find the sub-lattice or sub-class that is automatically merged with super-class of the taxonomy.

4.4.1 The Ontology Construction

We design the ontology construction based on the theory of FCA [8, 28]. The details of the process are described in section 4.4.1.1 and the schematic view of the process is presented in section 4.4.1.2.

4.4.1.1 The Process of Ontology Construction

The process of ontology construction consists of five steps:

1. Object Initialization process
2. Attribute Initialization process
3. Formal Context Generation process
4. Formal Concept Generation process
5. Formal Lattice Generation process

The details of each step are described as follow:

STEP 1: Object Initialization Process

The object initialization process is designed to initiate the object extracted from the prepared database via the itemsets extractor. The itemsets extractor groups the occurred itemsets and represents in term of objects set.

STEP 2: Attribute Initialization Process

The attribute initialization process is designed to get the user input data via the user input module. The acquired data are initiated as the primary attributes set and determined with the existing attribute in the ontology repository to obtain the domain of attribute set.

STEP 3: Formal Context Generation Process

The formal context generation process is designed to extract the common feature of the formal context [36]. We referred to the theory of the formal context as follows:

Theorem 1: Formal Context (K) “Formal Context K is a set of triplet $K(O, A, I)$; where O is a set of Objects; A is a set of Attributes; and I is a set of binary relation between O and A , $I \subseteq O \times A$ ”

From Theorem1, we define K as follows:

- $K(O, A, I)$ is a Formal Context.
- $O = \{o_1, o_2, \dots, o_m \mid o = \text{Objects in Relation}\}$
- $A = \{a_1, a_2, \dots, a_n \mid a = \text{Attributes in Relation}\}$
- $I = \{i_{11}, i_{12}, \dots, i_{m \times n} \mid i = \text{Binary Relation}, I \subseteq O \times A, \\ m = |O|, \\ n = |A|\}$

The cross-table is used to represent the perspective of the binary relations among the objects and the attributes of the formal context K , as depicted in Figure 4-4. The attributes in the figure are Fruit, Dairy_Product, and Meat, -- and the objects are Grape, Pear, Apple, Milk, Bread, Butter, Beef, Pork and Chicken.

A : attributes →

	Fruit	Dairy_product	Meat
Grape	☆		
Pear	☆		
Apple	☆		
Milk		☆	
Bread		☆	
Butter		☆	
Beef			☆
Pork			☆
Chicken			☆

↓ **O : objects**

Figure 4-4: The Cross-Table of the Formal Context K .

From the cross-table of the formal context K , we illustrate an example of the formal context that can be extracted from the database relation, as depicted in Figure 4-5.

$$K(O, A, I) = (\{ \text{Grape}_1, \text{Pear}_2, \text{Apple}_3, \text{Milk}_4, \text{Bread}_5, \text{Butter}_6, \text{Beef}_7, \text{Pork}_8, \text{Chicken}_9 \}, \\ \{ \text{Fruit}_1, \text{Dairy_Product}_2, \text{Meat}_3 \}, \\ \{ i_{11}, i_{21}, i_{31}, i_{42}, i_{52}, i_{62}, i_{73}, i_{83}, i_{93} \})$$

Figure 4-5: An Example of the Formal Context K .

STEP 4: Formal Concept Generation Process

The formal concept generation process is designed to form the set of concepts of the taxonomy. This process plays an important role for the ontology construction module. Additionally, this process is a main focus of the FCA technique to generate the concept of the ontology.

The core theory of the FCA provides a way to identify a group of “*Objects*” that has “*Common Attributes*”, and a group of “*Attributes*” that belongs to the “*Common Objects*”.

The following definitions, namely, *Objects Concept*, *Attributes Concept*, *Mapping Function*, *Common Objects Concept*, and *Common Attributes Concept* are defined as follows:

Definition 1: Attribute Concept (A) “Attribute Concept A is a set of class member (or properties, or characteristics) of each Object in the Database Relation.”

From Definition 1, attribute concept (A) is defined as follows:

- $A = \{a \in A \mid \forall o \in O, (o, a) \in I\}$
- $A = \{a_1, a_2, \dots, a_n \mid a = \text{Attributes in Relation}, \\ n = |A|, \text{ which } a_n \text{ occur in } K\}$

Definition 2: Object Concept (O) “Object Concept O is a set of Items (or Variables, or Entities) in the Database Relation.”

From Definition 2, object concept (O) is defined as follows:

- $O = \{o \in O \mid \forall o \in O, (o, a) \in I\}$
- $O = \{o_1, o_2, \dots, o_m \mid o = \text{Objects in Relation}, \\ m = |O|, \text{ which } o_m \text{ occur in } K\}$

Definition 3: Mapping Function $f()$, $g()$ “The mapping function $f()$ maps a set of object into a set of common object, whereas $g()$ maps a set of attribute into a set of common attributes.”

From Definition 3, mapping function $f()$ and $g()$ are defined as follows:

- $f: f(O) \rightarrow O' = \{ o \in O \mid \forall a \in A, (o, a) \in I \}$
- $g: g(A) \rightarrow A' = \{ a \in A \mid \forall o \in O, (o, a) \in I \}$

Definition 4: Common Attributes Concept (A') “Common Attribute Concept A' is a set of Attributes Concept, which common to Objects Concept”

From Definition 4, common attributes concept (A') is defined as follows:

- $A' = \{ a \in A \mid \forall o \in O, (o, a) \in I \}$
- $A' = \{ a_1, a_2, \dots, a_j \mid a = \text{Attributes in Relation, } j = |A|, \text{ which } a_j \text{ occur in } K \}$

Definition 5: Common Object Concept (O') “Common Object Concept O' is a set of Objects Concept, which is common to Attributes Concept”

From Definition 5, common object concept (O') is defined as follows:

- $O' = \{ o \in O \mid \forall a \in A, (o, a) \in I \}$
- $O' = \{ o_1, o_2, \dots, o_k \mid o = \text{Objects in Relation, } k = |O|, \text{ which } o_k \text{ occur in } K \}$

Theoretically, the definitions above simplify that a formal concept C of a formal context K is defined as a pair of the objects concept and attributes concept. Hence, from the formal context K , we referred to the theory of formal concept as follows:

Theorem 2: Formal Concept (C) “ Formal Concept C is a set of an order pair (O, A) if $A = A'$ and $O = O'$ ”

From the Theorem2, formal concept (C) is defined as follows:

- $C(O, A)$ is a Formal Concept
- $O = \{ o_1, o_2, \dots, o_m \mid o = \text{Objects in Relation} \}$
- $A = \{ a_1, a_2, \dots, a_n \mid a = \text{Attributes in Relation} \}$
- $A' = \{ a \in A \mid \forall o \in O, (o, a) \in I \}$

- $O' = \{o \in O \mid \forall a \in A, (o, a) \in I\}$
- $C = \{c \in C \mid \forall o \in O \text{ and } \forall a \in A, (o, a) \in I\}$
- $c \in C \Leftrightarrow A=A' \text{ and } O = O'$
- $C = (\{o_1, o_2, \dots, o_m\}, \{a_1, a_2, \dots, a_n\}) \mid$ o = Objects in Relation,
 a = Attributes in Relation,
 $m = |O|$, which o_m occur in K
 $n = |A|$, which a_n occur in K }

Now, we introduce an algorithm to construct the formal concept, namely, the GEN_FC algorithm. The GEN_FC is used to generate the formal concept (C), which is performed by the formal concept generation module in the generalization part of the proposed framework. In order to describe the process of the algorithm, we give an example 4-1 as follows:

Example 4-1 Let $O = \{\text{Grape, Pear, Milk}\}$, and $A = \{\text{Fruit, Dairy_Product}\}$ be the subset of object, and attribute of the formal context K that depicted in Figure 4-5. The binary relationship of the formal context: $I = \{i_{11}, i_{21}, i_{32}\}$

From the example 4-1, the details of the algorithm are described as follows:

STEP 4.1: Generate the PowerSET of Objects and Attributes

The first step of the GEN_FC algorithm is to generate the PowerSET of objects and attributes of the input data. Based on the set theory of the mathematics, if A is a set, then the set of all subsets of A is called the “PowerSET” of A . We now give an example of generating the PowerSET of objects O and Attribute A as follows:

Table 4-1: An Example of the PowerSET of Objects and Attributes

Objects: PowerSET(O)
= {{Grape} ₁ , {Pear} ₂ , {Milk} ₃ , {Grape, Pear} ₄ , {Grape, Milk} ₅ , {Pear, Milk} ₆ , {Grape, Pear, Milk} ₇ , {} ₈ }
Attributes: PowerSET(A)
= {{Fruit} ₁ , {Dairy_Product} ₂ , {Fruit, Dairy_Product} ₃ , {} ₄ }

STEP 4.2: Class Extraction Process

According to the Definitions 4 and 5, to extract the common attributes (A') and common objects (O'), we design the process of class extraction process to perform this task. The process will go through the PowerSET of object and attribute to perceive the binary relationship (I) on the formal context K of each object and attribute.

The class extraction process determines the characteristic of each class in two actions; perceive the binary relationship (I), and take the sub-set property.

- **Action 1: Percept the Binary Relationship I** , from the formal context K . This action perceives the binary relationship from the cross-table to determine the relationship between object and attribute.

IF (Object, Attribute) is in Binary Relation I , THEN return percept = TRUE

- **Action 2: Take Sub-Set Property.** This action determines the empty set “{}” is a sub-set of every set.

IF Object or Attribute is an empty set {}, THEN take sub-set = TRUE

Table 4-2, and 4-3 demonstrate an example of the class extraction mechanism. The details are listed below:

Table 4-2: An Example of the Common Attributes (A') Extractor

Objects (O) $\rightarrow$ Common Attributes: (A')					
		Extracted Pair	Percept I	Take Subset	Result
i=1	j=1	{Grape} ₁ , {Fruit} ₁	T	F	$O_1 = \{\text{Grape}\}, A'_1 = \{\text{Fruit}\}$ (*)
	j=2	{Grape} ₁ , {Dairy_Product} ₂	F	F	
	j=3	{Grape} ₁ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Grape} ₁ , {} ₄	F	T	
i=2	j=1	{Pear} ₂ , {Fruit} ₁	T	F	$O_2 = \{\text{Pear}\}, A'_2 = \{\text{Fruit}\}$ (*)
	j=2	{Pear} ₂ , {Dairy_Product} ₂	F	F	
	j=3	{Pear} ₂ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Pear} ₂ , {} ₄	F	T	
i=3	j=1	{Milk} ₃ , {Fruit} ₁	F	F	$O_3 = \{\text{Milk}\}, A'_3 = \{\text{Dairy_Product}\}$ (*)
	j=2	{Milk} ₃ , {Dairy_Product} ₂	T	F	
	j=3	{Milk} ₃ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Milk} ₃ , {} ₄	F	T	
i=4	j=1	{Grape, Pear} ₄ , {Fruit} ₁	T	F	$O_4 = \{\text{Grape, Pear}\}, A'_4 = \{\text{Fruit}\}$ (*)
	j=2	{Grape, Pear} ₄ , {Dairy_Product} ₂	F	F	
	j=3	{Grape, Pear} ₄ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Grape, Pear} ₄ , {} ₄	F	T	
i=5	j=1	{Grape, Milk} ₅ , {Fruit} ₁	F	F	$O_5 = \{\text{Grape, Milk}\}, A'_5 = \{\}$
	j=2	{Grape, Milk} ₅ , {Dairy_Product} ₂	F	F	
	j=3	{Grape, Milk} ₅ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Grape, Milk} ₅ , {} ₄	F	T	
i=6	j=1	{Pear, Milk} ₆ , {Fruit} ₁	F	F	$O_6 = \{\text{Pear, Milk}\}, A'_6 = \{\}$
	j=2	{Pear, Milk} ₆ , {Dairy_Product} ₂	F	F	
	j=3	{Pear, Milk} ₆ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Pear, Milk} ₆ , {} ₄	F	T	
i=7	j=1	{Grape, Pear, Milk} ₇ , {Fruit} ₁	F	F	$O_7 = \{\text{Grape, Pear, Milk}\}, A'_7 = \{\}$
	j=2	{Grape, Pear, Milk} ₇ , {Dairy_Product} ₂	F	F	
	j=3	{Grape, Pear, Milk} ₇ , {Fruit, Dairy_Product} ₃	F	F	
	j=4	{Grape, Pear, Milk} ₇ , {} ₄	F	T	
i=8	j=1	{}, {Fruit} ₁	F	T	$O_8 = \{\}, A'_8 = \{\text{Fruit, Dairy_Product}\}$ (*)
	j=2	{}, {Dairy_Product} ₂	F	T	
	j=3	{}, {Fruit, Dairy_Product} ₃	F	T	
	j=4	{}, {} ₄	F	T	

Table 4-3: An Example of the Common Objects (O') Extractor

Attribute (A) $\rightarrow$ Common Objects: (O')					
		Extracted Pair	Percept I	Take Subset	Result
i=1	j=1	{Fruit} ₁ , {Grape} ₁	T	F	$A_1 = \{\text{Fruit}\}, O'_1 = \{\text{Grape, Pear}\} \quad (*)$
	j=2	{Fruit} ₁ , {Pear} ₂	T	F	
	j=3	{Fruit} ₁ , {Milk} ₃	F	F	
	j=4	{Fruit} ₁ , {Grape, Pear} ₄	T	F	
	j=5	{Fruit} ₁ , {Grape, Milk} ₅	F	F	
	j=6	{Fruit} ₁ , {Pear, Milk} ₆	F	F	
	j=7	{Fruit} ₁ , {Grape, Pear, Milk} ₇	F	F	
	j=8	{Fruit} ₁ , {} ₈	F	T	
i=2	j=1	{Dairy_Product} ₂ , {Grape} ₁	F	F	$A_2 = \{\text{Dairy_Product}\}, O'_2 = \{\text{Milk}\} \quad (*)$
	j=2	{Dairy_Product} ₂ , {Pear} ₂	F	F	
	j=3	{Dairy_Product} ₂ , {Milk} ₃	T	F	
	j=4	{Dairy_Product} ₂ , {Grape, Pear} ₄	F	F	
	j=5	{Dairy_Product} ₂ , {Grape, Milk} ₅	F	F	
	j=6	{Dairy_Product} ₂ , {Pear, Milk} ₆	F	F	
	j=7	{Dairy_Product} ₂ , {Grape, Pear, Milk} ₇	F	F	
	j=8	{Dairy_Product} ₂ , {} ₈	F	T	
i=3	j=1	{Fruit, Dairy_Product} ₃ , {Grape} ₁	F	F	$A_3 = \{\text{Fruit, Dairy_Product}\}, O'_3 = \{\}$
	j=2	{Fruit, Dairy_Product} ₃ , {Pear} ₂	F	F	
	j=3	{Fruit, Dairy_Product} ₃ , {Milk} ₃	F	F	
	j=4	{Fruit, Dairy_Product} ₃ , {Grape, Pear} ₄	F	F	
	j=5	{Fruit, Dairy_Product} ₃ , {Grape, Milk} ₅	F	F	
	j=6	{Fruit, Dairy_Product} ₃ , {Pear, Milk} ₆	F	F	
	j=7	{Fruit, Dairy_Product} ₃ , {Grape, Pear, Milk} ₇	F	F	
	j=8	{Fruit, Dairy_Product} ₃ , {} ₈	F	T	
i=4	j=1	{ ₄ , {Grape} ₁	F	T	$A_4 = \{\}, O'_4 = \{\text{Grape, Pear, Milk}\} \quad (*)$
	j=2	{ ₄ , {Pear} ₂	F	T	
	j=3	{ ₄ , {Milk} ₃	F	T	
	j=4	{ ₄ , {Grape, Pear} ₄	F	T	
	j=5	{ ₄ , {Grape, Milk} ₅	F	T	
	j=6	{ ₄ , {Pear, Milk} ₆	F	T	
	j=7	{ ₄ , {Grape, Pear, Milk} ₇	F	T	
	j=8	{ ₄ , {} ₈	F	T	

(*) **NOTE:** In case of many perceptions, we use the longest member of object set. For example, in the last case: “i=4”, there are eight sub-sets. Hence, the longest member of object is {Grape, Pear, Milk}.

STEP 4.3: Class Matching Process

The class matching process matches all combinations of the objects, attributes, common objects, and common attributes. An example of the class matching process is illustrated in Table 4-4.

Table 4-4: An Example of the Class Matching Process

		Common Attributes	Common Object	Match	Result
i=1	j=1	$A_1 = \{\text{Fruit}\}$ $A'_1 = \{\text{Fruit}\}$	$O_1 = \{\text{Grape}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 = A'_1$ $O_1 \neq O'_1$	$(\{\text{Grape}\}, \{\text{Fruit}\}) \notin C$
	j=2	$A_1 = \{\text{Fruit}\}$ $A'_2 = \{\text{Fruit}\}$	$O_2 = \{\text{Pear}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 = A'_2$ $O_2 \neq O'_1$	$(\{\text{Pear}\}, \{\text{Fruit}\}) \notin C$
	j=3	$A_1 = \{\text{Fruit}\}$ $A'_3 = \{\text{Dairy_Product}\}$	$O_3 = \{\text{Milk}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 \neq A'_3$ $O_3 \neq O'_1$	$(\{\text{Milk}\}, \{\text{Fruit}\}) \notin C$
	j=4	$A_1 = \{\text{Fruit}\}$ $A'_4 = \{\text{Fruit}\}$	$O_4 = \{\text{Grape, Pear}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 = A'_4$ $O_4 = O'_1$	$(\{\text{Grape, Pear}\}, \{\text{Fruit}\}) \in C$
	j=5	$A_1 = \{\text{Fruit}\}$ $A'_5 = \{\}$	$O_5 = \{\text{Grape, Milk}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 \neq A'_5$ $O_5 \neq O'_1$	$(\{\text{Grape, Milk}\}, \{\text{Fruit}\}) \notin C$
	j=6	$A_1 = \{\text{Fruit}\}$ $A'_6 = \{\}$	$O_6 = \{\text{Pear, Milk}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 \neq A'_6$ $O_6 \neq O'_1$	$(\{\text{Pear, Milk}\}, \{\text{Fruit}\}) \notin C$
	j=7	$A_1 = \{\text{Fruit}\}$ $A'_7 = \{\}$	$O_7 = \{\text{Grape, Pear, Milk}\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 \neq A'_7$ $O_7 \neq O'_1$	$(\{\text{Grape, Pear, Milk}\}, \{\text{Fruit}\}) \notin C$
	j=8	$A_1 = \{\text{Fruit}\}$ $A'_8 = \{\text{Fruit, Dairy_Product}\}$	$O_8 = \{\}$ $O'_1 = \{\text{Grape, Pear}\}$	$A_1 \neq O'_8$ $O_8 \neq O'_1$	$(\{\}, \{\text{Fruit}\}) \notin C$
i=2	j=1	$A_2 = \{\text{Dairy_Product}\}$ $A'_1 = \{\text{Fruit}\}$	$O_1 = \{\text{Grape}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_1$ $O_1 \neq O'_2$	$(\{\text{Grape}\}, \{\text{Dairy_Product}\}) \notin C$
	j=2	$A_2 = \{\text{Dairy_Product}\}$ $A'_2 = \{\text{Fruit}\}$	$O_2 = \{\text{Pear}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_2$ $O_2 \neq O'_2$	$(\{\text{Pear}\}, \{\text{Dairy_Product}\}) \notin C$
	j=3	$A_2 = \{\text{Dairy_Product}\}$ $A'_3 = \{\text{Dairy_Product}\}$	$O_3 = \{\text{Milk}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 = A'_3$ $O_3 = O'_2$	$(\{\text{Milk}\}, \{\text{Dairy_Product}\}) \in C$
	j=4	$A_2 = \{\text{Dairy_Product}\}$ $A'_4 = \{\text{Fruit}\}$	$O_4 = \{\text{Grape, Pear}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_4$ $O_4 \neq O'_2$	$(\{\text{Grape, Pear}\}, \{\text{Dairy_Product}\}) \notin C$
	j=5	$A_2 = \{\text{Dairy_Product}\}$ $A'_5 = \{\}$	$O_5 = \{\text{Grape, Milk}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_5$ $O_5 \neq O'_2$	$(\{\text{Grape, Milk}\}, \{\text{Dairy_Product}\}) \notin C$
	j=6	$A_2 = \{\text{Dairy_Product}\}$ $A'_6 = \{\}$	$O_6 = \{\text{Pear, Milk}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_6$ $O_6 \neq O'_2$	$(\{\text{Pear, Milk}\}, \{\text{Dairy_Product}\}) \notin C$
	j=7	$A_2 = \{\text{Dairy_Product}\}$ $A'_7 = \{\}$	$O_7 = \{\text{Grape, Pear, Milk}\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_7$ $O_7 \neq O'_2$	$(\{\text{Grape, Pear, Milk}\}, \{\text{Dairy_Product}\}) \notin C$
	j=8	$A_2 = \{\text{Dairy_Product}\}$ $A'_8 = \{\text{Fruit, Dairy_Product}\}$	$O_8 = \{\}$ $O'_2 = \{\text{Milk}\}$	$A_2 \neq A'_8$ $O_8 \neq O'_2$	$(\{\}, \{\text{Dairy_Product}\}) \notin C$

Table 4-4: An Example of the Class Matching Process (Cont.)

		Common Attributes	Common Object	Match	Result
i=3	j=1	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_1 = \{\text{Fruit}\}$	$O_1 = \{\text{Grape}\}$ $O'_3 = \{\}$	$A_3 \neq A'_1$ $O_1 \neq O'_3$	$(\{\text{Grape}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=2	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_2 = \{\text{Fruit}\}$	$O_2 = \{\text{Pear}\}$ $O'_3 = \{\}$	$A_3 \neq A'_2$ $O_2 \neq O'_3$	$(\{\text{Pear}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=3	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_3 = \{\text{Dairy_Product}\}$	$O_3 = \{\text{Milk}\}$ $O'_3 = \{\}$	$A_3 \neq A'_3$ $O_3 \neq O'_3$	$(\{\text{Milk}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=4	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_4 = \{\text{Fruit}\}$	$O_4 = \{\text{Grape, Pear}\}$ $O'_3 = \{\}$	$A_3 \neq A'_4$ $O_4 \neq O'_3$	$(\{\text{Grape, Pear}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=5	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_5 = \{\}$	$O_5 = \{\text{Grape, Milk}\}$ $O'_3 = \{\}$	$A_3 \neq A'_5$ $O_5 \neq O'_3$	$(\{\text{Grape, Milk}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=6	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_6 = \{\}$	$O_6 = \{\text{Pear, Milk}\}$ $O'_3 = \{\}$	$A_3 \neq A'_6$ $O_6 \neq O'_3$	$(\{\text{Pear, Milk}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=7	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_7 = \{\}$	$O_7 = \{\text{Grape, Pear, Milk}\}$ $O'_3 = \{\}$	$A_3 \neq A'_7$ $O_7 \neq O'_3$	$(\{\text{Grape, Pear, Milk}\}, \{\text{Fruit, Dairy_Product}\}) \notin C$
	j=8	$A_3 = \{\text{Fruit, Dairy_Product}\}$ $A'_8 = \{\text{Fruit, Dairy_Product}\}$	$O_8 = \{\}$ $O'_3 = \{\}$	$A_3 = A'_8$ $O_8 = O'_3$	$(\{\}, \{\text{Fruit, Dairy_Product}\}) \in C$
i=4	j=1	$A_4 = \{\}$ $A'_1 = \{\text{Fruit}\}$	$O_1 = \{\text{Grape}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_1$ $O_1 \neq O'_4$	$(\{\text{Grape}\}, \{\}) \notin C$
	j=2	$A_4 = \{\}$ $A'_2 = \{\text{Fruit}\}$	$O_2 = \{\text{Pear}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_2$ $O_2 \neq O'_4$	$(\{\text{Pear}\}, \{\}) \notin C$
	j=3	$A_4 = \{\}$ $A'_3 = \{\text{Dairy_Product}\}$	$O_3 = \{\text{Milk}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_3$ $O_3 \neq O'_4$	$(\{\text{Milk}\}, \{\}) \notin C$
	j=4	$A_4 = \{\}$ $A'_4 = \{\text{Fruit}\}$	$O_4 = \{\text{Grape, Pear}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_4$ $O_4 \neq O'_4$	$(\{\text{Grape, Pear}\}, \{\}) \notin C$
	j=5	$A_4 = \{\}$ $A'_5 = \{\}$	$O_5 = \{\text{Grape, Milk}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_5$ $O_5 \neq O'_4$	$(\{\text{Grape, Milk}\}, \{\}) \notin C$
	j=6	$A_4 = \{\}$ $A'_6 = \{\}$	$O_6 = \{\text{Pear, Milk}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_6$ $O_6 \neq O'_4$	$(\{\text{Pear, Milk}\}, \{\}) \notin C$
	j=7	$A_4 = \{\}$ $A'_7 = \{\}$	$O_7 = \{\text{Grape, Pear, Milk}\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 = A'_7$ $O_7 = O'_4$	$(\{\text{Grape, Pear, Milk}\}, \{\}) \in C$
	j=8	$A_4 = \{\}$ $A'_8 = \{\text{Fruit, Dairy_Product}\}$	$O_8 = \{\}$ $O'_4 = \{\text{Grape, Pear, Milk}\}$	$A_4 \neq A'_8$ $O_8 \neq O'_4$	$(\{\}, \{\}) \notin C$

From the example 4-1, the formal concept C is constructed as depicted in Figure 4-6.

$$C(O, A) = \{ C_1(\{\}, \{\text{Fruit, Dairy_Product}\}), \quad C_2(\{\text{Grape, Pear}\}, \{\text{Fruit}\}), \\ C_3(\{\text{Milk}\}, \{\text{Dairy_Product}\}), \quad C_4(\{\text{Grape, Pear, Milk}\}, \{\}) \}$$

Figure 4-6: An Example of the Formal Concept C .

STEP 5: Formal Lattice Generation Process

The order of the taxonomy is defined over the concepts of the formal concept C . The class hierarchy of the taxonomy is the super-class / sub-class relationship between such a concept C_i . From this point of view, the perspective view of the relationship between the concepts is the significant role to form the taxonomy.

To build a rich perspective of how the taxonomy is actually done in practice, we then refer to the theory of the formal lattice as follows:

Theorem 3: Formal Lattice (L) “A set of all Formal Concept C of the Formal Context K forms a ‘Partial Order’ via: $(O_1, A_1) \leq (O_2, A_2) \Leftrightarrow O_1 \subseteq O_2$ and $A_2 \subseteq A_1$ ”

From the Theorem 3, we concentrate on the study of the class hierarchy of the taxonomy by which the class hierarchical relationships are constructed from the “Partial Order” relations.

We refer to the theory of the “Partially Ordered Sets” or “POSet” [29] to state the definition of the “Class-Hierarchical Operator”. Formally, based on the “Discrete Mathematics” [29], the mathematical framework to describe the kind of organization of sets is the theory of the “Relations”. To make clear about the theory of the Relations, we describe the related terms as follow:

- **Order:** An Order is a set whose members can be compared with each other.
- **Partial Order:** Let S be a set. The relation is called a “Partial Order” on set S if it satisfies the following properties:

- o *Reflexive*, $s \preceq s$ for every s in S ;
- o *Antisymmetric*, $s \preceq t$ and $t \preceq s$ imply $s = t$;
- o *Transitive*, $s \preceq t$ and $t \preceq u$ imply $s \preceq u$.
- **POSet**: if $\preceq$ is a “Partial Order” on S the pair $(S, \preceq)$ is called a “Partially Ordered Set” or “POSet”.

Based on the POSet theory, we applied the core concept of POSet to formalize the class hierarchical relationship between set of the Formal Concept via the “Class-Hierarchical Operator”.

Hence, the definition of the “Class-Hierarchical Operator” is defined to illustrate the class hierarchical relationship as follow:

Definition 6: Class-Hierarchical Operator “A Formal Concept $c_1(o_1, a_1)$ is a *Sub-Concept* of another Formal Concept $c_2(o_2, a_2)$ if and only if $o_1 \subseteq o_2$, and $a_2 \subseteq a_1$, we denoted ‘ $\leq$ ’ as a *Sub-Class Operator*. Vice versa, a Formal Concept $c_2(o_2, a_2)$ is a *Super-Concept* of Formal Concept $c_1(o_1, a_1)$ if and only if $o_2 \supseteq o_1$, and $a_1 \supseteq a_2$, we denoted ‘ $\geq$ ’ as a *Super-Class Operator*.”

As mentioned earlier, we simplify the core idea of the formal lattice construction process by proposing the “Class-Hierarchical Operator”, which will be used to construct the taxonomy.

The class hierarchical of the taxonomy is constructed by forming the “Associated Concept”. The definition of the associated concepts is defined as follows:

Definition 7: Associated Concept “A Formal Concept $c_1(o_1, a_1)$ is an *Associated Concept* of $c_2(o_2, a_2)$, if it satisfies the Class-Hierarchical Operator. The form $(c_1(o_1, a_1), c_2(o_2, a_2), \leq)$ is the Class-Hierarchy of the Formal Lattice, we called $c_1(o_1, a_1)$ is a *Sub-Concept* of $c_2(o_2, a_2)$. Vice versa, the form $(c_2(o_2, a_2), c_1(o_1, a_1), \geq)$ is the Class-Hierarchy of the Formal Lattice, we called $c_2(o_2, a_2)$ is a *Super-Concept* of $c_1(o_1, a_1)$.”

From the Definition 7, the *Associated Concept* is defined as follow:

- $(c_1(o_1, a_1), c_2(o_2, a_2), \leq) \Leftrightarrow o_1 \subseteq o_2, \text{ and } a_2 \subseteq a_1$
- $(c_1(o_1, a_1), c_2(o_2, a_2), \geq) \Leftrightarrow o_1 \supseteq o_2, \text{ and } a_2 \supseteq a_1$

To give greater insight into the method of the formal lattice generation process, we use an example of the formal concept that depicted in Figure 4-6, as follows:

Example 4-2 Let $C_1 = (\{\}, \{\text{Fruit, Dairy_Product}\})$
 $C_2 = (\{\text{Grape, Pear}\}, \{\text{Fruit}\})$
 $C_3 = (\{\text{Milk}\}, \{\text{Dairy_Product}\})$
 $C_4 = (\{\text{Grape, Pear, Milk}\}, \{\})$

In example 4-2, the GEN_FL algorithm is used to support the process of formal lattice generation. The detail of the algorithm is illustrated as follows:

STEP 5.1: Generate PowerSET of Formal Concept C

The first step of the GEN_FL algorithm is to generate the PowerSET of the formal concept C . Based on the set theory of the mathematics, we now give an example of the execution of the PowerSET of formal concept as follows:

Table 4-5: An Example of the PowerSET of Formal Concept C

PowerSET(C)
$= \{\{C_1\}, \{C_2\}, \{C_3\}, \{C_4\}, \{C_1 C_2\}, \{C_1 C_3\}, \{C_1 C_4\}, \{C_2 C_3\}, \{C_2 C_4\}, \{C_3 C_4\}, \{C_1 C_2 C_3\},$ $\{C_1, C_2, C_4\}, \{C_1 C_3 C_4\}, \{C_2 C_3 C_4\}, \{C_1 C_2 C_3 C_4\}, \{\}\}$

STEP 5.2: Order Pair Generation Process

The order pair generation process is defined to generate a couple of the element of the PowerSET. The process chooses the PowerSET element that has only two members to generate the order pair. The definition of the order pair is defined as follows:

Definition 8: Order Pair “An order pair is a set of a pair of objects with an order associated with them. If objects are represented by X and Y , then we write an order pair as $X \rightarrow Y$.”

An example of the order pair is shown in Table 4-6.

Table 4-6: An Example of the Order Pair Generation Process

	Two member element of PowerSET	Order Pair
1	$\{C_1 C_2\}$	$\{C_1\} \rightarrow \{C_2\}$
2	$\{C_1 C_3\}$	$\{C_1\} \rightarrow \{C_3\}$
3	$\{C_1 C_4\}$	$\{C_1\} \rightarrow \{C_4\}$
4	$\{C_2 C_3\}$	$\{C_2\} \rightarrow \{C_3\}$
5	$\{C_2 C_4\}$	$\{C_2\} \rightarrow \{C_4\}$
6	$\{C_3 C_4\}$	$\{C_3\} \rightarrow \{C_4\}$

STEP 5.3: Associated Concept Checking

The class-hierarchical operator is used to define the relationship between two formal concepts. Referred to the Definitions 6 and 7, the process of associated concept checking determines the associated concept property by using the class-hierarchical operator. The process will be separated in two ways: super-concept, and sub-concept. Table 4-7 shows an example of the associated concept checking process.

Table 4-7: An Example of the Associated Concept Checking

Order Pair	Obj. / Attr.	Determine Associated Concept	Associated Concept Result
$C_1 \rightarrow C_2$	$O_1 = \{\}$ $O_2 = \{\text{Grape, Pear}\}$ $A_1 = \{\text{Fruit, Dairy_Product}\}$ $A_2 = \{\text{Fruit}\}$	$O_2 \supseteq O_1$ and $A_1 \supseteq A_2$ $O_1 \subseteq O_2$ and $A_2 \subseteq A_1$	C_2 is a Super Concept of C_1 C_1 is a Sub Concept of C_2
$C_1 \rightarrow C_3$	$O_1 = \{\}$ $O_3 = \{\text{Milk}\}$ $A_1 = \{\text{Fruit, Dairy_Product}\}$ $A_3 = \{\text{Dairy_Product}\}$	$O_3 \supseteq O_1$ and $A_1 \supseteq A_3$ $O_1 \subseteq O_3$ and $A_3 \subseteq A_1$	C_3 is a Super Concept of C_1 C_1 is a Sub Concept of C_3
$C_1 \rightarrow C_4$	$O_1 = \{\}$ $O_4 = \{\text{Grape, Pear, Milk}\}$ $A_1 = \{\text{Fruit, Dairy_Product}\}$ $A_4 = \{\}$	$O_4 \supseteq O_1$ and $A_1 \supseteq A_4$ $O_1 \subseteq O_4$ and $A_4 \subseteq A_1$	C_4 is a Super Concept of C_1 C_1 is a Sub Concept of C_4

Table 4-7: An Example of the Associated Concept Checking (Cont.)

Order Pair	Obj. / Attr.	Determine Associated Concept	Associated Concept Result
$C_2 \rightarrow C_3$	$O_2 = \{\text{Grape, Pear}\}$ $O_3 = \{\text{Milk}\}$ $A_2 = \{\text{Fruit}\}$ $A_3 = \{\text{Dairy_Product}\}$	$O_3 \not\supseteq O_2$ and $A_2 \not\supseteq A_3$ $O_2 \not\subseteq O_3$ and $A_3 \not\subseteq A_2$	- (Not Associated)
$C_2 \rightarrow C_4$	$O_2 = \{\text{Grape, Pear}\}$ $O_4 = \{\text{Grape, Pear, Milk}\}$ $A_2 = \{\text{Fruit}\}$ $A_4 = \{\}$	$O_4 \supseteq O_2$ and $A_2 \supseteq A_4$ $O_2 \subseteq O_4$ and $A_4 \subseteq A_2$	C_4 is a Super Concept of C_2 C_2 is a Sub Concept of C_4
$C_3 \rightarrow C_4$	$O_3 = \{\text{Milk}\}$ $O_4 = \{\text{Grape, Pear, Milk}\}$ $A_3 = \{\text{Dairy_Product}\}$ $A_4 = \{\}$	$O_4 \supseteq O_3$ and $A_3 \supseteq A_4$ $O_3 \subseteq O_4$ and $A_4 \subseteq A_3$	C_4 is a Super Concept of C_3 C_3 is a Sub Concept of C_4

Finally, as the result of the formal lattice generation process of the example 4-2, to represent the completed taxonomy, we can then formulate the graphical representation of the taxonomy as shown in Figure 4-7.

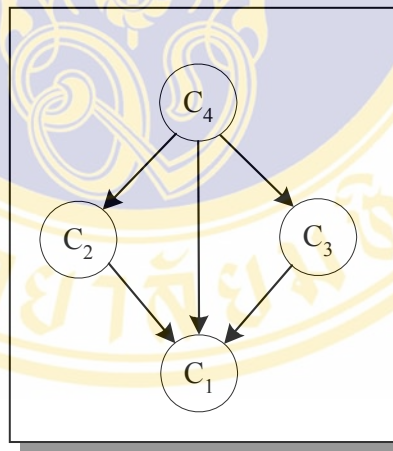


Figure 4-7: The Result of the Formal Lattice Generation process of the Example 4-2

5.4.1.2 The Schematic View of the Ontology Construction Process

This section summarizes the process of the ontology construction (described in section 5.4.1.1) which consists of five processes, namely, the Object Initialization, the Attribute Initialization, the Formal Context Generation, the Formal Concept Generation and the Formal Lattice Generation process. The schematic view of the ontology construction process is depicted in Figure 4-8.

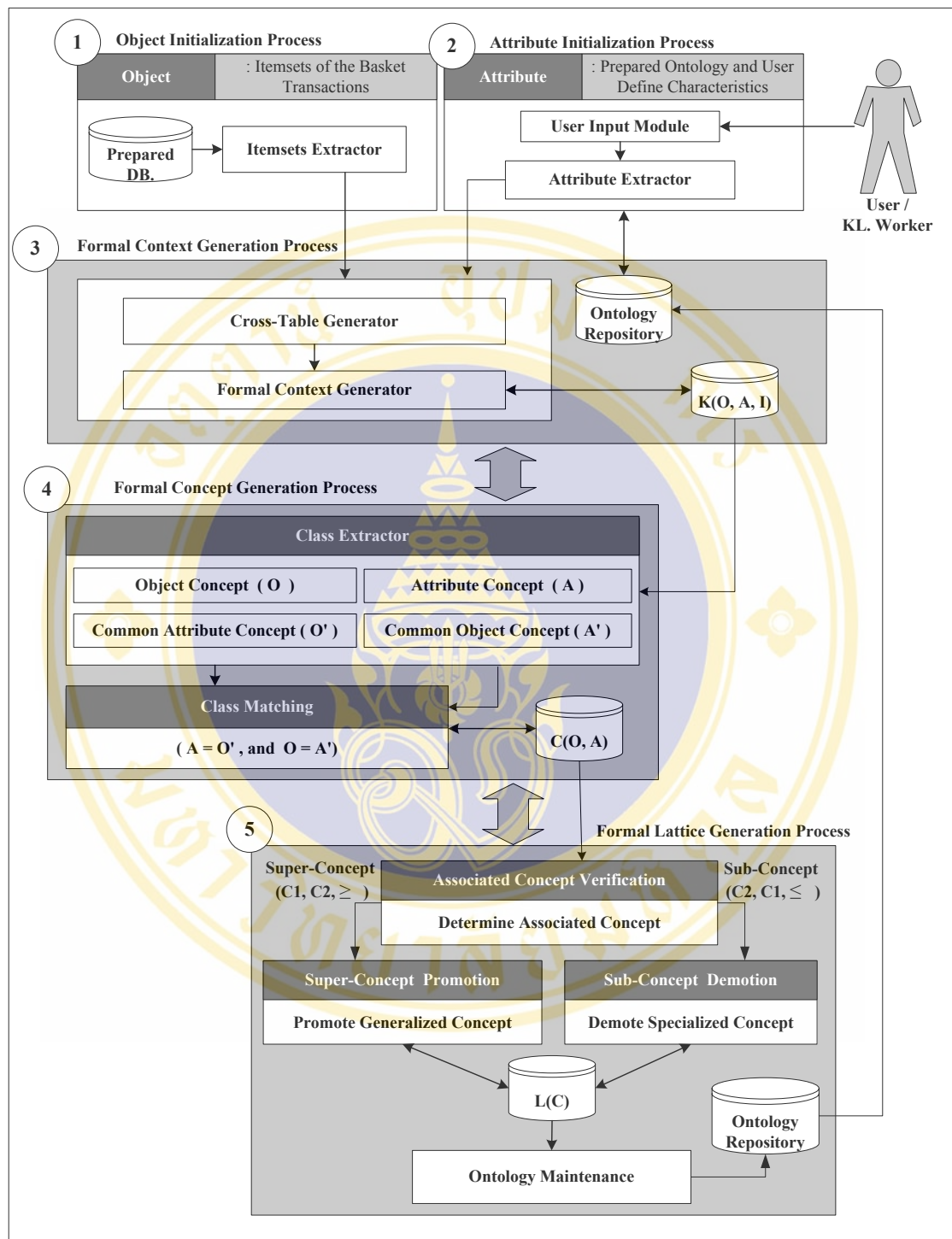


Figure 4-8: The Schematic View of the Ontology Construction Process

4.5 THE THIRD COMPONENT: Meta-Rules Guided Mining Construction

Module

After the completed taxonomy is generated, the process of the Meta-Rules Guided Mining Construction will be performed. To describe the whole process of this module, we state the theory of the “*Rules Implication on Lattice*” to implicate the related concepts of the taxonomy into the form of meta-rules [51].

4.5.1 Rules Implication on Lattice

The root idea of the rules implication on lattice is based on the “*Association Rules*” that state the relationship among the “*Itemsets*” of the transaction data. The association rules mining is closely related to the formal concept analysis, which was utilized in the second component (see section 4.4.1).

The association rules deal with the discovery of the association patterns among items in terms of the statistical definition. The statistics play an important role to discover the patterns. The quantitative measures are utilized to evaluate the discovered patterns, namely, Support and Confidence (referred to Section 2.2.2). An association rules is an implication $A \Rightarrow B$, where A and B are Itemsets. The intuitive meaning of such a rule is that the transactions contain A also contain B . The rules $A \Rightarrow B$ holds with confidence value of the transactions that contain A also contain B .

Identically, the rules implication on lattice is an implication rule on the formal concept C of the formal lattice L . As the same fashion of the association rules, we thus formalized the idea to simplify the rules implication on lattice, as the implication rule $X \Rightarrow Y$, where X and Y are the “*Attribute Concept*” elements of the formal lattice. The associated concept (also called the relation of the concepts) is involved in the formal lattice, and furthermore, which is represented in terms of the binary relation of the formal context. The associated concepts of the formal lattice are eventuated in the same path of the formal lattice.

Hence, the definition of the rules implication on lattice is defined as follows:

Definition 9: Rules Implication on Lattice “Let $X, Y \in \mathcal{A}$ be two sets of Attributes Concept. We say that X implied Y , if and only if any Object Z with the Attributes in X also has the Attributes in Y ”.

From the Definition 9, the *Rules Implication on Lattice* is defined as follows:

- $X \Rightarrow Y \Leftrightarrow \forall a \in A : (\forall x \in X : (o, a) \in I) \Rightarrow (\forall y \in Y : (o, a) \in I)$

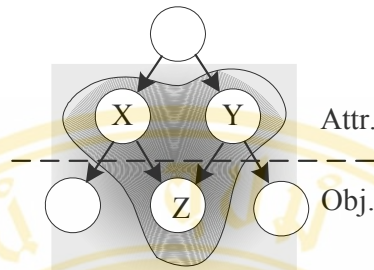


Figure 4-9: An example of Rules Implication on Lattice

As seen in Figure 4-9, $X \Rightarrow Y$, in case of both X and Y share the same object Z.

4.5.2 The Process of Meta-Rules Guided Mining Construction

We refer to the theory of the “*Itemset Discovery based on the FCA*” [27]. The relationships of the concepts in the formal lattice are determined to generate the associated concepts. The meta-rules are constructed via the associated concepts, which are similar to the association rules mining method.

The definition of the meta-rules guided mining is defined as follows:

Definition 10: Meta-Rules Guided Mining (MR) “An implication $X \Rightarrow Y$ is a triplet (X, Y, p) , where $X, Y \in \mathcal{A}$ are sets of Attributes Concept, and the *Precision* p is a Precision Value”

From the Definition 10, the Meta-Rules Guided Mining **MR** is defined as follows:

- $MR(X, Y, p)$ is a Meta-Rules Guided Mining
- $X, Y \in \mathcal{A} \mid \forall x \in X \text{ and } \forall y \in Y, (x, y) \in L(C)$
- $p = \text{Precision Value}.$

In addition, we apply the main idea of the association rules mining that aim to use the Min_Conf as the threshold values to determine the “*Frequent Itemset*” while process in the step of the rules generation. Consequently, we adapted the “*Precision Value*” [51] to be an important criterion to measure the strength of the meta-rules, and to determine the “*Frequent Concept*” of the formal lattice.

Return to the theoretical association rules mining method, we tend to discover the associated concepts that exceeded the precision value. The precision value is defined below:

Definition 11: Precision Value (p) “Precision Value is a quantitative measurement of the Meta-Rules which is used to determine the strength of the Meta-Rules.”

$$p(X, Y) = P(Y|X) = |\text{ext}(X \cup Y)| / |\text{ext}(X)|$$

Where $\text{ext}(X)$ is the extension part of the concept X

In addition, we state the definition of the extension as follows:

Definition 12: Extension “A formal concept of the formal context is defined as a pair (O, A) , where O is the object set of the formal concept. O is called the extension of the formal concept. Vice versa, A is the attribute set of the formal concept. A is called the intension of the formal concept.”

The Meta-Rules Guided Mining is constructed from the process of an implication on lattice which is also called an “*Inference Process*” that could be generated new meta-rules from the formal lattice.

To clearly understand the process of meta-rules guided mining construction module, consider an example of the “*Basket Transaction*” as depicted in Figure 4-10.

Basket Transaction	
Transactions	Items
1	Grape, Pear, Apple, Pork, Chicken
2.	Apple, Milk, Pork, Chicken
3.	Grape, Apple, Pork, Chicken
4.	Grape, Pear, Beef, Chicken
5.	Apple, Milk, Chicken
6.	Apple, Butter

Figure 4-10: An Example of the Basket Transaction Database

As seen in the Figure 4-10, an example of the “Basket Transaction” database is listed. The formal lattice is completely constructed from the first module, Ontology Construction Module. After that we calculate the “Precision Value” of each pair of concept. The results are shown in Figure 4-11.

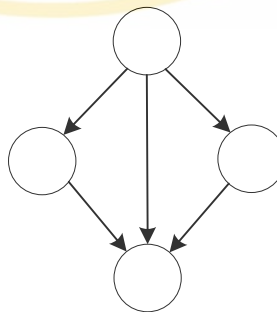


Figure 4-11: An Example of the Formal Lattice and Its Precision Value

As depicted in the Figure 4-11, the precision value is calculated from the basket transaction that shown in Figure 4-10 against the formal concepts that represent in example 4-2.

Let $C_1 = (\{\}, \{\text{Fruit, Dairy_Product}\})$

$C_2 = (\{\text{Grape, Pear}\}, \{\text{Fruit}\})$

$C_3 = (\{\text{Milk}\}, \{\text{Dairy_Product}\})$

$C_4 = (\{\text{Grape, Pear, Milk}\}, \{\})$

We can show the details of the calculation as below:

Ex1
$$p(C_4, C_1) = |\text{ext}(C_4 \cup C_1)| / |\text{ext}(C_4)|,$$

$$\text{ext}(C_4 \cup C_1) = \{\text{Grape, Pear, Apple}\} \cup \{\}, \text{ and}$$

$$\text{ext}(C_4) = \{\text{Grape, Pear, Apple}\}$$

$$= |\{\text{Grape, Pear, Apple}\} \cup \{\}| / |\{\text{Grape, Pear, Apple}\}|,$$

The basket transaction contains object sets $\{\text{Grape, Pear and Apple}\}$ which is transaction no 1,

$$|\{\text{Grape, Pear, Apple}\} \cup \{\}| = 1$$

and
$$|\{\text{Grape, Pear, Apple}\}| = 1$$

hence
$$= 1/1 = 1$$

Ex2
$$p(C_4, C_3) = |\text{ext}(C_4 \cup C_3)| / |\text{ext}(C_4)|$$

$$\text{ext}(C_4 \cup C_3) = \{\text{Grape, Pear, Apple}\} \cup \{\text{Milk}\}, \text{ and}$$

$$\text{ext}(C_4) = \{\text{Grape, Pear, Apple}\}$$

The basket transaction contain object set $\{\text{Grape, Pear and Apple}\} \cup \{\text{Milk}\}$ is NONE, hence $|\{\text{Grape, Pear, Apple}\} \cup \{\text{Milk}\}| = 0$

We summarize the precision value of each pair of concept of the taxonomy, as shown in the Figure 4-12.

Precision Value of each pair of concept	
$p(C_4, C_1)$	$= \text{ext}(C_4 \cup C_1) / \text{ext}(C_4) = 1$
$p(C_4, C_2)$	$= \text{ext}(C_4 \cup C_2) / \text{ext}(C_4) = 1$
$p(C_4, C_3)$	$= \text{ext}(C_4 \cup C_3) / \text{ext}(C_4) = 0$
$p(C_2, C_1)$	$= \text{ext}(C_2 \cup C_1) / \text{ext}(C_2) = 0$
$p(C_3, C_1)$	$= \text{ext}(C_3 \cup C_1) / \text{ext}(C_3) = 1$

Figure 4-12: An Example of the Method to Calculate the Precision Value of each pair of concept

According to our designed model, we repudiate the support value to purify the taxonomy. This implies that we concentrate our attention on the process of meta-rules guided mining construction that will generate the meta-rules from the entire concept element of the taxonomy. The meta-rules should be constructed from every concepts of the taxonomy. After that, we will traverse through the taxonomy in the level of “Generalized Level” to construct the meta-rules guided mining. We will calculate the “*Precision Value*” of each Meta-Rule as well.

4.5.2.1 The Process of Meta-Rules Guided Mining Construction

The GEN_MR algorithm is used to generate the meta-rules patterns (MR), which is performed by the Meta-Rules Guided Mining Construction module in the generalization part of the proposed framework. The algorithm consists of six steps:

1. Generate N-Large-Concepts
2. Generate the PowerSET for each element of N-Large-Concepts
3. Generate Order Pair of each PowerSET element
4. Extract Extent of each Order Pair from the Formal Concept
5. Calculate Precision value
6. Determine the Minimum Precision value (Min_P)

The details of each step and the examples are described as follows:

STEP 1: Generate N-Large-Concepts

The *N-Large-Concepts* are the number of “*Level*” in the taxonomy. In our methodology, the N value is greater than one. This implies that the ontology concept must be greater than one node. In worst case, if the generated ontology concept is equal to one, the GEN_MR algorithm will be discarded.

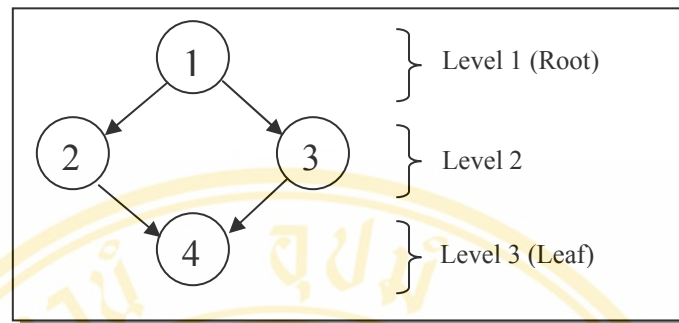


Figure 4-13: A General View of Ontology

According to the given general view of ontology in Figure 4-14, the N value equal to three, because it is the maximum number of the level in the taxonomy. The N-Large-Concepts can be derived as follows:

- 2-Large-Concept = $\{\{1,2\}, \{1,3\}, \{2,4\}, \{3,4\}\}$
- 3-Large-Concept = $\{\{1,2,4\}, \{1,3,4\}\}$

STEP 2: Generate the PowerSET for each element of N-Large-Concepts

The power set is designed to generate all combinations of each element in the N-Large-Concepts. For example, the PowerSET of each element of general view of taxonomy as depicted in Figure 4-14, are below:

- | | | |
|----|---------------|---|
| 1: | $\{1, 2\}$ | $\gg \{\{\}, \{1\}, \{2\}, \{1,2\}\}$ |
| 2: | $\{1, 3\}$ | $\gg \{\{\}, \{1\}, \{3\}, \{1,3\}\}$ |
| 3: | $\{2, 4\}$ | $\gg \{\{\}, \{2\}, \{4\}, \{2,4\}\}$ |
| 4: | $\{3, 4\}$ | $\gg \{\{\}, \{3\}, \{4\}, \{3,4\}\}$ |
| 5: | $\{1, 2, 4\}$ | $\gg \{\{\}, \{1\}, \{2\}, \{4\}, \{1,2\}, \{1,4\}, \{2,4\}, \{1,2,4\}\}$ |
| 6: | $\{1, 3, 4\}$ | $\gg \{\{\}, \{1\}, \{3\}, \{4\}, \{1,3\}, \{1,4\}, \{3,4\}, \{1,3,4\}\}$ |

STEP 3: Generate Order Pair of each PowerSET element.

The order pair of the PowerSET element is defined to generate all combinations of the elements. The example of the order pair of element $\{1,2\}$ and $\{1,3\}$ are shown below:

1: $\{1, 2\} \gg$

$\{\}$	$\rightarrow$	$\{1\}$
$\{\}$	$\rightarrow$	$\{2\}$
$\{\}$	$\rightarrow$	$\{1,2\}$
$\{1\}$	$\rightarrow$	$\{\}$
$\{1\}$	$\rightarrow$	$\{2\}$
$\{1\}$	$\rightarrow$	$\{1,2\}$
$\{2\}$	$\rightarrow$	$\{\}$
$\{2\}$	$\rightarrow$	$\{1\}$
$\{2\}$	$\rightarrow$	$\{1,2\}$
$\{1,2\}$	$\rightarrow$	$\{\}$
$\{1,2\}$	$\rightarrow$	$\{1\}$
$\{1,2\}$	$\rightarrow$	$\{2\}$

2: $\{1, 3\} \gg$

$\{\}$	$\rightarrow$	$\{1\}$
$\{\}$	$\rightarrow$	$\{3\}$
$\{\}$	$\rightarrow$	$\{1,3\}$
$\{1\}$	$\rightarrow$	$\{\}$
$\{1\}$	$\rightarrow$	$\{3\}$
$\{1\}$	$\rightarrow$	$\{1,3\}$
$\{3\}$	$\rightarrow$	$\{\}$
$\{3\}$	$\rightarrow$	$\{1\}$
$\{3\}$	$\rightarrow$	$\{1,3\}$
$\{1,3\}$	$\rightarrow$	$\{\}$
$\{1,3\}$	$\rightarrow$	$\{1\}$
$\{1,3\}$	$\rightarrow$	$\{3\}$

Note: No Self Determination, such as $\{1\} \rightarrow \{1\}$

STEP 4: Extract Extent of each Order Pair from the Formal Concept

According to the theory of the formal concept analysis (FCA), each formal concept consists of two parts, namely, extent part, and intent part. The extent is referred to the “*Objects*” while the intent is referred to the “*Attributes*” of the domain. The FCA theory specifies that we need to extract the extent from the formal concept to construct the meta-rules patterns.

STEP 5 Calculate Precision Value

The precision value of each element in the PowerSET of each N-Large-Concepts is calculated. The example equation is derived as follow:

$$\text{Precision Value (P)} = \frac{\text{Count}(\text{Extent}(\text{LHS}) \cup \text{Extent}(\text{RHS}))}{\text{Count}(\text{Extent}(\text{LHS}))}$$

$$1: \quad P(\{\} \rightarrow \{1\}) = \frac{\text{Count}(\text{Extent}(\{\}) \cup \text{Extent}(\{1\}))}{\text{Count}(\text{Extent}(\{\}))}$$

$$2: \quad P(\{1\} \rightarrow \{2\}) = \frac{\text{Count}(\text{Extent}(\{1\}) \cup \text{Extent}(\{2\}))}{\text{Count}(\text{Extent}(\{1\}))}$$

STEP 6 Determine Minimum Precision Value (Min_P)

The Min_P is the minimum precision value which is accepted from the user. The last step is to determine the Min_P value of each pair in the PowerSET. If it exceeds the Min_P, the MR patterns will be constructed.

Hence, for a more clearly understanding, consider the PowerSET $\{\{\}, \{1\}, \{2\}, \{1,2\}\}$, of an order pair (C_1, C_2) in the 2-Large-Concepts. The precision value of each pair of an element can be calculated, such as, precision value of the pair $\{\} \rightarrow \{1\}$ is equal to 60; precision value of $\{1\} \rightarrow \{2\}$ is equal to 40; precision value of $\{1\} \rightarrow \{1,2\}$ is equal to 78; and precision value of $\{2\} \rightarrow \{1,2\}$ is 80, respectively.

Suppose the Min_P is equal to 60, hence the pair $\{1\} \rightarrow \{2\}$ is discarded. However, the pair $\{\} \rightarrow \{1\}$, $\{1\} \rightarrow \{1, 2\}$, and $\{2\} \rightarrow \{1, 2\}$ can be determined as the Meta-Rules. The $\text{MR1}(C_1, C_1C_2, 78)$ is derived from the pair $\{1\} \rightarrow \{1, 2\}$, as well.

4.5.2.2 The Schematic View of the Meta-Rules Guided Mining Construction Process

In summary, we can summarize the Meta-Rules Guided Mining Construction process which consists of six processes, namely, the N-Large-Concepts Generation, the PowerSET of N-Large-Concepts Generation, the Order Pair of each PowerSET Element Generation, the Extension Extraction, the Precision Value Calculation and the Precision Value Determination. The schematic view of the Meta-Rules Guided Mining Construction process is presented in Figure 4-14.

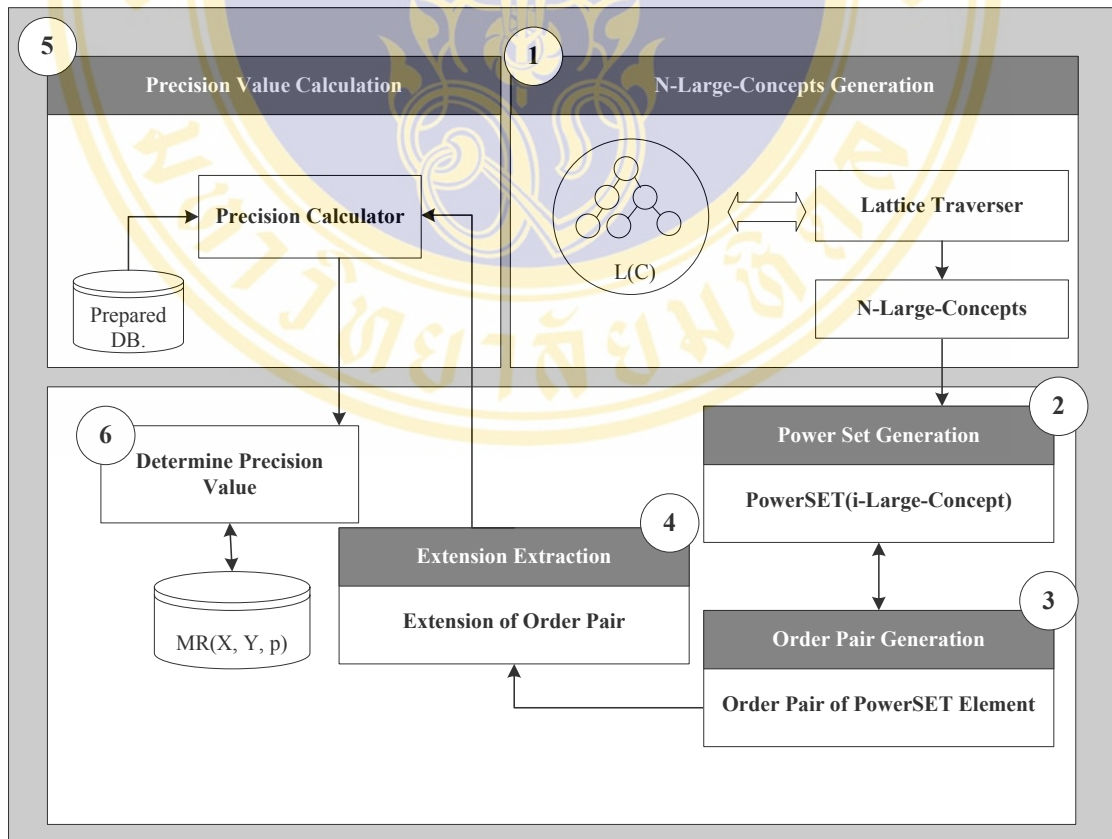


Figure 4-14: The Schematic View of the Meta-Rules Construction Process

4.6 THE FOURTH COMPONENT: Interestingness Patterns Matching Module

4.6.1 The Process of Interestingness Patterns Matching

The system matches each discovered pattern $dr_i \in DR$ against each meta-rules pattern $mr_j \in MR$ in order to obtain the degree of match for each discovered pattern dr_i against the set of MR . The discovered patterns in DR are then ranked according to their degree of match with MR .

Definition 13: Degree of Match

The degree of match, between two patterns $DR = (\{Ante\}, \{Cons\}, Supp, Conf)$ and $MR = (\{Ante\}, \{Cons\}, Precision)$ is defined as:

$$\text{Degree of Match} = \begin{cases} \frac{|\text{DR's Element match with MR's Element}|}{|\text{Total number of MR's Element}|} & \text{If the patterns match} \\ 0 & \text{Otherwise} \end{cases}$$

The degree of match is the statistical measurement of the matching process between the selected MR patterns and DR Patterns. The aim of this measurement is to evaluate the strength of the interestingness. If there is a high degree of match generated during the process of interestingness patterns matching, then the higher the degree of match the more interesting the pattern is.

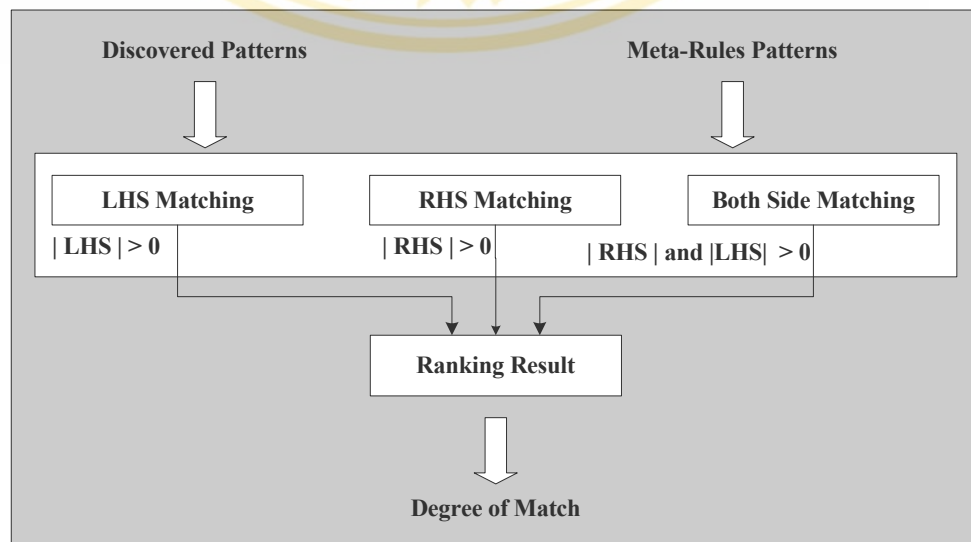


Figure 4-15: The process of interestingness patterns matching

As seen in Figure 4-15, the matching process consists of four sub modules, namely, LHS Matching, RHS Matching, Both Side Matching, and Ranking Result.

□ **LHS Matching**

The left hand side matching is determined when the left hand side of the Discovered Pattern's element matches with the left hand side of the Meta-Rules Pattern's element.

□ **RHS Matching**

The right hand side matching is determined when the right hand side of the DR's element matches with the right hand side of the Meta-Rules Pattern's element.

□ **Both Side Matching**

The both side matching is determined when the left and right hand side of the DR's element matches with the left and right hand side of the Meta-Rules Pattern's element.

□ **The Ranking Result Module**

For the ranking of the matching result of this stage, the result of LHS Matching, RHS Matching, and Both Side Matching are combined. In this combination, the system produces the degree of match, conforming to the Definition 10.

Moreover, the system ranks the interestingness results with two evaluation factors to obtain the subjective interestingness patterns conforming to the user's needs, namely, the recall of interestingness and the precision of interestingness. The detail of each factor is described below:

1. Recall of Interestingness

The recall of interestingness is the proportion of relevant patterns that are actually retrieved in answer to a search request. The recall of interestingness can answer the user question, i.e., have user found all relevant information? The equation for calculation the recall value is shown below:

$$\text{Recall}(\text{"Interestingness"}) = \frac{\text{\#Results that are classified as Interestingness and are actually Interestingness}}{\text{\#Results that are classified automatically and are actually Interestingness}}$$

2. Precision of Interestingness

The precision of interestingness is the proportion of retrieved patterns that are actually relevant to user needs. The precision of interestingness can answer the user question, i.e., how many interestingness results are really relevant to user needs? The equation for calculating the precision value is shown below:

$$\text{Precision ("Interestingness")} = \frac{\text{\#Results that are classified as Interestingness and are actually Interestingness}}{\text{\#Results that are classified as Interestingness}}$$

For the implementation of the two last factors, we described the details in the section 6.4.3.

CHAPTER 5

DESIGN AND IMPLEMENTATION

In this chapter, we describe the design and implementation of our proposed model. This chapter firstly describes the overall implementation details and functional design of our prototype system, namely FCA2MR. The system design includes structure chart, data flow diagram, and ER diagram. Hereafter, the ontology component is designed to show how the ontology process is constructed and how it is used for the implementation. Finally, the FCA2MR interface design is provided.

5.1 The Implementation of A Unified Framework of Meta-Rules Guided Mining Based on Ontology

We demonstrate and verify our proposed model by developing the prototype system called “FCA2MR”. FCA stands for Formal Concept Analysis, the technique that we proposed to define the Ontology Construction Module. MR stands for Meta-Rules, the Meta-Rules Guided Mining technique is defined as the Generalized Rules Discovery Module which is used for generating the meta-rules patterns.

The prototype is designed to purify the discovered patterns which are generated from the traditional data mining task defined in the specialized rules discovery. According to the proposed framework presented in section 4.2, the main process of the framework is the generalized rules discovery. This main process contains the mechanism to construct the ontology which is used to represent the data schema. The comprehensive ontology is then used to develop the Meta-Rules patterns. Finally, the last process is the mapping process between the Meta-Rules patterns and the discovered patterns to purify the results as the interestingness patterns.

Moreover, our prototype system provides two browsers: Ontology Browser, and the Interestingness Browser, to view the ontology categorical concept of rules and the interestingness results of the system. User can analyze and preview the results autonomously.

The next section, we describe the system design of our prototype system.

5.2 System Design

5.2.1 Structure Chart

Figure 5-1 illustrates the structure chart of a Unified Framework of Meta-Rules Guided Mining based on Ontology, which consists of two main parts: specialized rules discovery, and generalized rules discovery.

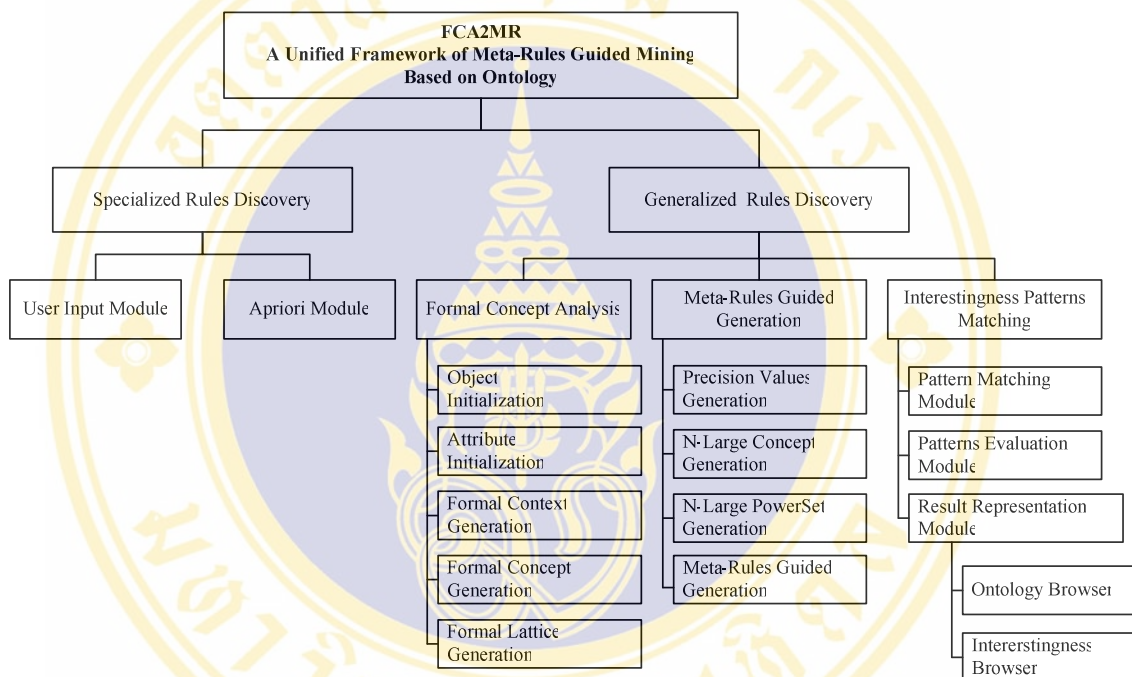


Figure 5-1: The Structure Chart of A Unified Framework of Meta-Rules Guided Mining Based on Ontology

In the next section, we describe the design and implementation of the proposed model in details.

5.2.2 Data Flow Diagrams (DFD)

Data flow diagrams (DFD) of the prototype system have been explored into four levels. Figure 5-2 shows the context diagram or DFD level 0. Figure 5-3 illustrates DFD level 1. Figure 5-4 to 5-5 shows DFD level 2, and Figure 5-6 to 5-8 shows DFD level 3, respectively.

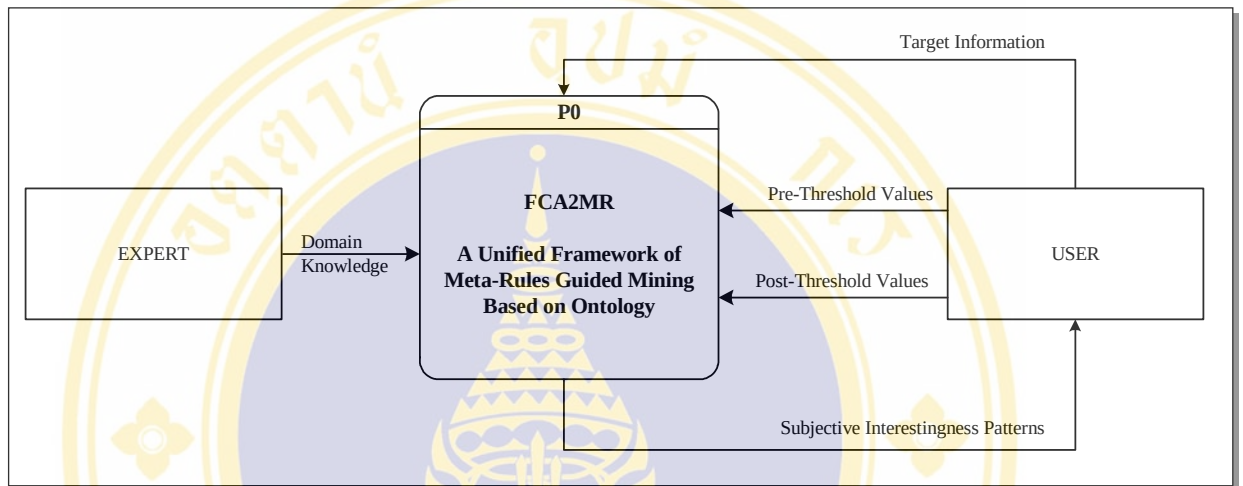


Figure 5-2: A Data Flow Diagram: The Context Diagram

The prototype system consists of two major processes described as follows:

Process 1: Specialized Rules Discovery

This process serves for the crucial task of data mining by using an Apriori algorithm to discover the patterns of data from prepared database. The results of this process will be matched with the Meta-Rules patterns to produce the interestingness results in the generalized rules discovery process.

Process 2: Generalized Rules Discovery

This process consists of three sub-processes, namely, the formal concept analysis (FCA), the meta-rules guided mining construction, and the interestingness patterns matching.

In contrast, our prototype system is based on the iterative and interactive process in the generalized rules discovery process while the traditional data mining system interacts with the users in the specialized rules discovery part. In our system,

the generalized rules represent the Meta-Rules patterns which are useful for the user to purify the data mining results in terms of the interestingness patterns.

The data flow diagram of the two major processes is illustrated in Figure 5-3.

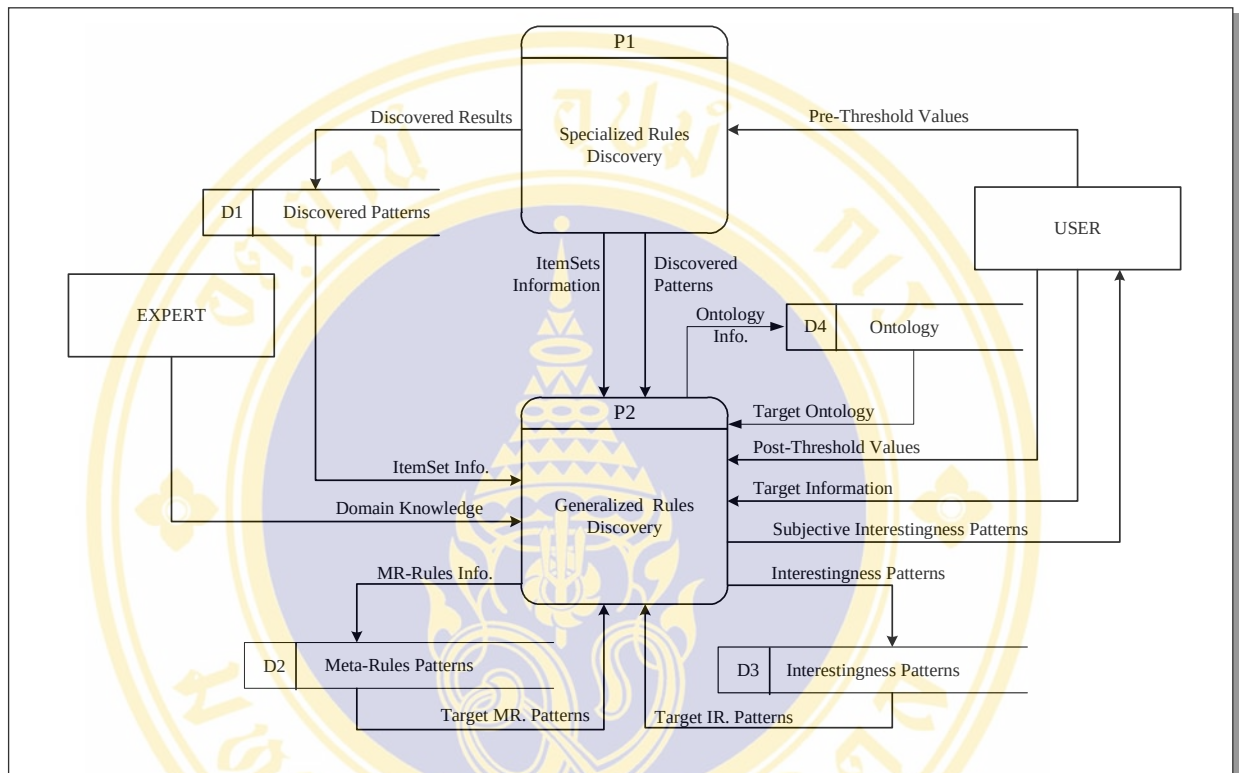


Figure 5-3: A Data Flow Diagram Level 1

First, the specialized rules discovery process consists of two sub modules, namely, the user input module and the Apriori module. The user input accepts the pre-threshold value from the user and sends the information to the Apriori module. The discovered patterns are generated and stored in the repository. This process is illustrated in Figure 5-4.

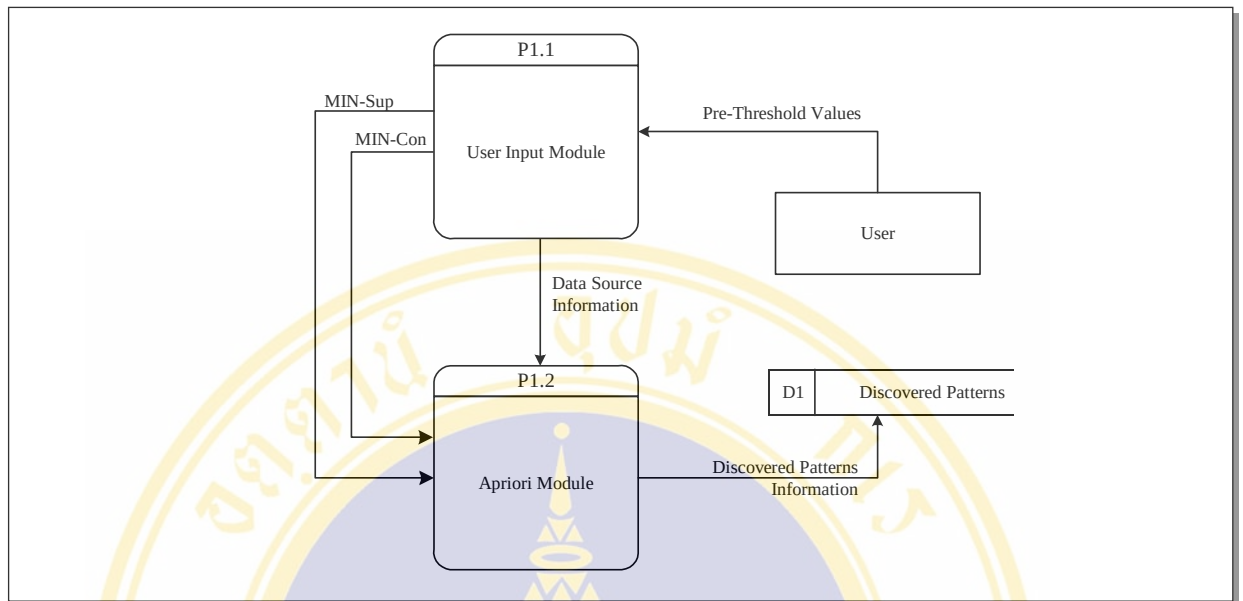


Figure 5-4: A Data Flow Diagram Level 2 – Process 1: Specialized Rules Discovery

Second, the generalized rules discovery process consists of three sub processes, namely, the Formal Concept Analysis, the Meta-Rules Guided Generation and the Interestingness Patterns Matching process. The Formal Concept Analysis process is designed to construct the ontology and the ontology results are then used to generate the Meta-Rules patterns. Finally, the Meta-Rules patterns will be matched with the discovered patterns to generate the interestingness patterns results. This process is illustrated in Figure 5-5.

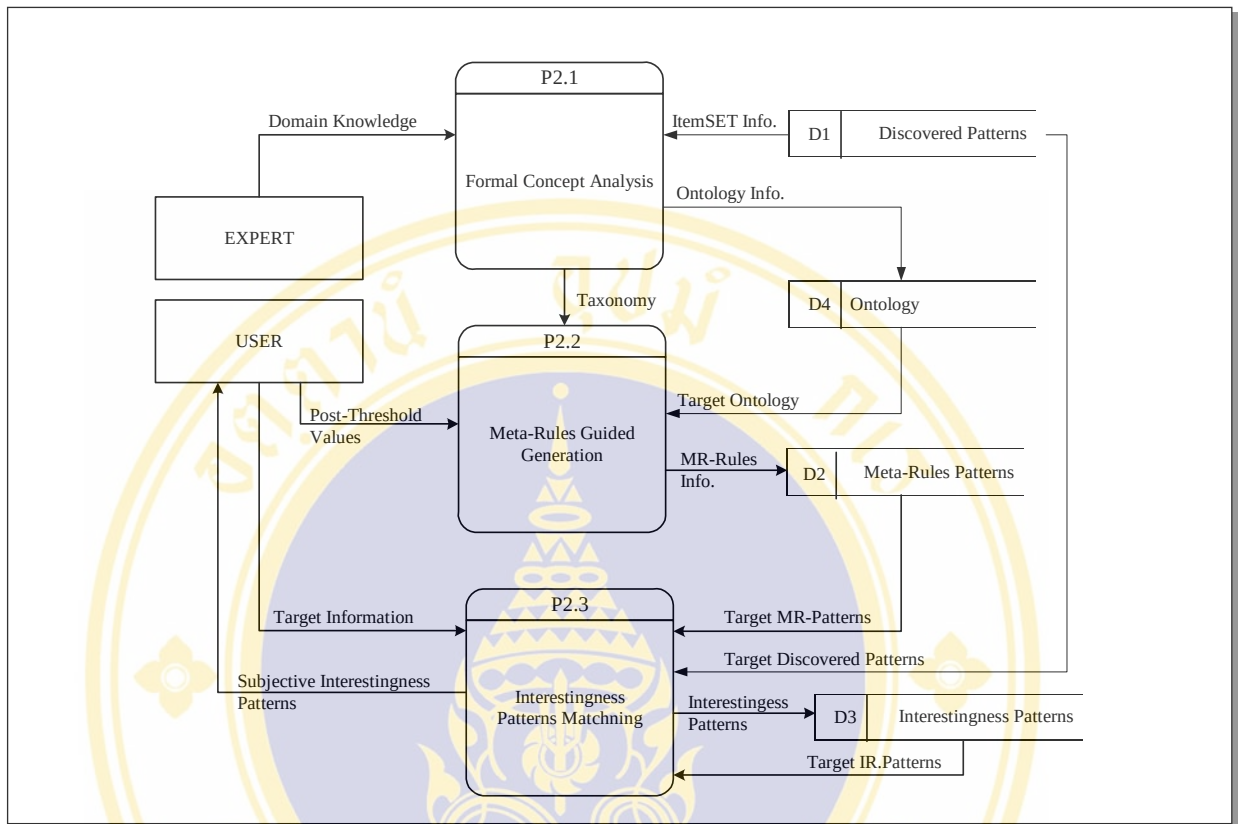


Figure 5-5: A Data Flow Diagram Level 2 – Process 2: Generalized Rules Discovery

The Formal Concept Analysis process consists of five sub processes, namely, the Object Initialization, the Attribute Initialization, the Formal Context Generation, the Formal Concept Generation and the Formal Lattice Generation. This process is illustrated in Figure 5-6.

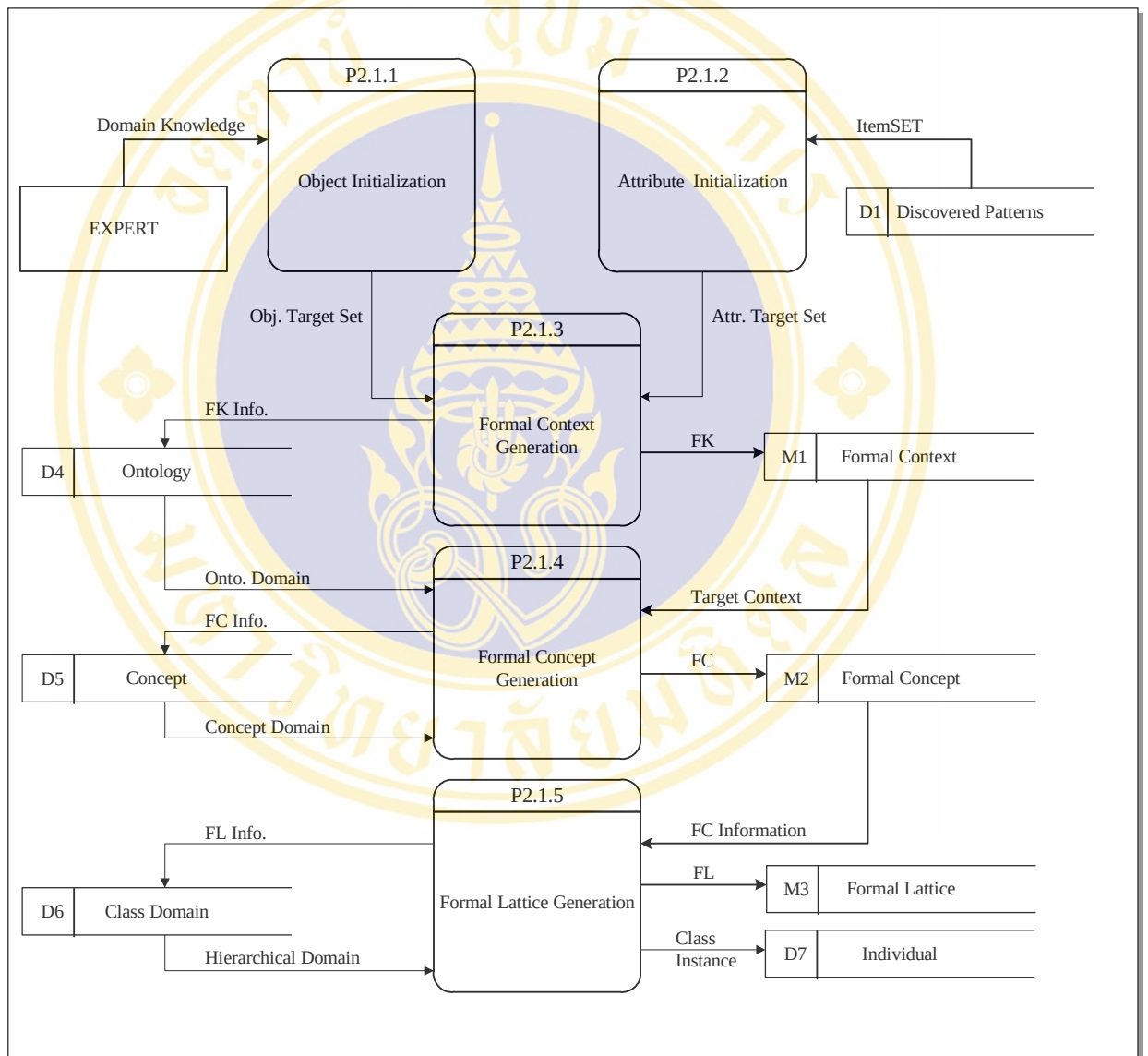


Figure 5-6: A Data Flow Diagram Level 3 – Process 2.1: Formal Concept Analysis

The Meta-Rules Guided Generation process consists of four sub processes, namely, the Precision Value Generation, the N-Large-Concept Generation, the N-Large Concept PowerSET Generation and the Meta-Rules Guided Generation. This process is illustrated in Figure 5-7.

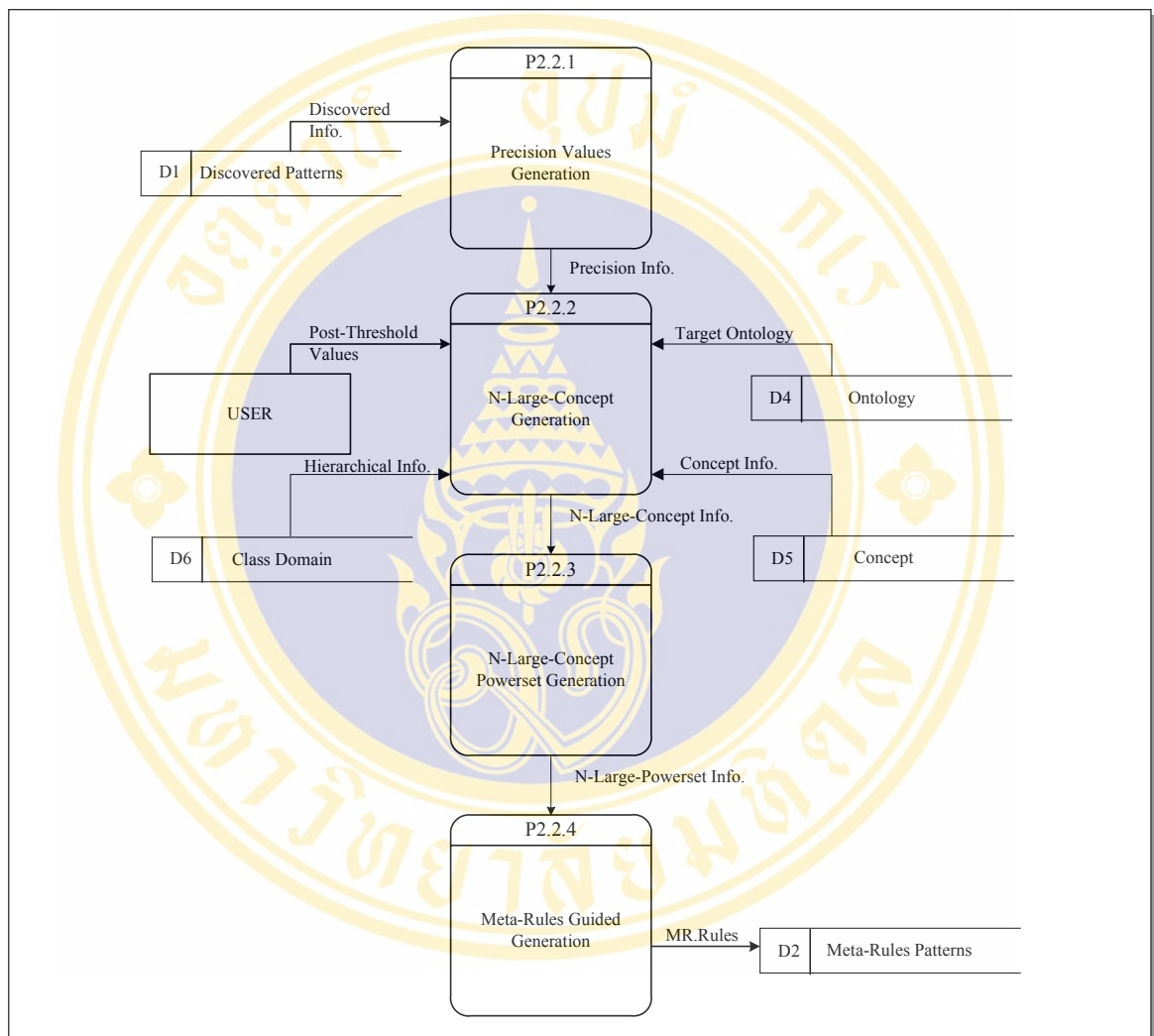


Figure 5-7: A Data Flow Diagram Level 3 – Process 2.2: Meta-Rules Guided Generation

The Interestingness Patterns Matching process consists of three modules, namely, the Pattern Matching Module, the Patterns Evaluation Module and the Result Representation Module. This process is illustrated in Figure 5-8.

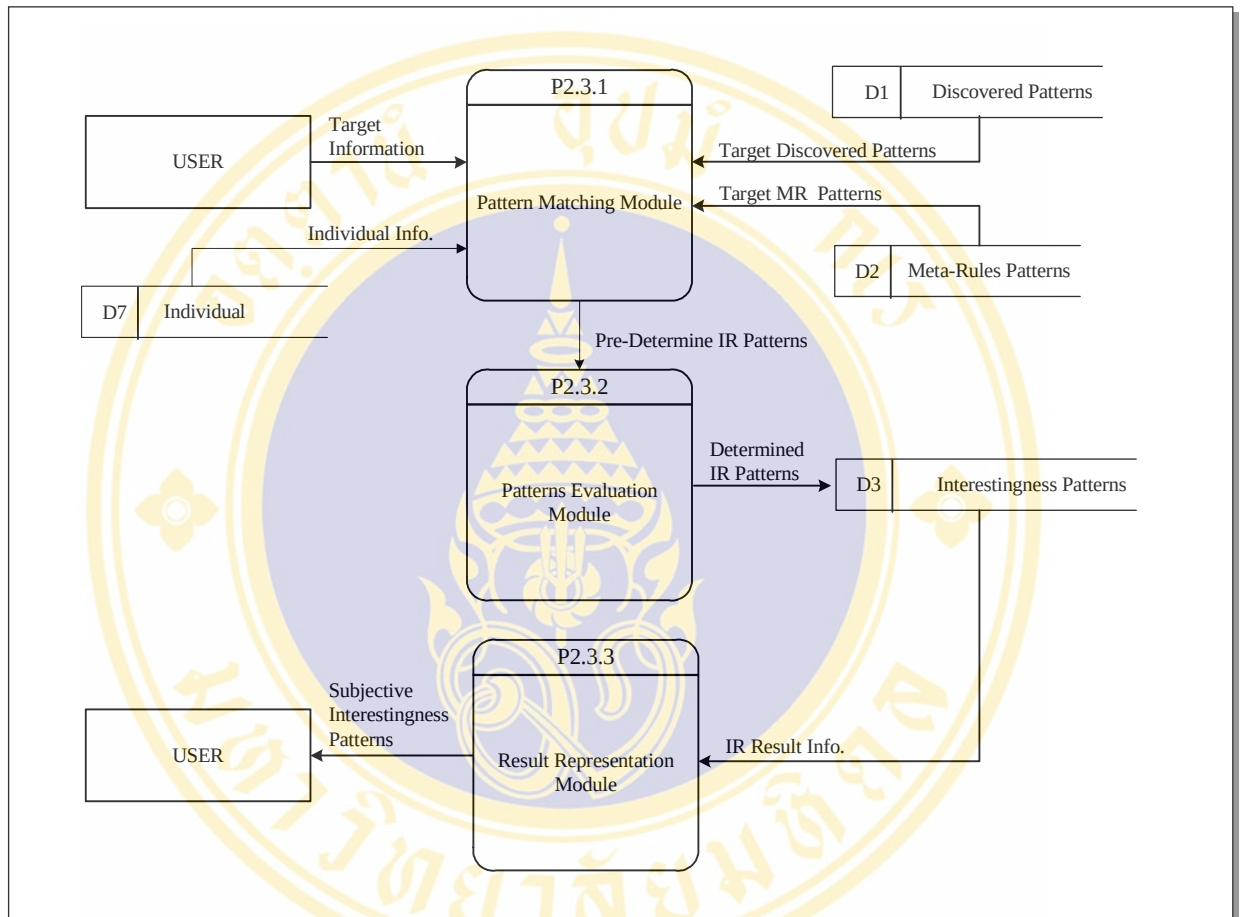


Figure 5-8: A Data Flow Diagram Level 3 – Process 2.3: Interestingness Pattern Matching

5.2.3 Entity Relationship Diagram (E-R Diagram)

We design and model the Entity Relationship to support the database design of our prototype system. An E-R model provides a synopsis view of the relationship among the entities. We use the derived entities to describe the on-the-fly entities that are generated during execution. The derived entities are drawn in the dotted line. Figure 5-9 presents the E-R diagram of the prototype system.

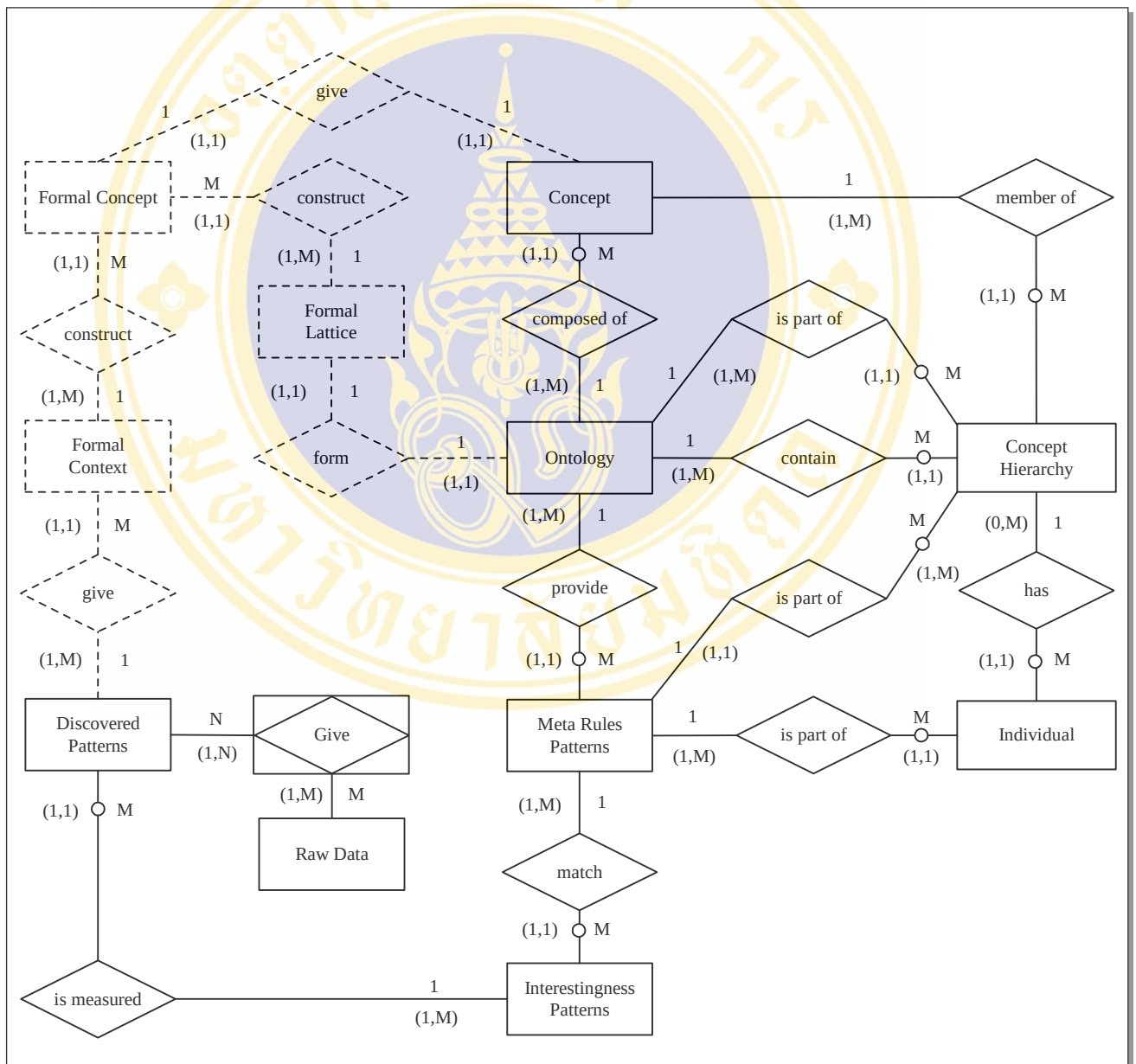


Figure 5-9: An Entity Relationship Diagram (E-R)

5.2.4 Database Designed Relation

The database designed relations are referred as the main entities of the prototype system. We define the seven relations schema to support the development process of the prototype system. The details are described below:

1) Ontology Relation

(**Domain**: String NOT NULL UNIQUE,
ObjMember: String,
AttrMember: String,
RelationMember: String,
NumberOfConcept: Integer,
CreateDate: Date Time)

An Example

<u>Domain</u>	<u>ObjMember</u>	<u>AttrMember</u>	<u>RelationMember</u>	<u>NumberOfConcept</u>	<u>CreateDate</u>
Animal	Cats, Dogs, Whales ...	Four-Legged, Animal ...	1:1, 1:3, 2:1 ...	6	2005-02-02
Food	Grape, Pare, Milk	Fruit, Meat, Daily_Product	1:1, 1:2, 1:3, 2:2 ...	12	2005-02-05

2) Concept Relation

(**CID**: Integer NOT NULL,
Domain: String NOT NULL,
ExtentMember: String,
IntentMember: String,
CreateDate: Date Time)

An Example

<u>CID</u>	<u>Domain</u>	<u>ExtentMember</u>	<u>IntentMember</u>	<u>CreateDate</u>
1	Animal	Whales	Marine, Animal	2005-02-02
2	Animal	Cats, Dogs	Four-Legged, Animal	2005-02-02
3	Animal	Human	Intelligent	2005-02-02
4	Animal	Cats, Dogs, Whales	Animal	2005-02-02

3) Concept Hierarchy Relation

(**RelationID**: Integer NOT NULL,
ParentID: String NOT NULL,
ChildID: String NOT NULL,
ParentExtent: String,
ParentIntent: String,
ChildExtent: String,
ChildIntent: String,
Domain: String NOT NULL)

An Example

<u>RelationID</u>	<u>ParentID</u>	<u>ChildID</u>	<u>ParentExtent</u>	<u>ParentIntent</u>	<u>ChildExtent</u>	<u>ChildIntent</u>	<u>Domain</u>
1	1	3	Pear, Grape	Fruit	Pear	Fruit	Mankind
2	2	4	Pare	Fruit	Pear	Fruit	Mankind
3	2	4	Pare	Fruit	Pear	Fruit	Mankind
4	2	5	Pare	Fruit	Pear	Fruit	Mankind
5	3	6	Mile	DailyProduct	Pear	Fruit	Mankind

4) Individual Relation

**(IndividualName: String NOT NULL,
InstanceOFClass: String NOT NULL)**

An Example

IndividualName	InstanceOFClass	Remark
PREECHA	Male	PREECHA is a Male
SOMSRI	Girl	SOMSRI is a Girl
PAIBOON	Boy	PAIBOON is a Boy
PAYAO	Female	PAYAO is a Female
JOE	Person	JOE is a Person

5) Discovered Patterns Relation

**(DID: Integer NOT NULL UNIQUE,
Ante: String NOT NULL,
Cons: String NOT NULL,
SUPP: Integer NOT NULL,
CONF: Integer NOT NULL,
MinS: Integer NOT NULL,
MinC: Integer NOT NULL,
DataSET: String NOT NULL)**

An Example

DID	Ante	Cons	SUPP	CONF	MinS	MinC	DataSET
1	Sex=Female	Hours=FullTime	33.2	64.2	20	50	Census
2	Age=Middle	Sex=Male	48.4	72.2	20	50	Census
3	Race=White	Hours=FullTime	66.8	80.5	20	50	Census

6) Meta Rules Patterns Relation

**(MID: Integer NOT NULL,
Domain: String NOT NULL,
Ante: String NOT NULL,
Cons: String NOT NULL,
Precision: Integer NOT NULL,
MinP: Integer NOT NULL)**

An Example

MID	Domain	Ante	Cons	Precision	MinP
1	Mankind	Male	Mankind	80.5	50
2	Mankind	Father	Male	100	50
3	Food	Fruit	Daily_Product	60.2	20

7) Interestingness Patterns Relation

**(IID: Integer NOT NULL,
DID: Integer NOT NULL,
MID: Integer NOT NULL,
Precision: Integer NOT NULL,
MinP: Integer NOT NULL,
CreateDate: Date Time,
SatisfySTS: String)**

An Example

IID	DID	MID	Precision	MinP	CreateDate	SatisfySTS
1	1	2	87	50	2005-02-05	Y
2	2	2	55	50	2005-02-05	Y
3	5	2	98	50	2005-02-05	Y

5.2.5 System Defined Relation

According to the Entity-Relation Diagram (E-R), the system-defined relations is determined to support the development process of the prototype system. These relations are made up while the system is executed. The details are described below:

1) ElementSET Relation (UnitSET (): String)

An Example

Declaration	DIM objSET as ElementSET REDIM objSET(1 to 2)
Stored Values	objSET.UnitSET(1) = "object1" objSET.UnitSET(2) = "object2"

2) PowerSET Relation (Power: Integer, Element (): ElementSET, OutputMember: Integer)

An Example

Declaration	DIM PSet as PowerSET REDIM PSet.Element(1 to PSet.Power)
Stored Values	PSet.Power = 3 PSet.Element(1) = "A" PSet.Element(2) = "AB" ... PSet.OutPutMember = $2 \wedge \text{PSet.Power}$

3) FormalContext Relation

**(ObjSET(): String,
AttrSET(): String,
RelationSET(): String)**

An Example

Declaration	DIM FK as FormalContext REDIM FK.ObjSET(1 to 2) REDIM FK.AttrSET(1 to 2) REDIM FK.RelationSET(1 to 2)
Stored Values	FK.ObjSET(1) = "object1" FK.ObjSET(2) = "object2" FK.AttrSET(1) = "attribute1" FK.AttrSET(2) = "attribute2" FK.RelationSET (1) = "1:1" FK.RelationSET (2) = "2:3"

4) FormalConcept Relation

**(ObjSET(): String,
AttrSET(): String)**

An Example

Declaration	DIM FC as FormalConcept REDIM FC.ObjSET(1 to 2) REDIM FC.AttrSET(1 to 2)
Stored Values	FC.ObjSET(1) = "object1" FC.ObjSET(2) = "object2" FC.AttrSET(1) = "attribute1" FC.AttrSET(2) = "attribute2"

5) FormalLattice Relation

**(RelationID: Integer,
ParentID: Integer,
ChildID: Integer,
ParentExtentSET(): String,
ParentIntentSET(): String,
ChildExtentSET(): String,
ChildIntentSET(): String)**

An Example

Declaration	DIM FL as FormalLattice REDIM FL.ParentExtentSET(1 to 2) REDIM FL.ParentIntetSET (1 to 2) REDIM FL.ChildExtentSET (1 to 2) REDIM FL.ChildIntentSET (1 to 2)
-------------	---

Stored Values	FL.RelationID=1 FL.ParentID=2 FL.ChildID=3 FL.ParentExtentSET(1) = “Object1” FL.ParentIntentSET(1) = “Attribute2” FL.ChildExtentSET(1) = “Object2” FL.ChildIntentSET(1) = “Attribute1”
---------------	--

5.3 Ontology Components Design

We design the ontology components to support the process of the ontology construction based on the FCA theory. The ontology design contains the following components.

1. Ontology Concept

The ontology concept represents the entities of the domain. They can be either concrete or abstract. Accordingly, they can have their own meaning to conceptualize the domain of ontology.

The ontology concepts contain three types of information: Concept ID, Extent Member and Intent Member.

- **Concept ID** identifies the number of ontology concept.
- **Extent Member** represents the objects set of the domain of interest.
- **Intent Member** represents the attributes set of the domain of interest.

An Example of the Ontology Concept of the Car_C_2 Domain is described below:

Concept ID : 1

Extent Member : {Lug-Boot}

Intent Member : {Med, Big, Small}

2. Relationship

The relationship between the ontology concepts are characterized by their type, origin (source/parent), and destination (target/child) of the concepts. The cardinality relationship between the origin and destination is set to one-to-one. The following are details of the relations:

- **Relationship ID** identifies the number of the relations.
- **Relationship Type** represents the traditional relationship between the ontology concepts. In our system, we restrict the relationship type to “IS-A”.
- **Parent Concept** represents the origin or source concept.
- **Child Concept** represents the destination or target concept.

An Example of Relationship is described below:

Relationship ID: 1

Relationship Type: IS-A

Parent Concept: ConceptID7

Child Concept: ConceptID8

3. Individual

The individual describes the properties of the ontology concepts by considering the behavior model conceptualized by the certain business logic which is mapped into the specific ontology domain. In fact, one ontology domain is composed of several concept members having their own values which are inserted as the instance members. These values are interpreted as explicit values for the ontology concepts. The following are details of the individuals:

- **Individual ID** identifies the number of the individuals.
- **Individual Label** represents the intension of the ontology concept. The intension is represented in terms of the attribute set of the domain of interest.
- **Instance Member** represents the member of the object set of the ontology concept.

The following is the example of the individual of the Car_C_2 Domain:

An Example of Individuals

Individual ID: 1

Individual Label: Big

Instance Member: {Lug-Boot}

These ontology components are integrated to accomplish a concrete purpose of the main process of the ontology construction module in our proposed framework (defined in section 4.3). The prototype system uses these ontology components to generate the ontology and represent it which is presented in the graphical view via the ontology browser.

Using the ontology components at the development time enables the system to conceptualize a “higher” level of domain knowledge. Moreover, it enables the user to understand the domain knowledge using a common terminology. It also enables the user to concentrate on the structure of the domain and support the user transparency from the implementation details.

The next section, we describe the system function of the FCA2MR prototype system.

5.4 FCA2MR System Functions

According to the system design, the FCA2MR system functions are separated into two main processes: the specialization rules mining function and the generalization rules mining function. The details are described below.

5.4.1 Specialization Rules Mining Function

Based on the traditional data mining, the specialization rules mining function generates the discovered patterns by using the Apriori algorithm. There are two results from this task. First, the frequent itemsets are the results of the first step of the Apriori algorithm, and the association rules are the results of the last step of the Apriori algorithm.

The Apriori algorithm is decomposed into two main processes:

- 1) Discover the large itemsets, i.e., the sets of itemsets that have transaction support above a predetermined minimum support (Min-Supp).
- 2) Generate the association rules for the database using the large itemsets.

In the first process, the Apriori algorithm uses the “Support” as the statistical measurement to indicate the frequencies of the derived patterns in the rule. It is desirable to pay attention only those rules which have a reasonable large support. The results of this step are called the “Frequent itemsets”.

Figure 5-10 illustrates an example of the 5,850 frequent itemsets generated from the adult dataset. The minimum support used in this example is set to 5 percent.

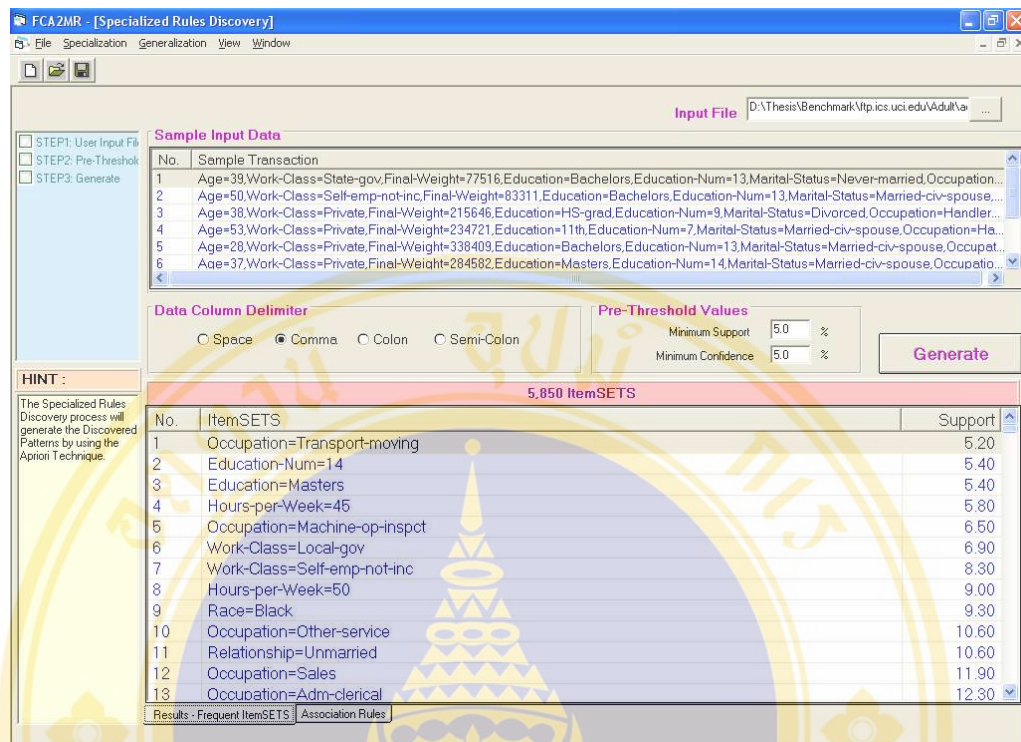


Figure 5-10: Frequent ItemSETS Results of Adult DataSET (Min_Supp=5)

Second, the Apriori algorithm uses the “Confidence” as the statistical measurement to denote the strength of the implication. The confidence is defined as the percentage of transactions that contain X also contain B. It indicates the degree of correlation between left-hand-side and right-hand-side of the rule.

Figure 5-11 illustrates an example of the 92,426 specialized rules results. The minimum confidence used in this example is setting to 5 percent. This is the last step of the Apriori algorithm, which aims to produce the order pair of association rules.

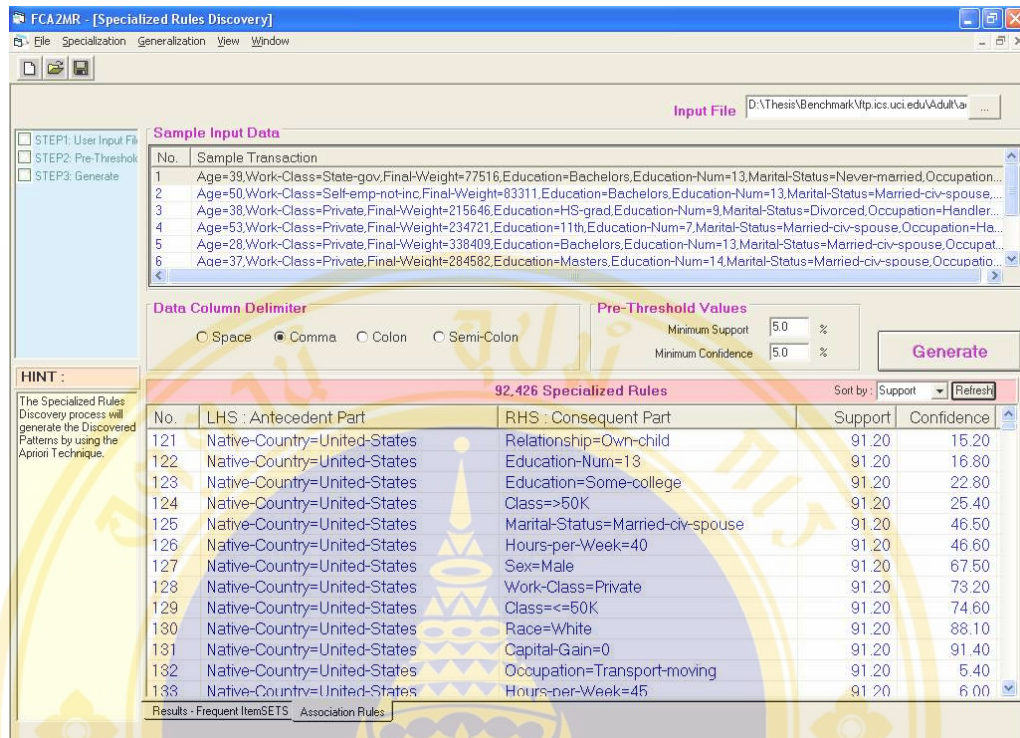


Figure 5-11: The Association Rules Results of Adult DataSET (Min_Conf=5)

5.4.2 Generalization Rules Mining Function

The generalization rules mining function consists of three sub functions, namely, FCA function, Meta-Rules Guided Mining Generation function, and Pattern Matching function. The details of each function are described below.

5.4.2.1 FCA function

Based on the formal concept analysis theory [22, 24], we defined five sub-tasks of the FCA function to produce the ontology, namely, the Object Initialization, the Attribute Initialization, the Formal Context Generation, the Formal Concept Generation, and the Formal Lattice Generation.

□ The Object Initialization

First, the object initialization task will extract the object set from the input data file. All of the objects that contain in the input file will be grouped and summarized in terms of the unique object value. Users can select some of the target objects from this set if they want to analyze them. Figure 5-12 shows the screen of this step.

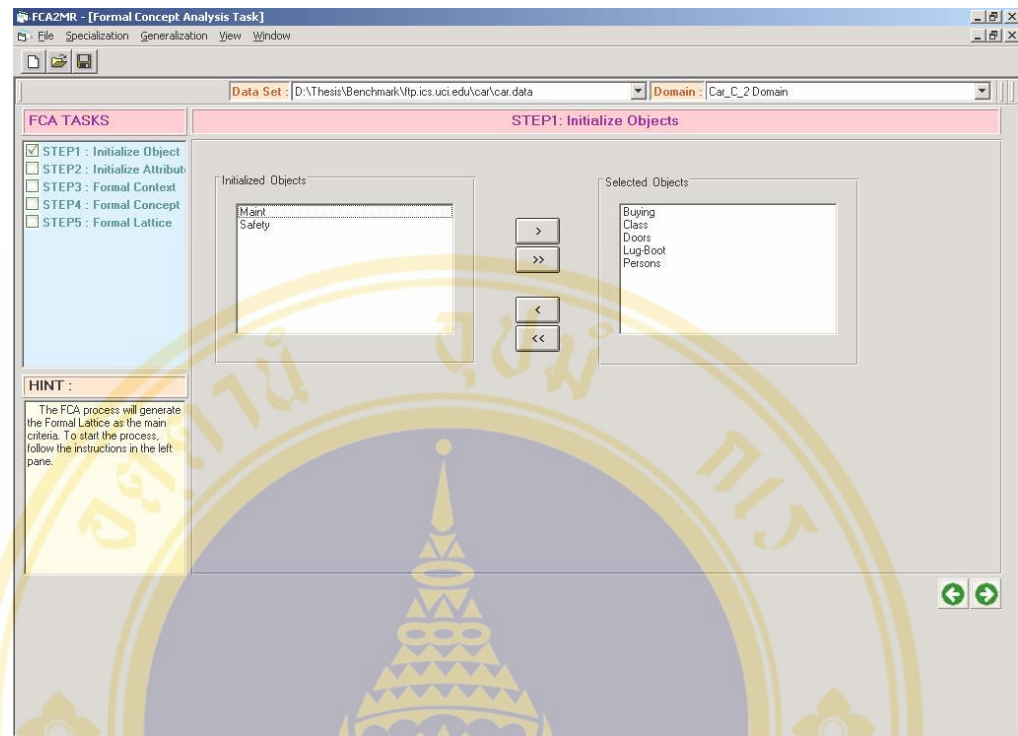


Figure 5-12: The Object Initialization Screen.

❑ The Attribute Initialization

The attribute initialization task provides an interface for the user to select the set of attributes by choosing two types of data source.

- New attributes, based on the user background knowledge against the input data file, user can type the new attribute value that he/she want to add into the system.
- Existing attributes, the attributes of this option are extracted from the input data file, as the same way of the object initialization.

Figure 5-13 displays the screen of this step.

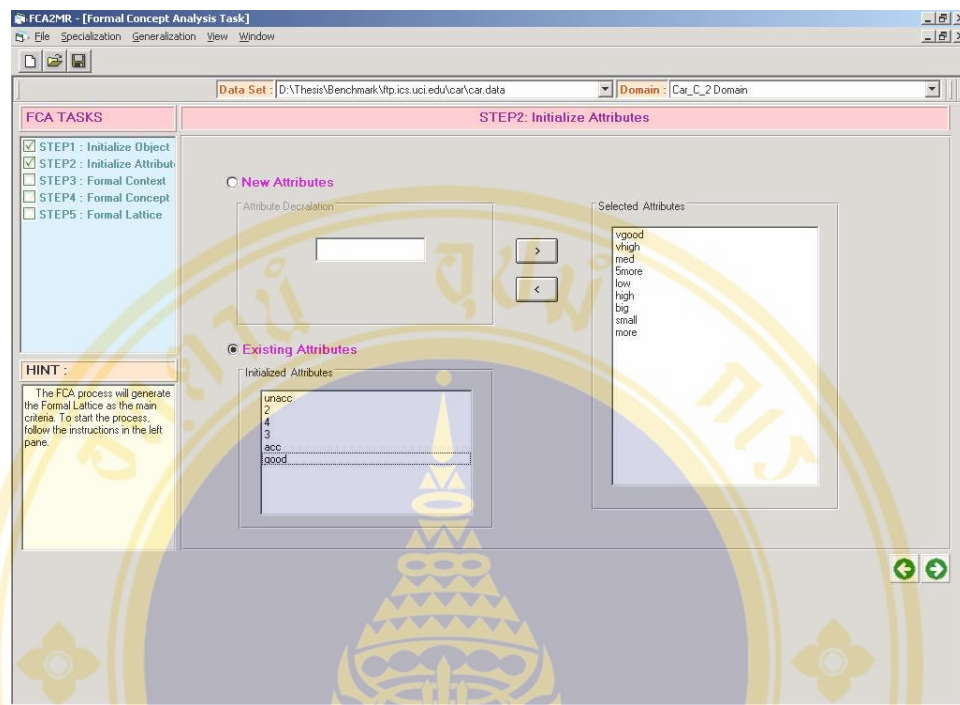


Figure 5-13: The Attribute Initialization Screen.

□ Formal Context Generation

After the user selects the objects and attributes, the system will generate the cross-table automatically. The cross-table represents the relationship between object set and attribute set. In the third step of the FCA, the formal context generation process will be performed. If the object has the binary relationship with the attribute then the formal context is marked the relationship between this object and attribute. Figure 5-14 displays the screen of the cross table of the formal context generation process and Figure 5-15 displays the generated formal context.

FCA TASKS

- ☒ STEP1: Initialize Object
- ☒ STEP2: Initialize Attribute
- ☒ STEP3: Formal Context
- ☒ STEP4: Formal Concept
- ☒ STEP5: Formal Lattice

HINT:
The FCA process will generate the Formal Lattice as the main criteria. To start the process, follow the instructions in the left pane.

STEP3: Formal Context Generation

		O B J E C T								
		vgood	vhigh	med	5more	low	high	big	small	more
A T T R I B U T E	Buying		x	x		x	x			
	Class	x								
	Doors				x					x
	Lug-Boot			x				x	x	
	Persons									x

Cross Table **Formal Context**

Figure 5-14: The Cross Table of the Formal Context Generation Screen.

FCA TASKS

- ☒ STEP1: Initialize Object
- ☒ STEP2: Initialize Attribute
- ☒ STEP3: Formal Context
- ☒ STEP4: Formal Concept
- ☒ STEP5: Formal Lattice

HINT:
The FCA process will generate the Formal Lattice as the main criteria. To start the process, follow the instructions in the left pane.

STEP3: Formal Context Generation

$K(O,A,I) = (\{Buying, Class, Doors, Lug-Boot, Persons\},$
 $\{vgood, vhigh, med, 5more, low, high, big, small, more\},$
 $\{1:2, 1:3, 1:5, 1:6, 2:1, 3:4, 3:9, 4:3, 4:7, 4:8, 5:9\})$

Cross Table **Formal Context**

Figure 5-15: The Generated Formal Context.

□ Formal Concept Generation

The next step of the FCA function is the formal concept generation. The formal concept generation task is performed by checking the group of common attributes, and common objects. This process separates the results in two ways, namely, the formal concept results and not formal concept results. Figure 5-16 demonstrates the results of the formal concept generation process.

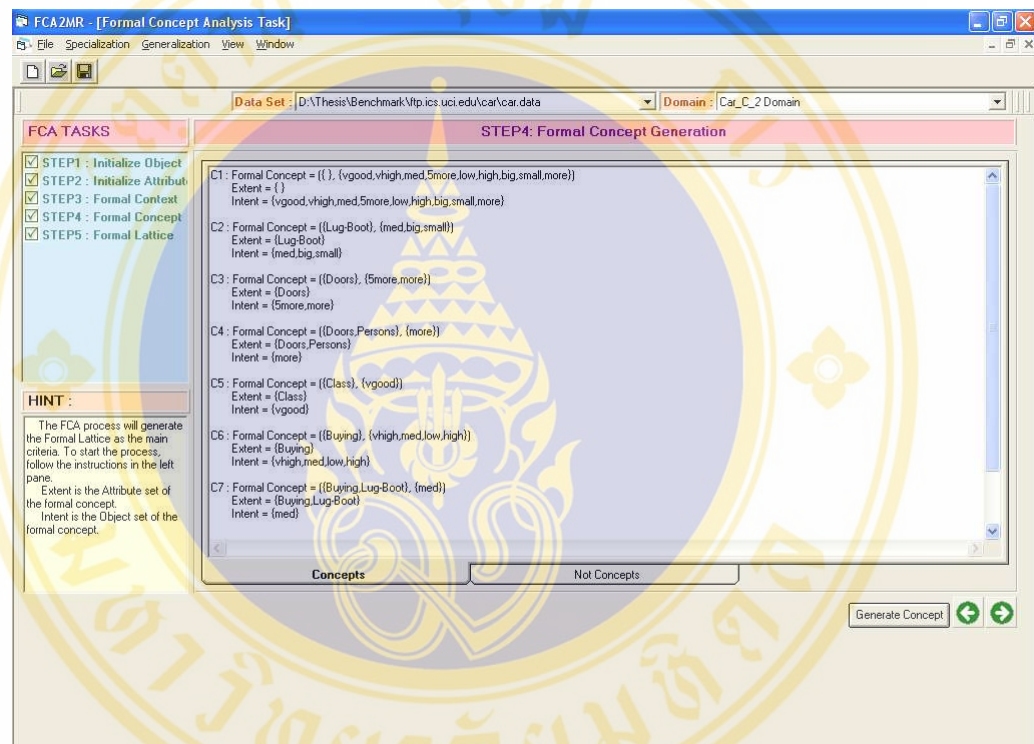


Figure 5-16: The Formal Concept Generation Screen.

□ Formal Lattice Generation

The last step of the FCA task is to generate the formal lattice of the ontology presented in the graphical view. Figure 5-17 shows the screen design of the formal lattice generation process.

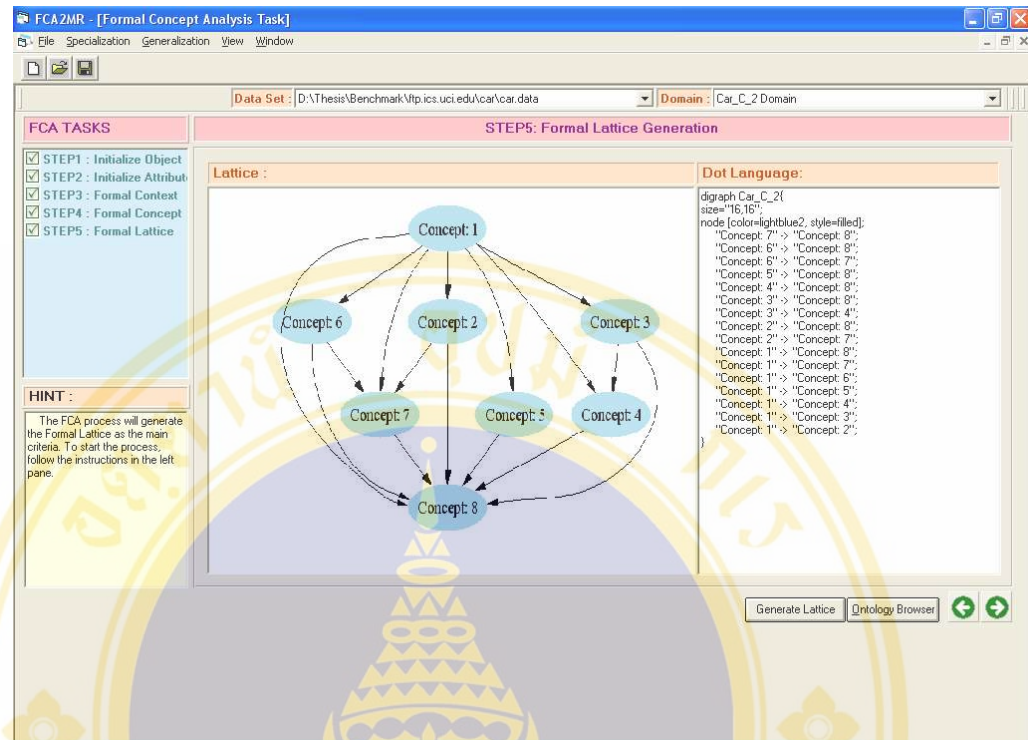


Figure 5-17: The Formal Lattice Generation Screen.

Based on the FCA theory used in the FCA task, the representation style of our ontology is separated into three views: the General View, the Extent View, and the Intent View. The details of each view are described below:

- The general view represents the formal concept of the ontology. Each node of the ontology is named by using the formal concept ID. This view presents the class hierarchical relationship of the formal concept in terms of super-class / sub-class relationship. Figure 5-18 shows the general view of the Car_C_2 ontology domain.

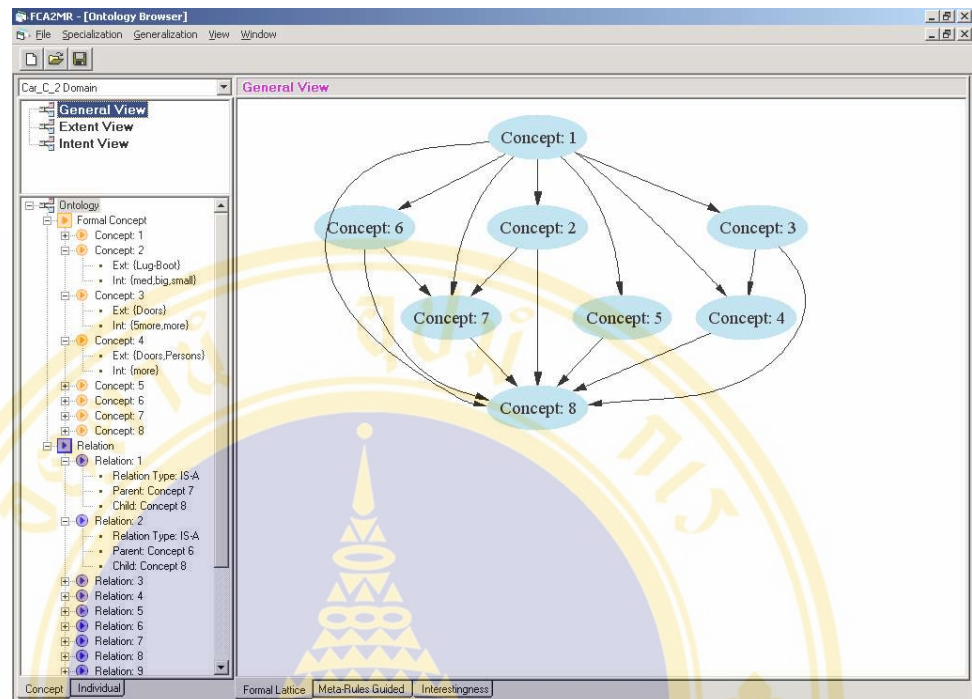


Figure 5-18: Ontology Browser: General View with Formal Concept and Relation Properties

- The extent view represents the extension part of the formal concept. The extension is represented in terms of “Objects set” in the domain of interest. Figure 5-19 shows the extent view of the Car_C_2 ontology domain.

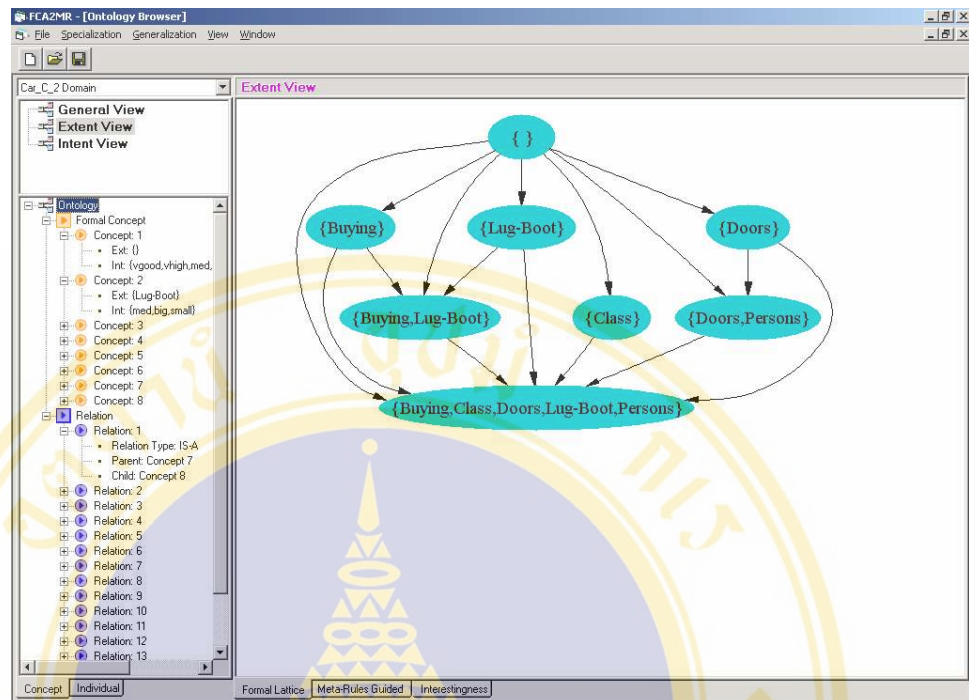


Figure 5-19: Ontology Browser: The Extent View with Formal Concept and Relation Properties.

- The Intent View represents the intension part of the formal concept. The intension is represented in terms of “Attributes” set in the domain of interest. Figure 5-20 shows the intent view of the Car_C_2 ontology domain.

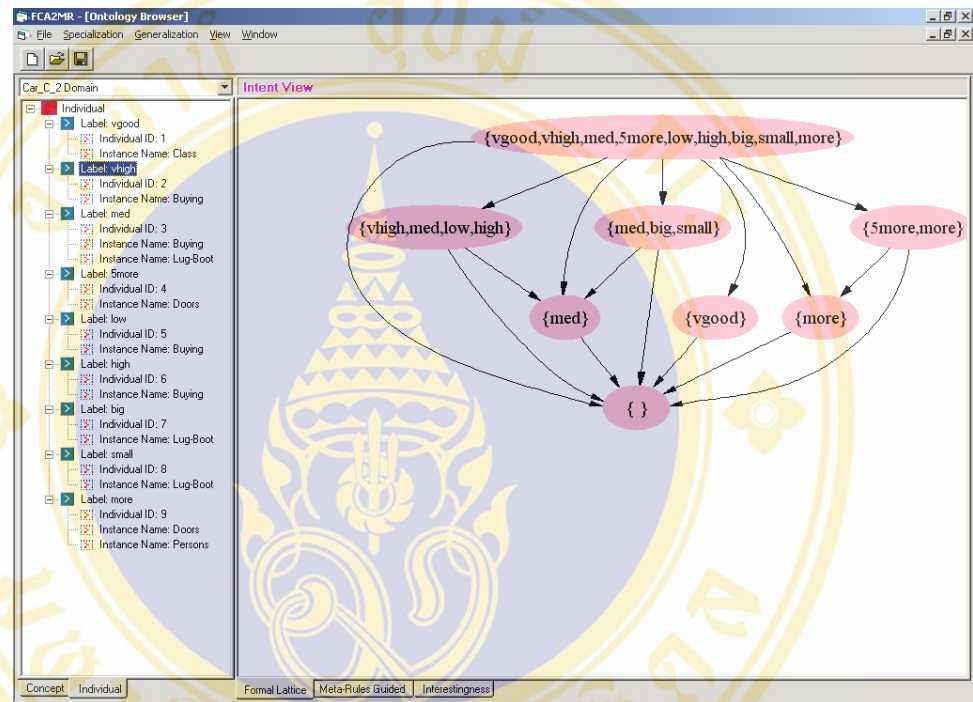
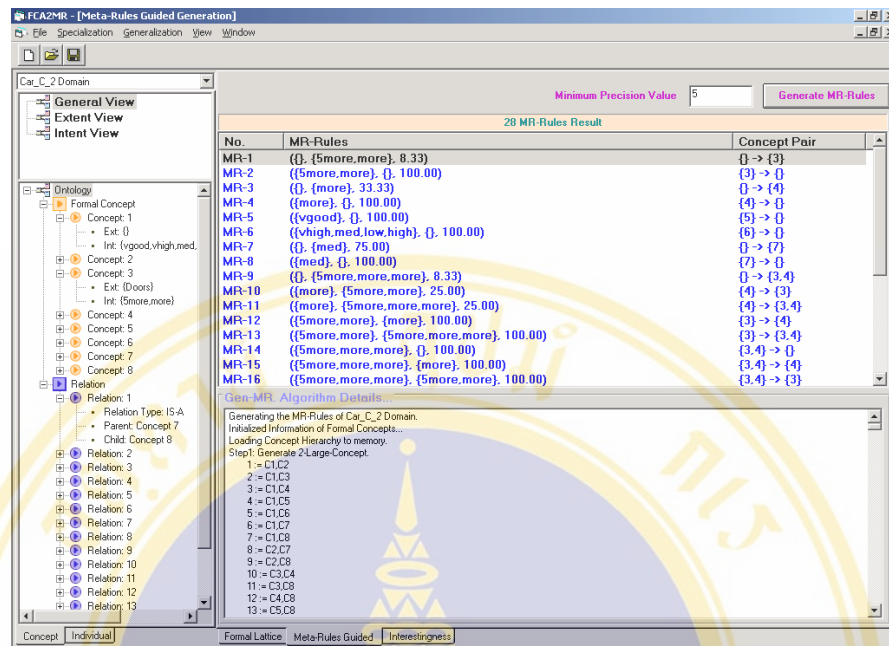


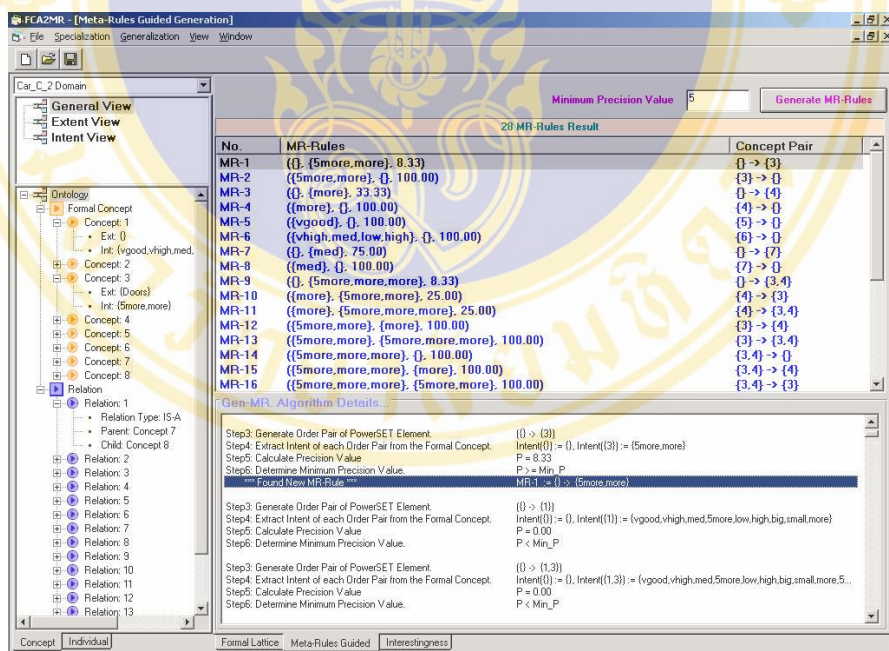
Figure 5-20: Ontology Browser: The Intent View with Individual Properties

5.4.2.2 Meta-Rules Guided Mining Function

The meta-rules guided mining function is developed to generate the meta-rules (MR) which is defined as the generalized rules level. The inference process used in this function will extract and generate the MR by using the rules implication on lattice (referred to section 4.4.1: Rules Implication on Lattice). Figure 5-21 (a) and (b) present the screenshot of this function.



(a)



(b)

Figure 5-21: The Meta-Rules Result Screen (min_P=5.00 %)

5.4.2.3 Interestingness Pattern Matching Function

This is the last function of the FCA2MR prototype system. The interestingness pattern matching function is designed to match the pattern between the Meta-Rules Patterns and Discovered Patterns to produce the Interestingness results of the system. This function matches each Discovered Patterns against each Meta-Rules Patterns in order to obtain the degree of matching which is used to evaluate the strength of the interestingness results. In addition, this function ranks the results according to their Meta-Rules patterns. Figure 5-22 illustrates the screen design of this function.

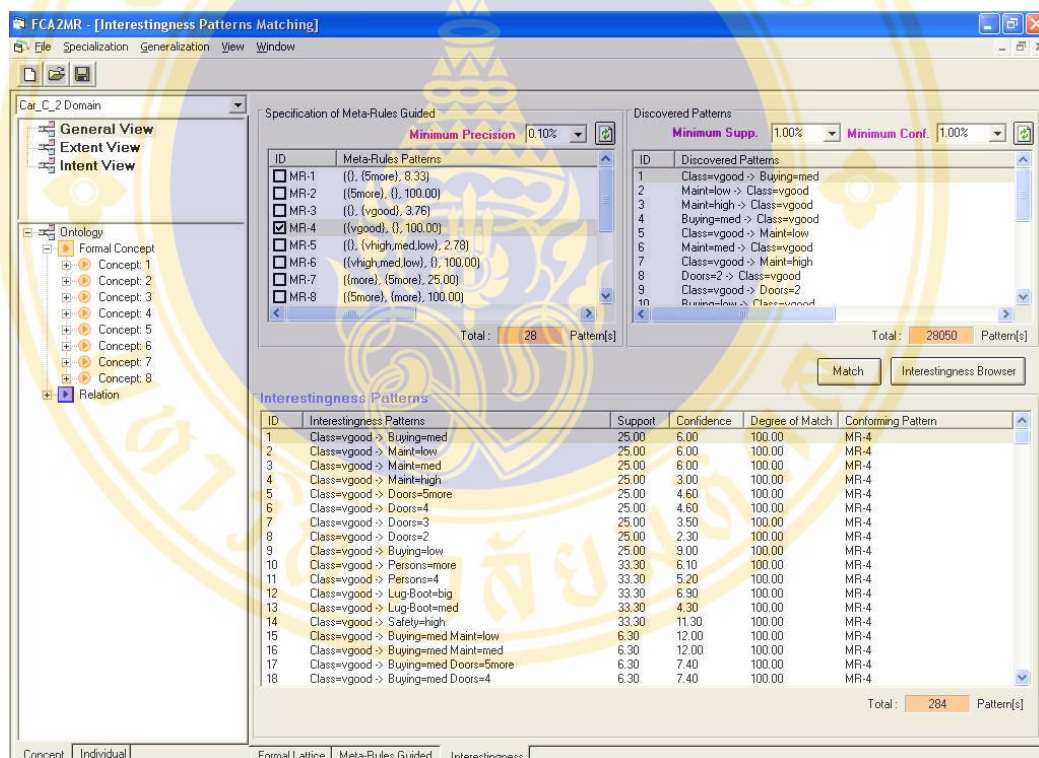


Figure 5-22: The Interestingness Patterns Result Screen

Moreover, the interestingness browser is developed to support the user to be convenient in tracking the interestingness results. User can analyze the result by specifying the sorting attribute and specifying the number of result records. The interestingness browser is displayed in Figure 5-23.

FCA2MR - [Interestingness Browser]

File Specialization Generalization View Window

Interestingness Patterns

Sort By: Degree of Match

Result Rec: All Rows

Load

ID	Ante	Cons	Support	Confidence	Degree of Match
1	Class=vgood	Safety=high	33.30	11.30	100.00
2	Class=vgood	Lug-Boot=big	33.30	6.90	100.00
3	Class=vgood	Persons=more	33.30	6.10	100.00
4	Class=vgood	Persons=4	33.30	5.20	100.00
5	Class=vgood	Lug-Boot=med	33.30	4.30	100.00
6	Class=vgood	Buying=low	25.00	9.00	100.00
7	Class=vgood	Buying=med	25.00	6.00	100.00
8	Class=vgood	Maint=low	25.00	6.00	100.00
9	Class=vgood	Maint=med	25.00	6.00	100.00
10	Class=vgood	Doors=5more	25.00	4.60	100.00
11	Class=vgood	Doors=4	25.00	4.60	100.00
12	Class=vgood	Doors=3	25.00	3.50	100.00
13	Class=vgood	Maint=high	25.00	3.00	100.00
14	Class=vgood	Doors=2	25.00	2.30	100.00
15	Class=vgood	Lug-Boot=big Safety=high	11.10	20.80	100.00
16	Class=vgood	Persons=more Safety=high	11.10	18.20	100.00
17	Class=vgood	Persons=4 Safety=high	11.10	15.60	100.00
18	Class=vgood	Lug-Boot=med Safety=high	11.10	13.00	100.00
19	Class=vgood	Persons=4 Lug-Boot=big	11.10	10.40	100.00
20	Class=vgood	Persons=more Lug-Boot=big	11.10	10.40	100.00
21	Class=vgood	Persons=more Lug-Boot=med	11.10	7.80	100.00
22	Class=vgood	Persons=4 Lug-Boot=med	11.10	5.20	100.00
23	Class=vgood	Buying=low Safety=high	8.30	27.10	100.00
24	Class=vgood	Maint=low Safety=high	8.30	18.10	100.00
25	Class=vgood	Buying=med Safety=high	8.30	18.10	100.00
26	Class=vgood	Maint=med Safety=high	8.30	18.10	100.00
27	Class=vgood	Buying=low Lug-Boot=big	8.30	16.70	100.00
28	Class=vgood	Buying=low Persons=more	8.30	14.60	100.00
29	Class=vgood	Doors=5more Safety=high	8.30	13.90	100.00
30	Class=vgood	Doors=4 Safety=high	8.30	13.90	100.00
31	Class=vgood	Buying=low Persons=4	8.30	12.50	100.00
32	Class=vgood	Maint=low Lug-Boot=big	8.30	11.10	100.00
33	Class=vgood	Maint=med Lug-Boot=big	8.30	11.10	100.00
34	Class=vgood	Buying=med Lug-Boot=big	8.30	11.10	100.00

Total: 284 Pattern[s]

Figure 5-23: The Interestingness Browser Screen

CHAPTER 89

EXPERIMENT AND RESULT

6.1 Scope of the Experiment

The experiment is conducted to verify our proposed idea of the novel approach of a unified data mining framework, namely, the Meta-Rules Guided Mining based on Ontology. The proposed framework is designed to discover the subjective interestingness patterns, in particular, actionable criteria.

The chapter begins with the experimental framework, the experimental setting, and the experimental results. Finally, the experimental evaluations are shown in the last section.

6.2 Experimental Framework

In our experiment, the main process deals with the development of prototype functions that conform to the proposed framework in which we described in chapter 4 and designed in chapter 5. The experimental framework is illustrated in Figure 6-1.

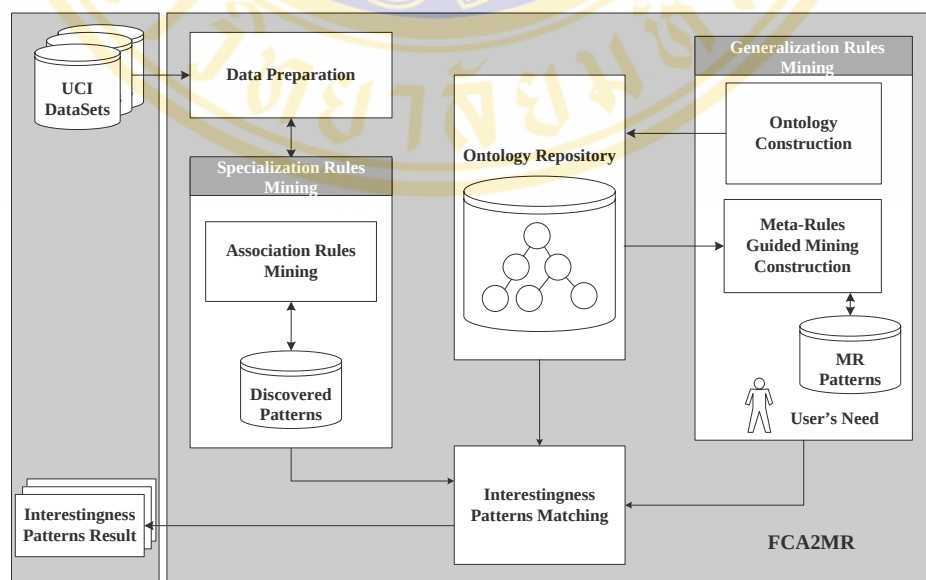


Figure 6-1: An Experimental Framework

From Figure 6-1, the experimental framework starts with the data preparation module. The specialization rule mining is a part which contains the association rules mining module for generating a set of discovered patterns of the particular domain of interest. While the generalization rules mining constructs the ontology via the ontology construction module. The ontology results are stored in the ontology repository and use to generate the Meta-Rules patterns by performing the meta-rules guided mining construction module. Finally, the interestingness pattern matching is the last module to match the interestingness patterns result.



6.2.1 Experimental Setting

6.2.1.1 DataSet

The dataset we used in the experiments are not actually focused on the large data size since the evaluation objective is to analyze and gain the qualitative result pertaining to the subjective interestingness of the number of discovered patterns.

In the experiments, we used three datasets from the UCI Machine Learning repository [2]. Most of the datasets are frequently used in the literature [14, 29, 31, 46]. The dataset and their details are listed in Table 6-1.

Table 6-1: The Details of Datasets

DataSet	Attributes	Tuples	Missing Value
Car	7	1,728	None
An Example of Car Dataset Buying=vhigh,Maint=vhigh,Doors=2,Persons=2,Lug-Boot=small, Safety=low,Class=unacc			
Nursery	8	12,960	None
An Example of Nursery Dataset Parents=usual,Has_Nurs=proper,Form=complete,Children=1,Housing=convenient, Finance=convenient,Social=nonprob,Health=recommended,Class=recommend			
Adult	14	30,162	None
An Example of Adult Dataset Age=39,Work-Class=State-gov,Final-Weight=77516,Education=Bachelors, Education-Num=13,Marital-Status=Never-married,Occupation=Adm-clerical, Relationship=Not-in-family,Race=White,Sex=Male,Capital-Gain=2174, Capital- Loss=0,Hours-per-Week=40,Native-Country=United-States, Class=<=50K			

6.2.1.2 Data Model Setting

For each dataset, we separate into three data models. The details are described in Table 6-2.

Table 6-2: The Details of Data Model

DataSet	Data Model	Specified Domain
Car	C-1	Car_C_1
	An Example of C_1 data model Object Member = 5 Object Set = {Buying,Lug-Boot,Persons,Safety,Maint} Attribute Member = 5 Attribute Set = {High,Big,Med,More,Good}	
	C-2	Car_C_2
	An Example of C_2 data model Object Member = 5 Object Set = {Buying,Class,Doors,Lug-Boot,Persons } Attribute Member = 9 Attribute Set = {VGood,VHigh,Med,5More,Low,High,Big,Small,More}	
	C-3	Car_C_3
	An Example of C_3 data model Object Member = 7 Object Set = {Buying,Doors,Lug-Boot,Maint,Persons,Safety,Class Attribute Member = 10 Attribute Set = {High,VHigh,Med,5More,Big,Small,Low,More,2,4}	
Nursery	N-1	Nursery_N_1
	An Example of N_1 data model Object Member = 5 Object Set = {Form,Parents,Social,Housing,Health} Attribute Member= 8 Attribute Set = {Convenient,Priority,Recommended,Problematic,Usual, Incomplete,Very_Recom,Complete}	
	N-2	Nursery_N_2
	An Example of N_2 data model Object Member = 7 Object Set = {Children,Class,Finance,Health,Housing,Social,Parents} Attribute Member = 10 Attribute Set = {Convenient,Priority,Critical,Pretentious,Problematic,Usual, Complete,Foster,Not_Recom,Recommended}	
	N-3	Nursery_N_3
	An Example of N_3 data model Object Member = 7 Object Set = {Children,Health,Finance,Housing,Parents,Social,Form} Attribute Member = 15 Attribute Set = {1,More,3,Recommend,Not_Recom,Convenient,Inconv, Critical,Less_Conv,Pretentious,Problematic,Usual,NonProb, Completed,Incomplete}	

Table 6-2: The Details of Data Model (Cont.)

DataSet	Data Model	Specified Domain
Adult	A-1	Adult_A_1
	An Example of A_1 data model Object Member = 3 Object Set = {Sex, Education, Age} Attribute Member = 3 Attribute Set = {Bachelors, Male, 17}	
	A-2	Adult_A_2
	An Example of A_2 data model Object Member = 5 Object Set = {Marital-Status, Relationship, Education, Native-Country, Sex} Attribute Member = 8 Attribute Set = {United-States, Never-Married, SomeCollege, Male, Husband, Female, Masters, Bachelors}	
	A-3	Adult_A_3
	An Example of A_3 data model Object Member = 7 Object Set = {Education, Occupation, Relationship, Hours-Per-Week, Age, MaritalStatus, Native-Country} Attribute Member = 9 Attribute Set = {Bachelors, Masters, Sales, Other-Service, Husband, Wife, 25, Never-Married, Divorced, United-States}	

Each data model is determined to specify a domain of existing knowledge for each dataset, including objects and attributes. These two elements are used to construct the ontology by using the FCA technique, and to view the ranking results in the last module of our system, the interestingness pattern matching module.

6.2.1.3 Threshold Setting

The threshold values used in this experiment are separated into two values, namely, pre-threshold value, and post-threshold value.

▪ Pre-Threshold Value

The pre-threshold value is used to measure the association rule-mining task in the first step of the framework. The pre-threshold value is divided into two values:

- ❑ Minimum Support Value (Min_Supp): the range values cover 1, 5, and 10.

- Minimum Confidence Value (Min_Conf): the range values cover 1, 5, and 10.

- **Post-Threshold Value**

The system generates the MR patterns by determining the post-threshold value in the meta-rules guided mining construction module, we examine the post-threshold value as follows:

- Minimum Precision Value (Min_P): the range values cover 0.1, 0.5, 1, 5, and 10.

6.2.2 Experimental Methods

The experimental methods of the research focus on two stages: the experimental testing stage, and the experimental evaluation stage.

6.2.2.1 **The Experimental Testing** consists of three tests. The details are described below:

1. **The Data Mining Testing**, to test the result of the specialized rules discovery part. The system design adopts the *Apriori* algorithm to perform on this task. Moreover, we run the test case on the UCI benchmark datasets [2]. The results are shown in section 6.3.1.
2. **The Ontology Generator Testing**, we test the ontology generator by generating nine ontology domains against the benchmark dataset extracted from the UCI Benchmark data. The screen results of the ontology are shown in section 6.3.2
3. **The Interestingness Testing**, we test the interestingness results generated from our prototype system. The details of the testing are summarized and the comparisons of the interestingness/discarded results of each data domain are plotted. The results are shown in section 6.3.3.

6.2.2.2 **FCA2MR Functional Evaluation** consists of three evaluations. The details are described below:

1. The Functional Comparison of Three Ontology Generators,

we evaluate the ontology generated from the FCA2MR by comparing with two other tools, namely, Protégé2000, and ConImp. The results of the comparison are shown in section 6.4.1

2. Interestingness Evaluation, we evaluate the efficiency of our interestingness function by comparing the results with the Magnum Opus Demo program [45, 46, 47]. The evaluation aims to validate the correctness and consistency of interestingness patterns, as well as to measure the degree of variation of a set of rules yielded from our prototype and benchmark tool.

As the data size allowed for the test with Magnum Opus is limited up to 1000 tuples, we thus use the same 1000 tuples of datasets for the interestingness evaluation. The comparisons of interestingness results are shown in section 6.4.2.

3. User Participation Evaluation, we evaluate the interestingness results against the user participation by preparing the questionnaire to get the users interestingness. The questionnaires are composed of three active datasets prepared from Mahidol University, namely, library dataset, ICT student dataset, and short course training dataset. Each questionnaire contains two major themes, and each theme contains ten cases of problem. Moreover, each questionnaire is collected from 100 peoples who are relevant to the data domain. The summarized results of each questionnaire are presented in section 6.4.3

6.3 Experimental Results

The experimental results are summarized and gained from the test and evaluation of the functional system of the FCA2MR prototype. The results show that the system can support all the functions of the proposed framework.

All experimental results described below were generated through the process of the FCA2MR system. We describe the detail of the experimental results in three ways, namely, the data mining testing results, the ontology generator testing results and the interestingness testing results.

6.3.1 The Data Mining Testing Results

The three datasets of UCI machine learning repository [2] are used to generate the discovered patterns of the specialization module. Table 6-3 shows the results of the data mining testing.

Table 6-3: Summary Results of Three Datasets in Specialization Module

DataSets	Min_Supp	Min_Conf	Number of Discovered ItemSets	Number of Discovered Patterns
Car	1.00	1.00	2,291	28,050
	1.00	5.00	2,291	27,986
	1.00	10.00	2,291	27,669
	5.00	1.00	349	5,649
	5.00	5.00	349	5,598
	5.00	10.00	349	5,499
	10.00	1.00	86	1,569
	10.00	5.00	86	1,552
	10.00	10.00	86	1,528
Nursery	1.00	1.00	8,094	134,019
	1.00	5.00	8,094	134,723
	1.00	10.00	8,094	132,609
	5.00	1.00	623	13,969
	5.00	5.00	623	13,806
	5.00	10.00	623	13,611
	10.00	1.00	169	4,126
	10.00	5.00	169	4,072
	10.00	10.00	169	4,029
Adult	1.00	1.00	41,741	1,623,767
	1.00	5.00	41,741	558,770
	1.00	10.00	41,741	379,587
	5.00	1.00	5,850	292,963
	5.00	5.00	5,850	92,426
	5.00	10.00	5,850	61,792
	10.00	1.00	2,249	133,523
	10.00	5.00	2,249	42,256
	10.00	10.00	2,249	28,361

6.3.2 The Ontology Generator Testing Results

To test the efficiency of the ontology generator, we run the process of the FCA task by selecting some target objects and attributes, and analyzing the shape of the lattice in terms of the ontology depth and the ontology concept. The details are described in Table 6-4.

Table 6-4: The Summary Results of Ontology Generation Process of each Data Model

Data Model	Domain	Selected Obj.	Selected Attr.	Ontology Depth	No. of Concept
C-1	Car_C_1	3	3	4	7
C-2	Car_C_2	5	8	4	8
C-3	Car_C_3	7	10	4	8
N-1	Nursery_N_1	5	8	3	7
N-2	Nursery_N_2	7	10	4	7
N-3	Nursery_N_3	10	15	4	7
A-1	Adult_A_1	3	3	3	5
A-2	Adult_A_2	5	8	3	7
A-3	Adult_A_3	7	10	3	7

The results of each data model reveal that the ontology generator task of the FCA2MR prototype can perform and generate the lattice completely. The ontology results screen of each data domain are shown as follows:

Case 1: Car_C_1 Domain

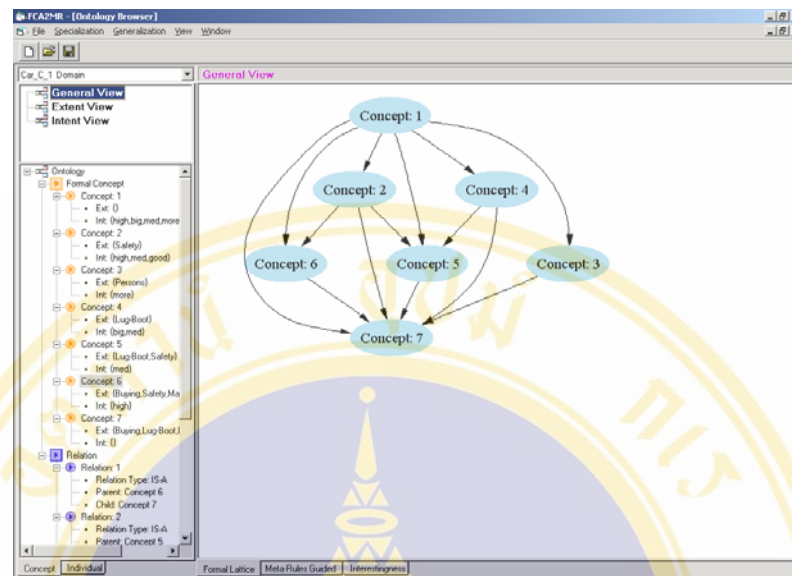


Figure 6-2: Ontology of the Car_C_1 domain

Figure 6-2 presents the generated ontology of Car_C_1 domain which consists of seven concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Buying, Lug-Boot, Persons, Safety, Maint},
{high, big, med, more, good},
{5:1,1:1,2:2,2:3,3:4,4:5,4:1,4:3})

○ Formal Concept (FC)

- C1 Formal Concept = ({ }, {high, big, med, more, good})
Extent = { }
Intent = {high, big, med, more, good}
- C2 Formal Concept = ({Safety}, {high, med, good})
Extent = {Safety}
Intent = {high, med, good}
- C3 Formal Concept = ({Persons}, {more})
Extent = {Persons}
Intent = {more}
- C4 Formal Concept = ({Lug-Boot}, {big, med})
Extent = {Lug-Boot}
Intent = {big, med}
- C5 Formal Concept = ({Lug-Boot, Safety}, {med})
Extent = {Lug-Boot, Safety}
Intent = {med}
- C6 Formal Concept = ({Buying, Safety, Maint}, {high})
Extent = {Buying, Safety, Maint}
Intent = {high}
- C7 Formal Concept = ({Buying, Lug-Boot, Persons, Safety, Maint}, { })
Extent = {Buying, Lug-Boot, Persons, Safety, Maint}
Intent = { }

Case 2: Car_C_2 Domain

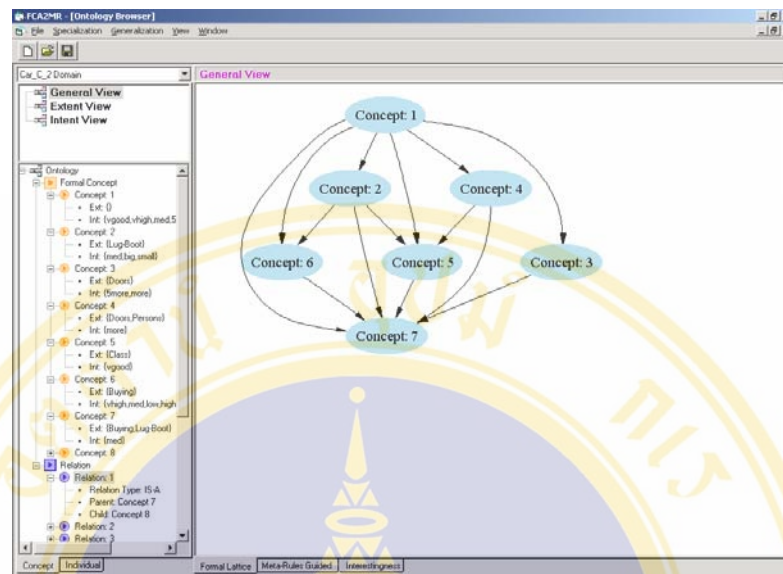


Figure 6-3: Ontology of the Car_C_2 domain

Figure 6-3 presents the generated ontology of Car_C_2 domain which consists of eight concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Buying, Class, Doors, Lug-Boot, Persons},
{vgood, vhigh, med, 5more, low, high, big, small, more},
{1:2,1:3,1:5,1:6,2:1,3:4,3:9,4:3,4:7,4:8,5:9})

○ Formal Concept (FC)

- C1 Formal Concept = ({ }, {vgood, vhigh, med, 5more, low, high, big, small, more})
Extent = { }
Intent = {vgood, vhigh, med, 5more, low, high, big, small, more}
- C2 Formal Concept = ({Lug-Boot}, {med, big, small})
Extent = {Lug-Boot}
Intent = {med, big, small}
- C3 Formal Concept = ({Doors}, {5more,more})
Extent = {Doors}
Intent = {5more,more}
- C4 Formal Concept = ({Doors, Persons}, {more})
Extent = {Doors, Persons}
Intent = {more}
- C5 Formal Concept = ({Class}, {vgood})
Extent = {Class}
Intent = {vgood}
- C6 Formal Concept = ({Buying}, {vhigh, med, low, high})
Extent = {Buying}
Intent = {vhigh, med, low, high}
- C7 Formal Concept = ({Buying, Lug-Boot}, {med})
Extent = {Buying, Lug-Boot}
Intent = {med}
- C8 Formal Concept = ({Buying, Class, Doors, Lug-Boot, Persons}, { })
Extent = {Buying, Class, Doors, Lug-Boot, Persons}
Intent = { }

Case 3: Car_C_3 Domain

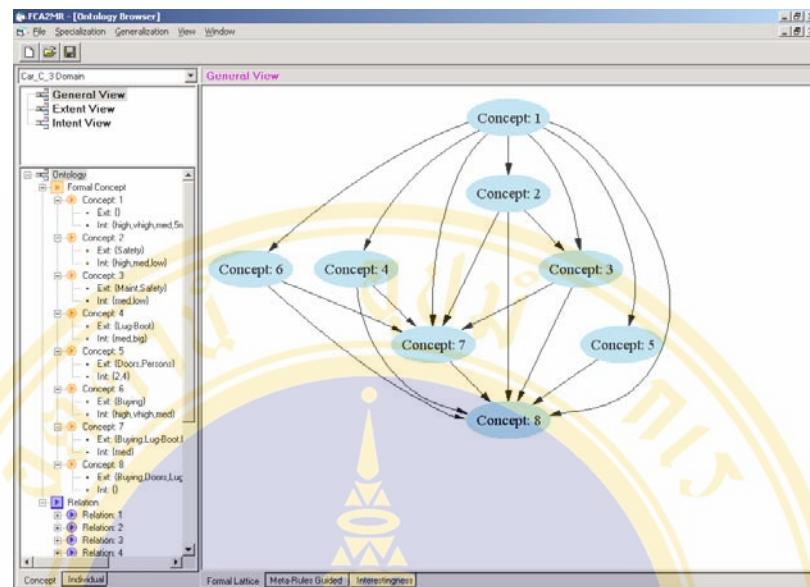


Figure 6-4: Ontology of the Car_C_3 domain

Figure 6-4 presents the generated ontology of Car_C_3 domain which consists of eight concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

{Buying, Doors, Lug-Boot, Maint, Persons, Safety, Class},
 {high, vhigh, med, 5more, big, small, low, more, 2, 4},
 {1:1,1:2,1:3,2:4,2:9,2:10,3:5,3:3,4:3,4:7,5:10,5:9,5:8,6:1,6:3,6:7})

○ Formal Concept (FC)

- C1 Formal Concept = ({ }, {high,vhigh,med,5more,big,small,low,more,2,4})
 Extent = { }
 Intent = {high,vhigh,med,5more,big,small,low,more,2,4}
- C2 Formal Concept = ({Safety}, {high, med, low})
 Extent = {Safety}
 Intent = {high, med, low}
- C3 Formal Concept = ({Maint, Safety}, {med, low})
 Extent = {Maint, Safety}
 Intent = {med, low}
- C4 Formal Concept = ({Lug-Boot}, {med, big})
 Extent = {Lug-Boot}
 Intent = {med, big}
- C5 Formal Concept = ({Doors, Persons}, {2,4})
 Extent = {Doors, Persons}
 Intent = {2,4}
- C6 Formal Concept = ({Buying}, {high, vhigh, med})
 Extent = {Buying}
 Intent = {high, vhigh, med}
- C7 Formal Concept = ({Buying, Lug-Boot, Maint, Safety}, {med})
 Extent = {Buying, Lug-Boot, Maint, Safety}
 Intent = {med}
- C8 Formal Concept = ({Buying, Doors, Lug-Boot, Maint, Persons, Safety, Class}, { })
 Extent = {Buying, Doors, Lug-Boot, Maint, Persons, Safety, Class}
 Intent = { }

Case 4: Nursery_N_1 Domain

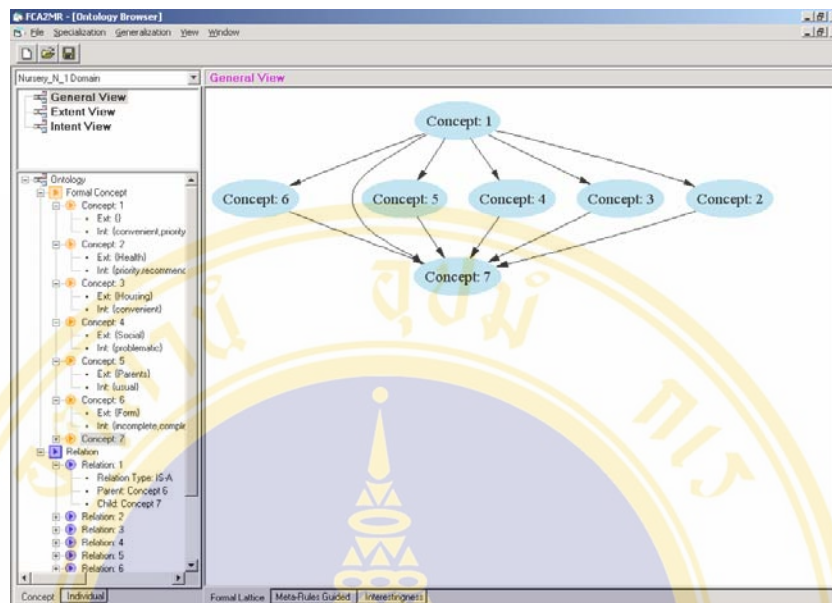


Figure 6-5: Ontology of the Nursery_N_1 domain

Figure 6-5 presents the generated ontology of Nursery_N_1 domain which consists of seven concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Form, Parents, Social, Housing, Health},
 {convenient,priority,recommended,problematic,usual,incomplete,very_recom, complete},
 {1:6,1:8,2:5,5:2,5:3,4:1,3:4})

○ Formal Concept (FC)

- C1 Formal Concept = ({ },
 {convenient,priority,recommended,problematic,usual,incomplete,very_recom,complete})
 Extent = { }
 Intent =
 {convenient,priority,recommended,problematic,usual,incomplete,very_recom,complete}
- C2 Formal Concept = ({Health}, {priority, recommended})
 Extent = {Health}
 Intent = {priority, recommended}
- C3 Formal Concept = ({Housing}, {convenient})
 Extent = {Housing}
 Intent = {convenient}
- C4 Formal Concept = ({Social}, {problematic})
 Extent = {Social}
 Intent = {problematic}
- C5 Formal Concept = ({Parents}, {usual})
 Extent = {Parents}
 Intent = {usual}
- C6 Formal Concept = ({Form}, {incomplete, complete})
 Extent = {Form}
 Intent = {incomplete, complete}
- C7 Formal Concept = ({Form,Parents,Social,Housing,Health}, { })
 Extent = {Form,Parents,Social,Housing,Health}
 Intent = { }

Case 5: Nursery_N_2 Domain

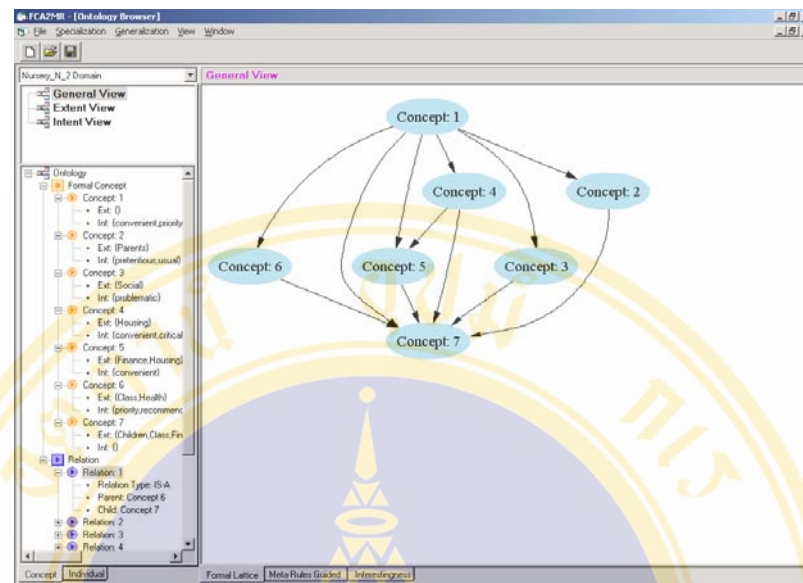


Figure 6-6: Ontology of the Nursery_N_2 domain

Figure 6-6 presents the generated ontology of Nursery_N_2 domain which consists of seven concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Children,Class,Finance,Health,Housing,Social,Parents},
 {convenient,priority,critical,pretentious,problematic,usual,complete,foster,
 not_recom, recommended},
 {5:1,4:2,4:10,7:6,6:5,2:9,2:10,2:2,3:1,5:3,7:4})

○ Formal Concept (FC)

C1 Formal Concept = ({ }, {convenient,priority,critical,pretentious,problematic,
 usual,complete,foster,not_recom, recommended})

Extent = { }

Intent = {convenient,priority,critical,pretentious,problematic,
 usual,complete,foster,not_recom, recommended}

C2 Formal Concept = ({Parents}, {pretentious, usual})

Extent = {Parents}

Intent = {pretentious, usual}

C3 Formal Concept = ({Social}, {problematic})

Extent = {Social}

Intent = {problematic}

C4 Formal Concept = ({Housing}, {convenient, critical})

Extent = {Housing}

Intent = {convenient, critical}

C5 Formal Concept = ({Finance, Housing}, {convenient})

Extent = {Finance, Housing}

Intent = {convenient}

C6 Formal Concept = ({Class, Health}, {priority, recommended})

Extent = {Class, Health}

Intent = {priority, recommended}

C7 Formal Concept = ({Children,Class,Finance,Health,Housing,Social,Parents}, { })

Extent = {Children,Class,Finance,Health,Housing,Social,Parents}

Intent = { }

Case 6: Nursery_N_3 Domain

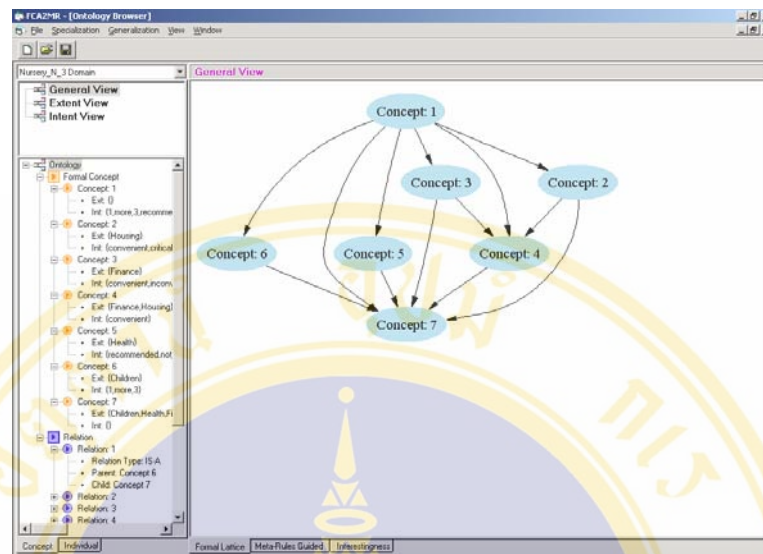


Figure 6-7: Ontology of the Nursery_N_3 domain

Figure 6-7 presents the generated ontology of Nursery_N_3 domain which consists of seven concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Children,Health,Finance,Housing,Parents,Social,Form},
 {1,more,3,recommended,not_recom,convenient,inconv,critical,
 less_conv,pretentious,problematic,nonprob,completed,incomplete},
 {1:1,1:2,1:3,2:4,2:5,3:6,3:7,4:6,4:8,4:9,5:10,6:11,6:12,7:13,7:14})

○ Formal Concept (FC)

- C1 Formal Concept = ({ }, {1,more,3,recommended,not_recom,convenient,
 inconv,critical, less_conv,pretentious,problematic,nonprob,
 completed,incomplete})
 Extent = { }
 Intent = {1,more,3,recommended,not_recom,convenient,inconv,critical,
 less_conv,pretentious,problematic,nonprob,completed,incomplete}
- C2 Formal Concept = ({Housing}, {convenient, critical, less_conv})
 Extent = {Housing}
 Intent = {convenient, critical, less_conv}
- C3 Formal Concept = ({Finance}, {convenient,inconv})
 Extent = {Finance}
 Intent = {convenient,inconv}
- C4 Formal Concept = ({Finance,Housing}, {convenient})
 Extent = {Finance,Housing}
 Intent = {convenient}
- C5 Formal Concept = ({Health}, {recommended,not_recom})
 Extent = {Health}
 Intent = {recommended,not_recom}
- C6 Formal Concept = ({Children}, {1,more,3})
 Extent = {Children}
 Intent = {1,more,3}
- C7 Formal Concept = ({Children,Health,Finance,Housing,Parents,Social,Form}, { })
 Extent = {Children,Health,Finance,Housing,Parents,Social,Form}
 Intent = { }

Case 7: Adult_A_1 Domain

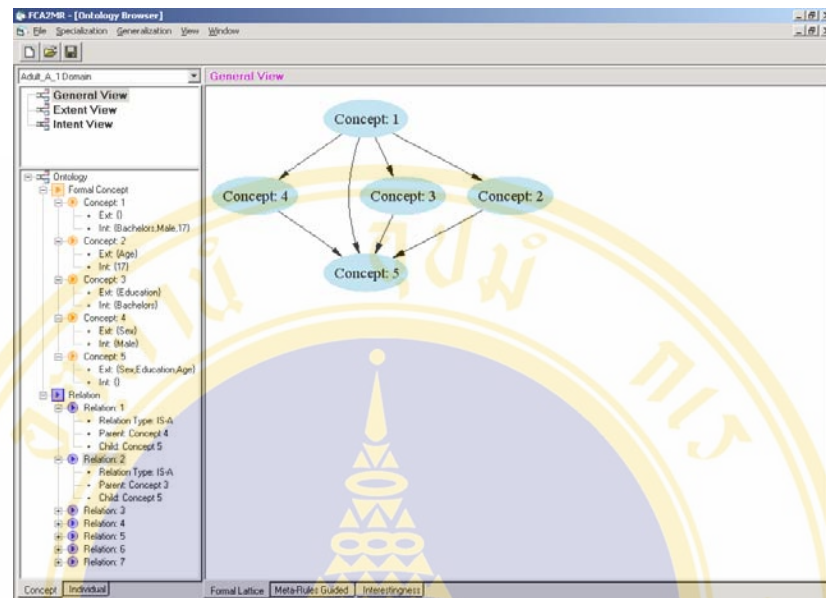


Figure 6-8: Ontology of the Adult_A_1 domain

Figure 6-7 presents the generated ontology of Adult_A_1 domain which consists of five concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

$(\{Sex, Education, Age\},$
 $\{Bachelors, Male, 17\},$
 $\{3:3, 2:1, 1:2\})$

○ Formal Concept (FC)

- C1 Formal Concept = $(\{ \}, \{Bachelors, Male, 17\})$
 Extent = $\{ \}$
 Intent = $\{Bachelors, Male, 17\}$
- C2 Formal Concept = $(\{Age\}, \{17\})$
 Extent = $\{Age\}$
 Intent = $\{17\}$
- C3 Formal Concept = $(\{Education\}, \{Bachelors\})$
 Extent = $\{Education\}$
 Intent = $\{Bachelors\}$
- C4 Formal Concept = $(\{Sex\}, \{Male\})$
 Extent = $\{Sex\}$
 Intent = $\{Male\}$
- C5 Formal Concept = $(\{Sex, Education, Age\}, \{ \})$
 Extent = $\{Sex, Education, Age\}$
 Intent = $\{ \}$

Case 8: Adult_A_2 Domain

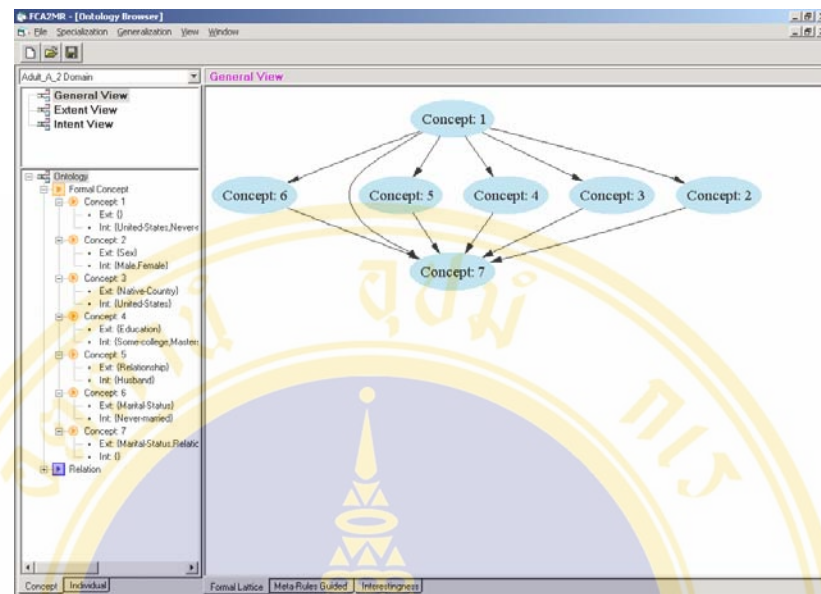


Figure 6-9: Ontology of the Adult_A_2 domain

Figure 6-9 presents the generated ontology of Adult_A_2 domain which consists of seven concepts, the details are described below:

■ Ontology Details

○ Formal Context (FK)

({Marital-Status, Relationship, Education, Native-Country, Sex},
 {United-States, Never-married, SomeCollege, Male, Husband, Female, Masters,
 Bachelors},
 {1:2, 2:5, 3:3, 5:4, 5:6, 3:7, 3:8, 4:1})

○ Formal Concept (FC)

- C1 Formal Concept = ($\{ \}$, {United-States, Never-married, Some-college, Male, Husband, Female, Masters, Bachelors})
 Extent = $\{ \}$
 Intent = {United-States, Never-married, Some-college, Male, Husband, Female, Masters, Bachelors}
- C2 Formal Concept = ({Sex}, {Male, Female})
 Extent = {Sex}
 Intent = {Male, Female}
- C3 Formal Concept = ({Native-Country}, {United-States})
 Extent = {Native-Country}
 Intent = {United-States}
- C4 Formal Concept = ({Education}, {Some-college, Masters, Bachelors})
 Extent = {Education}
 Intent = {Some-college, Masters, Bachelors}
- C5 Formal Concept = ({Relationship}, {Husband})
 Extent = {Relationship}
 Intent = {Husband}
- C6 Formal Concept = ({Marital-Status}, {Never-married})
 Extent = {Marital-Status}
 Intent = {Never-married}
- C7 Formal Concept = ({Marital-Status, Relationship, Education, Native-Country, Sex}, $\{ \}$)
 Extent = {Marital-Status, Relationship, Education, Native-Country, Sex}
 Intent = $\{ \}$

Case 9: Adult_A_3 Domain

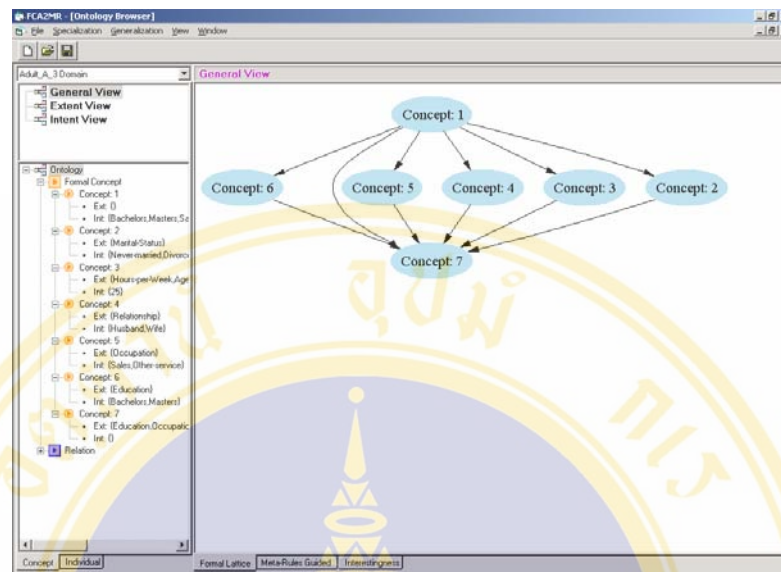


Figure 6-10: Ontology of the Adult_A_3 domain

Figure 6-10 presents the generated ontology of Adult_A_3 domain which consists of seven concepts, the details are described below:

- **Ontology Details**

- **Formal Context (FK)**

{(Education, Occupation, Relationship, Hours-per-Week, Age, Marital-Status, Native- Country),
 {Bachelors, Masters, Sales, Other-service, Husband, Wife, 25, Never-married, Divorced, United-States}, {1:1, 1:2, 2:3, 2:4, 3:5, 3:6, 4:7, 5:7, 6:8, 6:9}}

- **Formal Concept (FC)**

- C1 Formal Concept = ({ }, {Bachelors, Masters, Sales, Other-service, Husband, Wife, 25, Never-married, Divorced, United-States})
 Extent = { }
 Intent = {Bachelors, Masters, Sales, Other-service, Husband, Wife, 25, Never-married, Divorced, United-States}
- C2 Formal Concept = ({Marital-Status}, {Never-married, Divorced})
 Extent = {Marital-Status}
 Intent = {Never-married, Divorced}
- C3 Formal Concept = ({Hours-per-Week, Age}, {25})
 Extent = {Hours-per-Week, Age}
 Intent = {25}
- C4 Formal Concept = ({Relationship}, {Husband, Wife})
 Extent = {Relationship}
 Intent = {Husband, Wife}
- C5 Formal Concept = ({Occupation}, {Sales, Other-service})
 Extent = {Occupation}
 Intent = {Sales, Other-service}
- C6 Formal Concept = ({Education}, {Bachelors, Masters})
 Extent = {Education}
 Intent = {Bachelors, Masters}
- C7 Formal Concept = ({Education, Occupation, Relationship, Hours-per-Week, Age, Marital-Status, Native-Country}, { })
 Extent = {Education, Occupation, Relationship, Hours-per-Week, Age, Marital-Status, Native-Country}
 Intent = { }

6.3.3 The Interestingness Testing Results

After we run the ontology generator testing, we then generate the meta-rules guided mining patterns (MR Patterns) and test the interestingness of each data model. The nine data models are performed and the post threshold value is used to purify the result of the interestingness.

In practice, the results of the interestingness testing are performed step by step from the ontology construction module through the interestingness patterns matching module (defined in section 4.3 to 4.5.)

Table 6-5 summarizes the results of each data model in the interestingness testing.

Table 6-5: The Summary Results of the Interestingness Testing

Domain	Min_P	Total of MR Patterns	Selected MR Patterns	MAX values Degree of Match
Car-C-1	0.1	21	MR-2: ({Hight,Med,Good},{},100)	33.33%
	0.5	21	MR-6: ({Big,Med},{},100)	50%
	1.0	21	MR-4: ({More},{},100)	100%
	5.0	12	MR-7: ({High,Med,Good},{High,Med,Good},100)	100%
	10.0	18	MR-12: ({High,Med,Good},{High},100)	25%
Car-C-2	0.1	13	MR-4: ({VGood},{},100)	100%
	0.5	13	MR-2: ({5More},{},100)	100%
	1.0	13	MR-8: ({5More},{More},100)	100%
	5.0	9	MR-8: ({VHigh,Med,Low},{Med},100)	50%
	10.0	9	MR-9: ({VHigh,Med,Low},{VHigh,Low},100)	50%
Car-C-3	0.1	25	MR-4: ({Med,Low},{},100)	50%
	0.5	25	MR-6: ({Med,Big},{},100)	50%
	1.0	25	MR-8: ({2,4},{},100)	50%
	5.0	25	MR-10: ({High,VHigh,Med},{},100)	33.33%
	10.0	25	MR-20: ({Med,Big},{Med},100)	66.67%
Nursery-N-1	0.1	12	MR-4: ({Convenient},{},100)	100%
	0.5	12	MR-6: ({Problematic},{},100)	100%
	1.0	12	MR-8: ({Usual},{},100)	100%
	5.0	12	MR-2: ({Priority,Recommended},{},100)	50%
	10.0	12	MR-3: ({}, {Convenient},66.67)	100%
Nursery-N-2	0.1	5	MR-1: ({}, {Convenient,Critical}, 26,67)	100%
	0.5	5	MR-2: ({Convenient,Critical},{},100)	50%
	1.0	5	MR-3: ({Convenient},{Convenient,Critical},40)	100%
	5.0	5	MR-4: ({Convenient,Critical},{Convenient},100)	66.67%
	10.0	5	MR-5: ({Convenient,Critical},{Convenient,Critical},100)	75%
Nursery-N-3	0.1	10	MR-2: ({Convenient,Critical,Less_Conv},{},100)	33.33%
	0.5	10	MR-4: ({Convenient,InConv},{},100)	50%
	1.0	10	MR-9: ({Convenient,InConv},{Convenient},100)	66.67%
	5.0	8	MR-6: ({Convenient},{Convenient,InConv},100)	66.67%

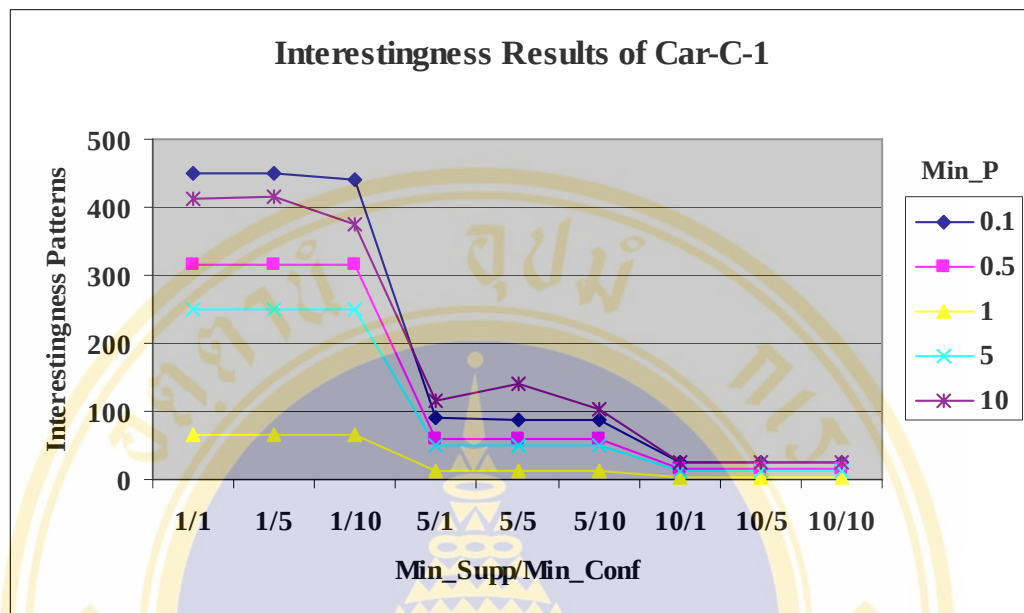
Table 6-5: The Summary Results of the Interestingness Testing (Cont.)

Domain	Min_P	Total of MR Patterns	Selected MR Patterns	MAX values Degree of Match
Adult-A-1	10.0	8	MR-4: ({Convenient, Critical, Less_Conv}, {Convenient}, 100)	50%
	0.1	18	MR-1: ({}, {17}, 6.86)	100%
	0.5	18	MR-13: ({Male}, {}, 100)	100%
	1.0	17	MR-8: ({Bachelor}, {}, 100)	100%
	5.0	15	MR-5: ({Bachelor, Male, 17}, {}, 100)	100%
Adult-A-2	10.0	13	MR-10: ({Male}, {Male}, 100)	100%
	0.1	28	MR-1: ({Male, Female}, {}, 100)	50%
	0.5	22	MR-4: ({Husband}, {}, 100)	100%
	1.0	22	MR-6: ({Never-Married}, {}, 100)	100%
	5.0	22	MR-15: ({}, {Husband}, 100)	100%
Adult-A-3	10.0	22	MR-9: ({Some-College, Master, Bachelor}, {}, 100)	33.33%
	0.1	28	MR-1: ({Never-married, Divorced}, {}, 100)	100%
	0.5	28	MR-6: ({25}, {}, 100)	100%
	1.0	28	MR-9: ({Husband, Wife}, {}, 100)	50%
	5.0	28	MR-12: ({Sales, Other-Service}, {}, 100)	100%
	10.0	28	MR-15: ({Bachelor, Master}, {}, 100)	100%

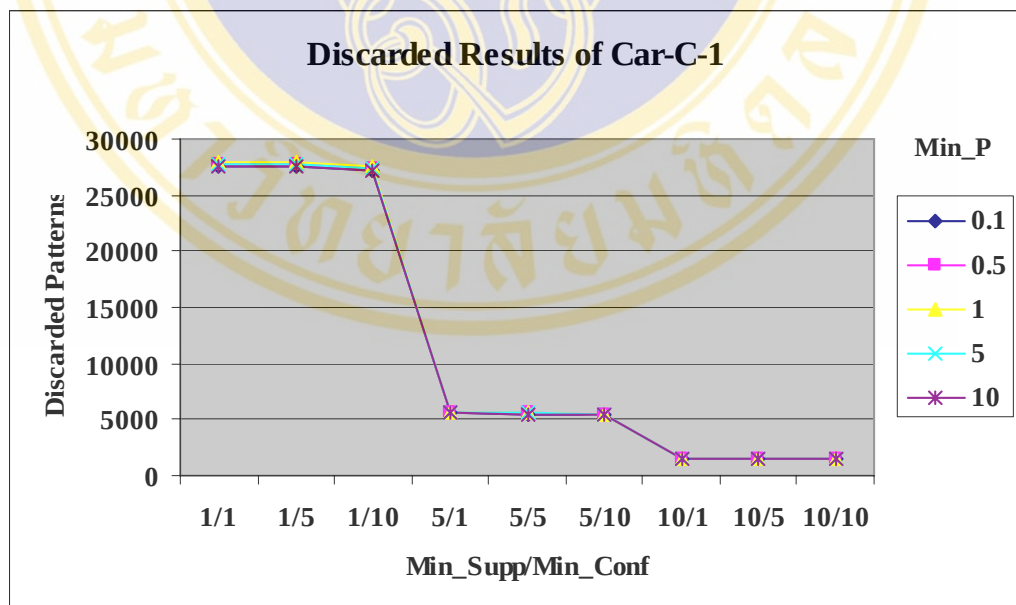
The details of the interestingness testing are summarized in Table 6-6 to 6-14 and the comparative interestingness/discarded results of each data domain are plotted in Figure 6-11 to 6-19.

Table 6-6: The Detail Results (Car-C-1 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of Discarded Patterns
Car-C-1	0.1	1.0	1.0	MR-2	28,050	33.33%	448	27,602
	0.5	1.0	1.0	MR-6	28,050	50%	314	27,736
	1.0	1.0	1.0	MR-4	28,050	100%	66	27,984
	5.0	1.0	1.0	MR-7	28,050	100%	250	27,800
	10.0	1.0	1.0	MR-12	28,050	25%	450	27,600
	0.1	1.0	5.0	MR-2	27,986	33.33%	449	27,537
	0.5	1.0	5.0	MR-6	27,986	50%	315	27,671
	1.0	1.0	5.0	MR-4	27,986	100%	66	27,920
	5.0	1.0	5.0	MR-7	27,986	100%	251	27,735
	10.0	1.0	5.0	MR-12	27,986	25%	449	27,537
	0.1	1.0	10.0	MR-2	27,669	33.33%	440	27,229
	0.5	1.0	10.0	MR-6	27,669	50%	315	27,354
	1.0	1.0	10.0	MR-4	27,669	100%	66	27,603
	5.0	1.0	10.0	MR-7	27,669	100%	250	27,419
	10.0	1.0	10.0	MR-12	27,669	25%	440	27,229
	0.1	5.0	1.0	MR-2	5,649	33.33%	90	5,559
	0.5	5.0	1.0	MR-6	5,649	50%	61	5,589
	1.0	5.0	1.0	MR-4	5,649	100%	13	5,637
	5.0	5.0	1.0	MR-7	5,649	100%	49	5,601
	10.0	5.0	1.0	MR-12	5,649	25%	90	5,559
	0.1	5.0	5.0	MR-2	5,598	33.33%	89	5,509
	0.5	5.0	5.0	MR-6	5,598	50%	61	5,538
	1.0	5.0	5.0	MR-4	5,598	100%	13	5,586
	5.0	5.0	5.0	MR-7	5,598	100%	49	5,550
	10.0	5.0	5.0	MR-12	5,598	25%	89	5,509
	0.1	5.0	10.0	MR-2	5,499	33.33%	89	5,410
	0.5	5.0	10.0	MR-6	5,499	50%	61	5,549
	1.0	5.0	10.0	MR-4	5,499	100%	13	5,487
	5.0	5.0	10.0	MR-7	5,499	100%	49	5,451
	10.0	5.0	10.0	MR-12	5,499	25%	89	5,410
	0.1	10.0	1.0	MR-2	1,569	33.33%	25	1,544
	0.5	10.0	1.0	MR-6	1,569	50%	16	1,553
	1.0	10.0	1.0	MR-4	1,569	100%	3	1,566
	5.0	10.0	1.0	MR-7	1,569	100%	13	1,556
	10.0	10.0	1.0	MR-12	1,569	25%	25	1,544
	0.1	10.0	5.0	MR-2	1,552	33.33%	25	1,527
	0.5	10.0	5.0	MR-6	1,552	50%	16	1,536
	1.0	10.0	5.0	MR-4	1,552	100%	3	1,549
	5.0	10.0	5.0	MR-7	1,552	100%	13	1,539
	10.0	10.0	5.0	MR-12	1,552	25%	25	1,527
	0.1	10.0	10.0	MR-2	1,528	33.33%	24	1,504
	0.5	10.0	10.0	MR-6	1,528	50%	16	1,512
	1.0	10.0	10.0	MR-4	1,528	100%	3	1,525
	5.0	10.0	10.0	MR-7	1,528	100%	13	1,515
	10.0	10.0	10.0	MR-12	1,528	25%	24	1,504



(a)

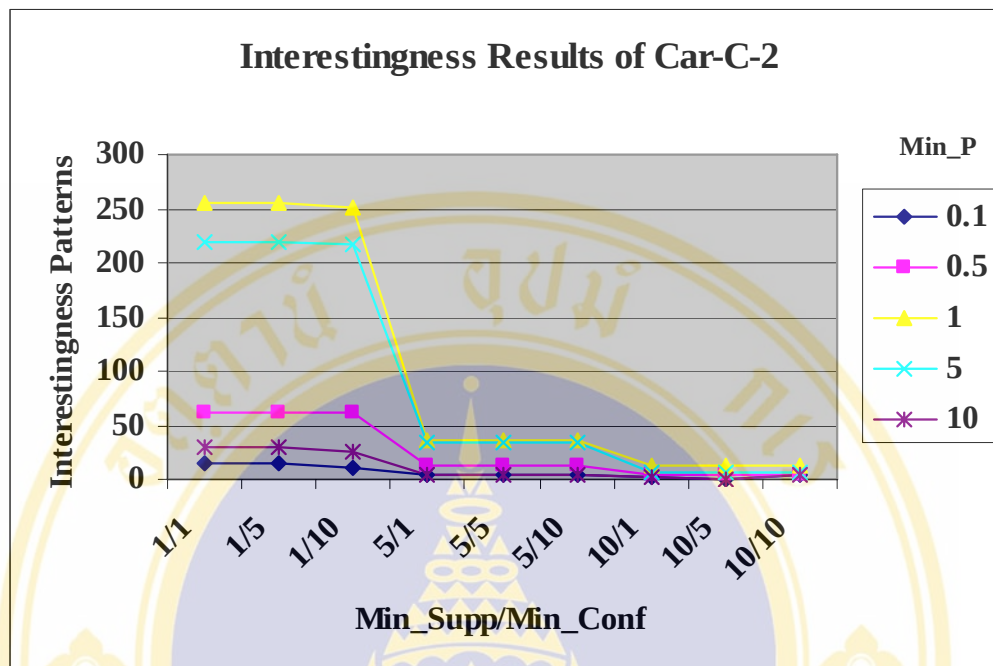


(b)

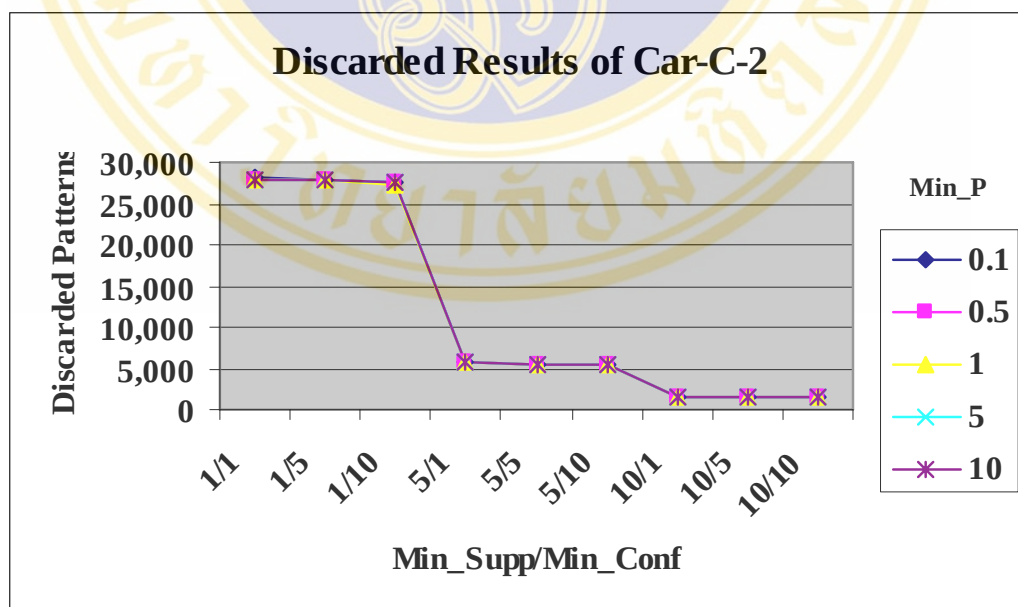
Figure 6-11: Comparative Interestingness/Discarded Results of Car-C-1 Domain

Table 6-7: The Detail Results (Car-C-2 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of Discarded Patterns
Car-C-2	0.1	1.0	1.0	MR-4	28,050	100%	15	28,035
	0.5	1.0	1.0	MR-2	28,050	100%	61	27,989
	1.0	1.0	1.0	MR-8	28,050	100%	256	27,794
	5.0	1.0	1.0	MR-11	28,050	50%	219	27,831
	10.0	1.0	1.0	MR-7	28,050	50%	31	28,019
	0.1	1.0	5.0	MR-4	27,986	100%	14	27,972
	0.5	1.0	5.0	MR-2	27,986	100%	61	27,925
	1.0	1.0	5.0	MR-8	27,986	100%	256	27,731
	5.0	1.0	5.0	MR-11	27,986	50%	219	27,767
	10.0	1.0	5.0	MR-7	27,986	50%	31	27,955
	0.1	1.0	10.0	MR-4	27,669	100%	10	27,659
	0.5	1.0	10.0	MR-2	27,669	100%	61	27,608
	1.0	1.0	10.0	MR-8	27,669	100%	250	27,419
	5.0	1.0	10.0	MR-11	27,669	50%	217	27,452
	10.0	1.0	10.0	MR-7	27,669	50%	27	27,642
	0.1	5.0	1.0	MR-4	5,649	100%	5	5,644
	0.5	5.0	1.0	MR-2	5,649	100%	12	5,637
	1.0	5.0	1.0	MR-8	5,649	100%	37	5,612
	5.0	5.0	1.0	MR-11	5,649	50%	34	5,616
	10.0	5.0	1.0	MR-7	5,649	50%	5	5,644
	0.1	5.0	5.0	MR-4	5,598	100%	4	5,594
	0.5	5.0	5.0	MR-2	5,598	100%	12	5,586
	1.0	5.0	5.0	MR-8	5,598	100%	36	5,562
	5.0	5.0	5.0	MR-11	5,598	50%	33	5,565
	10.0	5.0	5.0	MR-7	5,598	50%	4	5,594
	0.1	5.0	10.0	MR-4	5,499	100%	4	5,495
	0.5	5.0	10.0	MR-2	5,499	100%	12	5,487
	1.0	5.0	10.0	MR-8	5,499	100%	36	5,463
	5.0	5.0	10.0	MR-11	5,499	50%	33	5,466
	10.0	5.0	10.0	MR-7	5,499	50%	4	5,495
	0.1	10.0	1.0	MR-4	1,569	100%	1	1,568
	0.5	10.0	1.0	MR-2	1,569	100%	4	1,565
	1.0	10.0	1.0	MR-8	1,569	100%	12	1,557
	5.0	10.0	1.0	MR-11	1,569	50%	6	1,563
	10.0	10.0	1.0	MR-7	1,569	50%	1	1,568
	0.1	10.0	5.0	MR-4	1,552	100%	1	1,551
	0.5	10.0	5.0	MR-2	1,552	100%	4	1,548
	1.0	10.0	5.0	MR-8	1,552	100%	12	1,540
	5.0	10.0	5.0	MR-11	1,552	50%	6	1,546
	10.0	10.0	5.0	MR-7	1,552	50%	1	1,551
	0.1	10.0	10.0	MR-4	1,528	100%	2	1,526
	0.5	10.0	10.0	MR-2	1,528	100%	4	1,524
	1.0	10.0	10.0	MR-8	1,528	100%	12	1,516
	5.0	10.0	10.0	MR-11	1,528	50%	6	1,522
	10.0	10.0	10.0	MR-7	1,528	50%	5	1,523



(a)

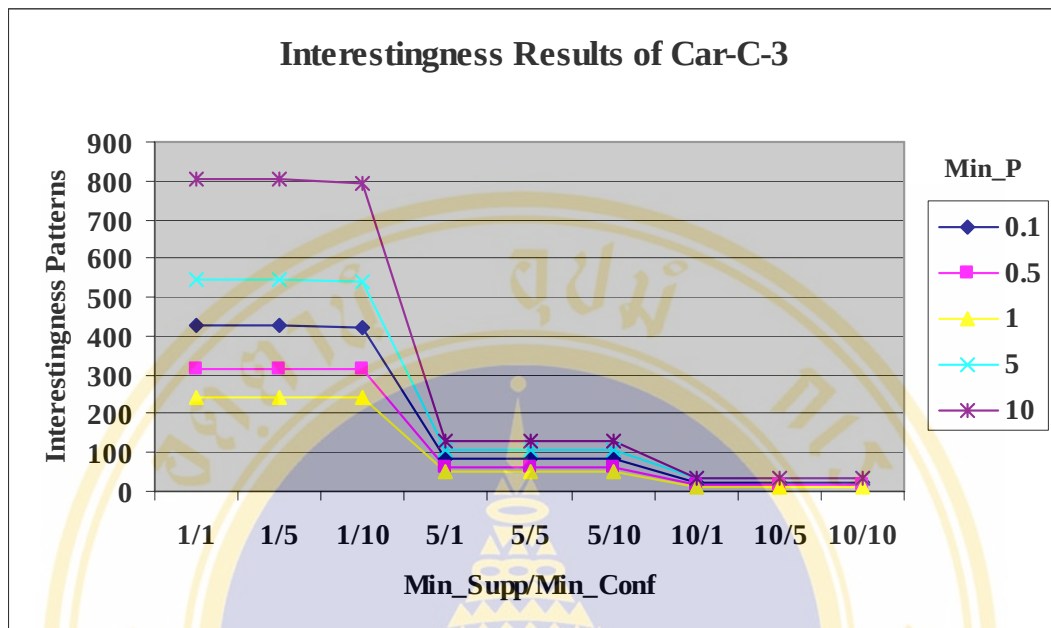


(b)

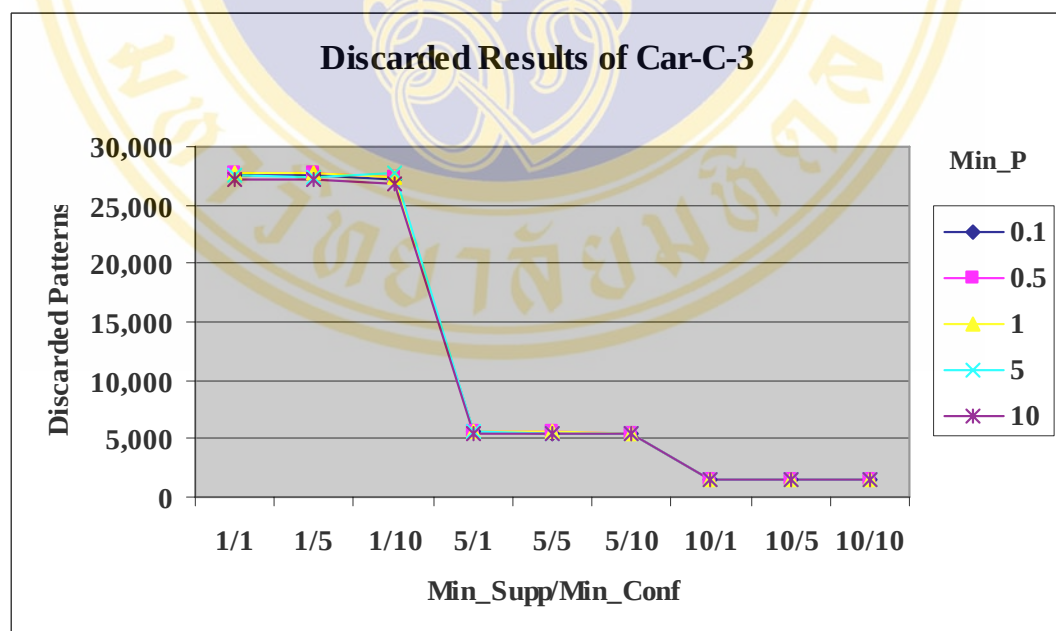
Figure 6-12: Comparative Interestingness/Discarded Result of Car-C-2 Domain

Table 6-8: The Detail Results (Car-C-3 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Car-C-3	0.1	1.0	1.0	MR-4	28,050	50%	426	27,624
	0.5	1.0	1.0	MR-6	28,050	50%	315	27,735
	1.0	1.0	1.0	MR-8	28,050	50%	244	27,806
	5.0	1.0	1.0	MR-10	28,050	33.33%	544	27,506
	10.0	1.0	1.0	MR-20	28,050	66.67%	806	27,245
	0.1	1.0	5.0	MR-4	27,986	50%	426	27,560
	0.5	1.0	5.0	MR-6	27,986	50%	315	27,671
	1.0	1.0	5.0	MR-8	27,986	50%	244	27,742
	5.0	1.0	5.0	MR-10	27,986	33.33%	544	27,442
	10.0	1.0	5.0	MR-20	27,986	66.67%	805	27,181
	0.1	1.0	10.0	MR-4	27,669	50%	424	27,245
	0.5	1.0	10.0	MR-6	27,669	50%	315	27,354
	1.0	1.0	10.0	MR-8	27,669	50%	244	27,425
	5.0	1.0	10.0	MR-10	27,669	33.33%	81	27,668
	10.0	1.0	10.0	MR-20	27,669	66.67%	795	26,874
	0.1	5.0	1.0	MR-4	5,649	50%	84	5,565
	0.5	5.0	1.0	MR-6	5,649	50%	61	5,589
	1.0	5.0	1.0	MR-8	5,649	50%	48	5,601
	5.0	5.0	1.0	MR-10	5,649	33.33%	109	5,540
	10.0	5.0	1.0	MR-20	5,649	66.67%	131	5,518
	0.1	5.0	5.0	MR-4	5,598	50%	84	5,514
	0.5	5.0	5.0	MR-6	5,598	50%	61	5,538
	1.0	5.0	5.0	MR-8	5,598	50%	48	5,550
	5.0	5.0	5.0	MR-10	5,598	33.33%	109	5,489
	10.0	5.0	5.0	MR-20	5,598	66.67%	131	5,467
	0.1	5.0	10.0	MR-4	5,499	50%	84	5,415
	0.5	5.0	10.0	MR-6	5,499	50%	61	5,439
	1.0	5.0	10.0	MR-8	5,499	50%	48	5,451
	5.0	5.0	10.0	MR-10	5,499	33.33%	109	5,390
	10.0	5.0	10.0	MR-20	5,499	66.67%	131	5,368
	0.1	10.0	1.0	MR-4	1,569	50%	23	1,546
	0.5	10.0	1.0	MR-6	1,569	50%	16	1,553
	1.0	10.0	1.0	MR-8	1,569	50%	13	1,556
	5.0	10.0	1.0	MR-10	1,569	33.33%	31	1,538
	10.0	10.0	1.0	MR-20	1,569	66.67%	33	1,536
	0.1	10.0	5.0	MR-4	1,552	50%	23	1,529
	0.5	10.0	5.0	MR-6	1,552	50%	16	1,536
	1.0	10.0	5.0	MR-8	1,552	50%	13	1,539
	5.0	10.0	5.0	MR-10	1,552	33.33%	31	1,521
	10.0	10.0	5.0	MR-20	1,552	66.67%	33	1,519
	0.1	10.0	10.0	MR-4	1,528	50%	23	1,505
	0.5	10.0	10.0	MR-6	1,528	50%	16	1,512
	1.0	10.0	10.0	MR-8	1,528	50%	13	1,515
	5.0	10.0	10.0	MR-10	1,528	33.33%	31	1,497
	10.0	10.0	10.0	MR-20	1,528	66.67%	32	1,496



(a)

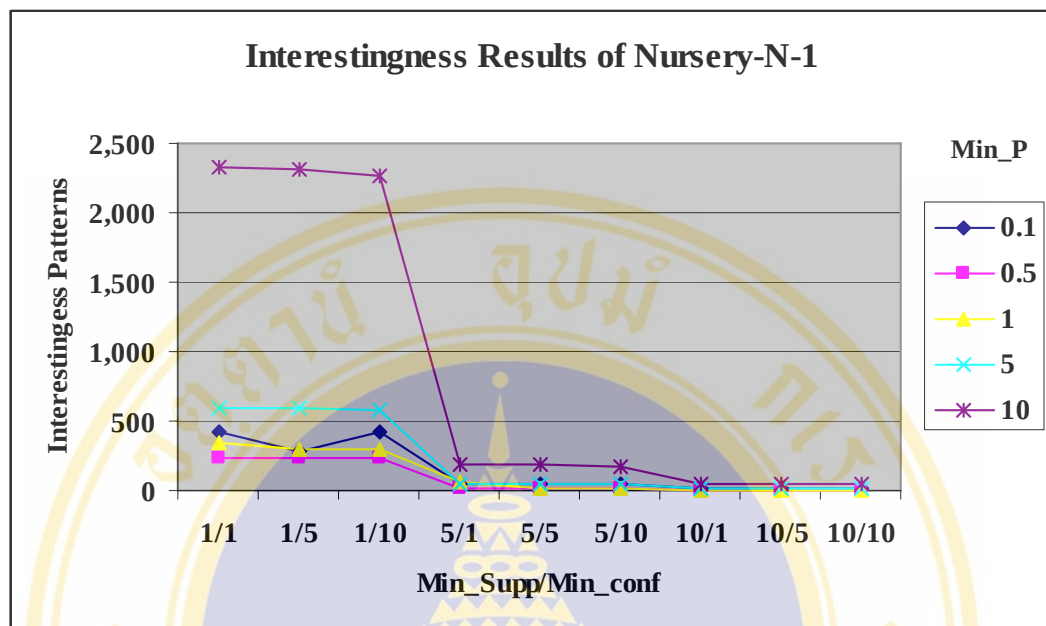


(b)

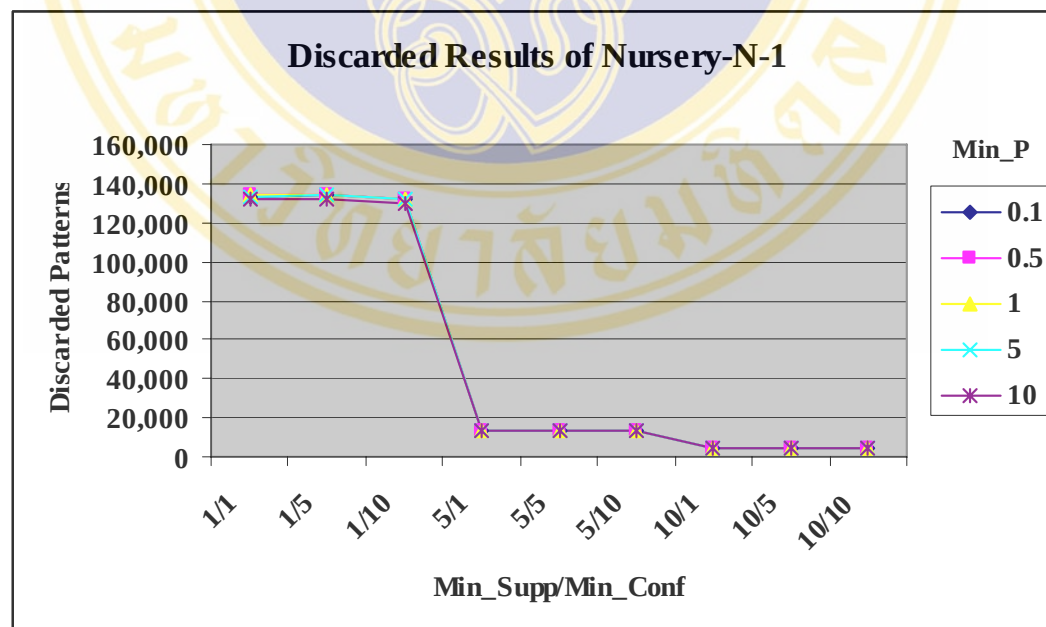
Figure 6-13: Comparative Interestingness/Discarded Result of Car-C-3 Domain

Table 6-9: The Detail Results (Nursery-N-1 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Nursery-N-1	0.1	1.0	1.0	MR-4	134,019	100%	425	133,594
	0.5	1.0	1.0	MR-6	134,019	100%	235	133,784
	1.0	1.0	1.0	MR-8	134,019	100%	233	133,786
	5.0	1.0	1.0	MR-2	134,019	50%	586	133,422
	10.0	1.0	1.0	MR-3	134,019	100%	2,330	131,689
	0.1	1.0	5.0	MR-4	134,723	100%	275	134,448
	0.5	1.0	5.0	MR-6	134,723	100%	234	134,489
	1.0	1.0	5.0	MR-8	134,723	100%	233	134,490
	5.0	1.0	5.0	MR-2	134,723	50%	591	134,132
	10.0	1.0	5.0	MR-3	134,723	100%	2,308	132,415
	0.1	1.0	10.0	MR-4	132,609	100%	425	132,184
	0.5	1.0	10.0	MR-6	132,609	100%	234	132,375
	1.0	1.0	10.0	MR-8	132,609	100%	233	132,376
	5.0	1.0	10.0	MR-2	132,609	50%	581	132,028
	10.0	1.0	10.0	MR-3	132,609	100%	2,273	130,336
	0.1	5.0	1.0	MR-4	13,969	100%	42	13,927
	0.5	5.0	1.0	MR-6	13,969	100%	22	13,947
	1.0	5.0	1.0	MR-8	13,969	100%	21	13,948
	5.0	5.0	1.0	MR-2	13,969	50%	55	13,914
	10.0	5.0	1.0	MR-3	13,969	100%	182	13,787
	0.1	5.0	5.0	MR-4	13,806	100%	41	13,765
	0.5	5.0	5.0	MR-6	13,806	100%	23	13,783
	1.0	5.0	5.0	MR-8	13,806	100%	24	13,782
	5.0	5.0	5.0	MR-2	13,806	50%	55	13,751
	10.0	5.0	5.0	MR-3	13,806	100%	180	13,626
	0.1	5.0	10.0	MR-4	13,611	100%	41	13,570
	0.5	5.0	10.0	MR-6	13,611	100%	23	13,588
	1.0	5.0	10.0	MR-8	13,611	100%	22	13,589
	5.0	5.0	10.0	MR-2	13,611	50%	54	13,557
	10.0	5.0	10.0	MR-3	13,611	100%	178	13,433
	0.1	10.0	1.0	MR-4	4,126	100%	12	4,114
	0.5	10.0	1.0	MR-6	4,126	100%	6	4,120
	1.0	10.0	1.0	MR-8	4,126	100%	6	4,120
	5.0	10.0	1.0	MR-2	4,126	50%	16	4,110
	10.0	10.0	1.0	MR-3	4,126	100%	47	4,079
	0.1	10.0	5.0	MR-4	4,072	100%	12	4,060
	0.5	10.0	5.0	MR-6	4,072	100%	6	4,066
	1.0	10.0	5.0	MR-8	4,072	100%	6	4,066
	5.0	10.0	5.0	MR-2	4,072	50%	16	4,056
	10.0	10.0	5.0	MR-3	4,072	100%	47	4,025
	0.1	10.0	10.0	MR-4	4,029	100%	12	4,017
	0.5	10.0	10.0	MR-6	4,029	100%	6	4,023
	1.0	10.0	10.0	MR-8	4,029	100%	6	4,023
	5.0	10.0	10.0	MR-2	4,029	50%	16	4,013
	10.0	10.0	10.0	MR-3	4,029	100%	47	3,982



(a)

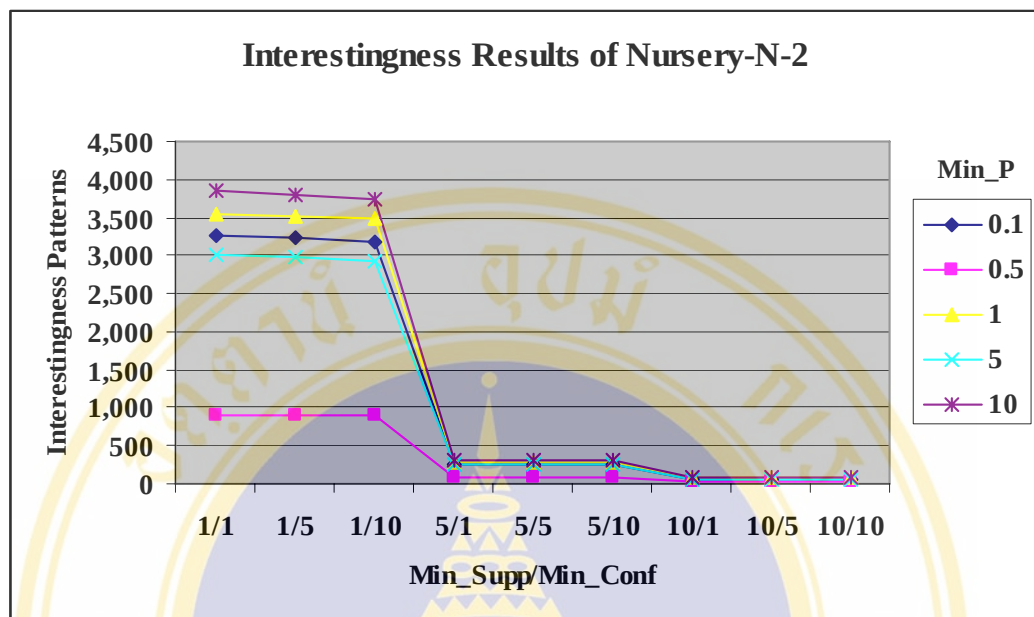


(b)

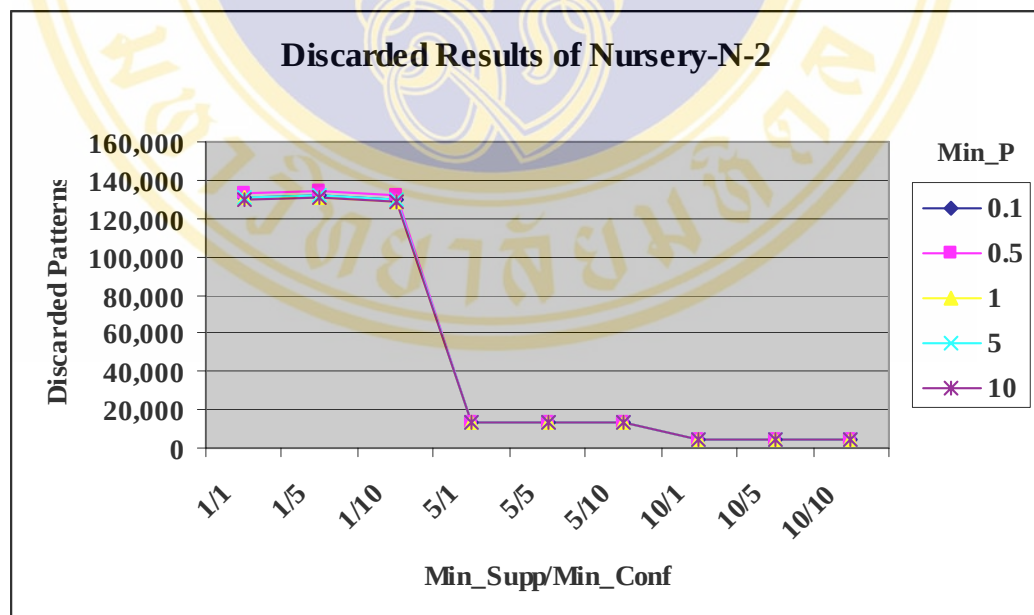
Figure 6-14: Comparative Interestingness/Discarded Result of Nursery-N-1 Domain

Table 6-10: The Detail Results (Nursery-N-2 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Nursery-N-2	0.1	1.0	1.0	MR-1	134,019	100%	3,261	130,758
	0.5	1.0	1.0	MR-2	134,019	50%	905	133,115
	1.0	1.0	1.0	MR-3	134,019	100%	3,553	130,466
	5.0	1.0	1.0	MR-4	134,019	66.67%	3,001	131,018
	10.0	1.0	1.0	MR-5	134,019	75%	3,841	130,178
	0.1	1.0	5.0	MR-1	134,723	100%	3,230	131,493
	0.5	1.0	5.0	MR-2	134,723	50%	900	133,823
	1.0	1.0	5.0	MR-3	134,723	50%	3,522	131,201
	5.0	1.0	5.0	MR-4	134,723	66.67%	2,975	131,748
	10.0	1.0	5.0	MR-5	134,723	75%	3,807	130,916
	0.1	1.0	10.0	MR-1	132,609	100%	3,186	129,423
	0.5	1.0	10.0	MR-2	132,609	50%	887	131,722
	1.0	1.0	10.0	MR-3	132,609	100%	3,477	129,132
	5.0	1.0	10.0	MR-4	132,609	66.67%	2,930	129,679
	10.0	1.0	10.0	MR-5	132,609	75%	3,755	128,854
	0.1	5.0	1.0	MR-1	13,969	100%	249	13,720
	0.5	5.0	1.0	MR-2	13,969	50%	89	13,880
	1.0	5.0	1.0	MR-3	13,969	100%	283	13,686
	5.0	5.0	1.0	MR-4	13,969	66.67%	255	13,714
	10.0	5.0	1.0	MR-5	13,969	75%	317	13,653
	0.1	5.0	5.0	MR-1	13,806	100%	247	13,559
	0.5	5.0	5.0	MR-2	13,806	50%	89	13,717
	1.0	5.0	5.0	MR-3	13,806	100%	280	13,526
	5.0	5.0	5.0	MR-4	13,806	67.67%	253	13,553
	10.0	5.0	5.0	MR-5	13,806	75%	313	13,493
	0.1	5.0	10.0	MR-1	13,611	100%	243	13,368
	0.5	5.0	10.0	MR-2	13,611	50%	89	13,522
	1.0	5.0	10.0	MR-3	13,611	100%	276	13,335
	5.0	5.0	10.0	MR-4	13,611	66.67%	250	13,361
	10.0	5.0	10.0	MR-5	13,611	75%	311	13,300
	0.1	10.0	1.0	MR-1	4,126	100%	64	4,062
	0.5	10.0	1.0	MR-2	4,126	50%	26	4,100
	1.0	10.0	1.0	MR-3	4,126	66.67%	75	4,051
	5.0	10.0	1.0	MR-4	4,126	66.67%	70	4,056
	10.0	10.0	1.0	MR-5	4,126	75%	86	4,040
	0.1	10.0	5.0	MR-1	4,072	100%	65	4,007
	0.5	10.0	5.0	MR-2	4,072	100%	26	4,046
	1.0	10.0	5.0	MR-3	4,072	66.67%	75	3,997
	5.0	10.0	5.0	MR-4	4,072	66.67%	68	4,004
	10.0	10.0	5.0	MR-5	4,072	75%	85	3,987
	0.1	10.0	10.0	MR-1	4,029	100%	65	3,965
	0.5	10.0	10.0	MR-2	4,029	50%	26	4,003
	1.0	10.0	10.0	MR-3	4,029	66.67%	73	3,956
	5.0	10.0	10.0	MR-4	4,029	66.67%	68	3,961
	10.0	10.0	10.0	MR-5	4,029	75%	84	3,945



(a)

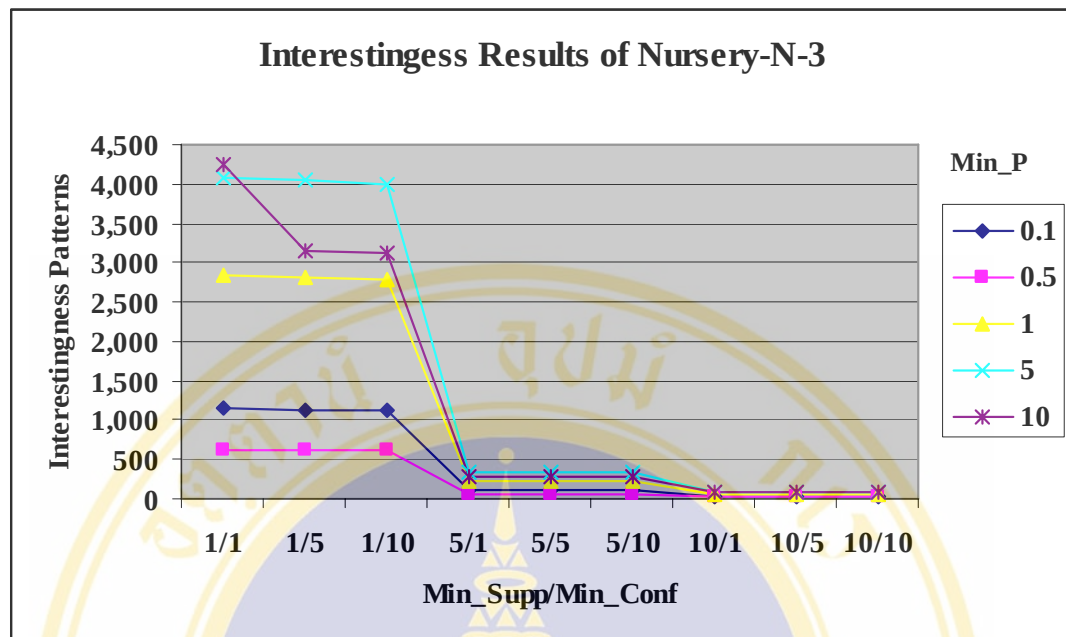


(b)

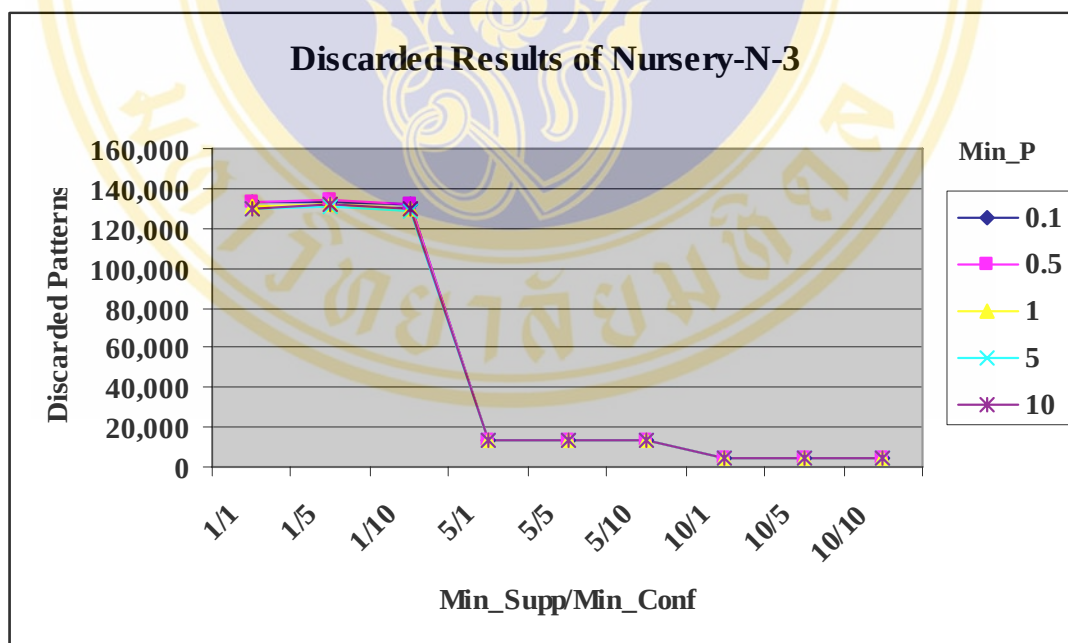
Figure 6-15: Comparative Interestingness/Discarded Results of Nursery-N-2 Domain

Table 6-11: The Detail Results (Nursery-N-3 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Nursery-N-3	0.1	1.0	1.0	MR-2	134,019	33.33%	1,141	132,878
	0.5	1.0	1.0	MR-4	134,019	50%	614	133,405
	1.0	1.0	1.0	MR-9	134,019	66.67%	2,839	131,180
	5.0	1.0	1.0	MR-6	134,019	66.67%	4,077	129,942
	10.0	1.0	1.0	MR-4	134,019	50%	4,242	129,777
	0.1	1.0	5.0	MR-2	134,723	33.33%	1,136	133,587
	0.5	1.0	5.0	MR-4	134,723	50%	614	134,109
	1.0	1.0	5.0	MR-9	134,723	66.67%	2,817	131,906
	5.0	1.0	5.0	MR-6	134,723	66.67%	4,044	130,679
	10.0	1.0	5.0	MR-4	134,723	50%	3,153	131,570
	0.1	1.0	10.0	MR-2	132,609	33.33%	1,123	131,486
	0.5	1.0	10.0	MR-4	132,609	50%	614	131,995
	1.0	1.0	10.0	MR-9	132,609	66.67%	2,781	129,828
	5.0	1.0	10.0	MR-6	132,609	66.67%	3,986	128,623
	10.0	1.0	10.0	MR-4	132,609	50%	3,108	129,501
	0.1	5.0	1.0	MR-2	13,969	33.33%	111	13,858
	0.5	5.0	1.0	MR-4	13,969	50%	60	13,909
	1.0	5.0	1.0	MR-9	13,969	66.67%	236	13,733
	5.0	5.0	1.0	MR-6	13,969	66.67%	335	13,634
	10.0	5.0	1.0	MR-4	13,969	50%	274	13,695
	0.1	5.0	5.0	MR-2	13,806	33.33%	112	13,694
	0.5	5.0	5.0	MR-4	13,806	50%	60	13,746
	1.0	5.0	5.0	MR-9	13,806	66.67%	234	13,572
	5.0	5.0	5.0	MR-6	13,806	66.67%	332	13,474
	10.0	5.0	5.0	MR-4	13,806	50%	271	13,535
	0.1	5.0	10.0	MR-2	13,611	33.33%	111	13,500
	0.5	5.0	10.0	MR-4	13,611	50%	60	13,551
	1.0	5.0	10.0	MR-9	13,611	66.67%	231	13,380
	5.0	5.0	10.0	MR-6	13,611	66.67%	328	13,283
	10.0	5.0	10.0	MR-4	13,611	50%	268	13,343
	0.1	10.0	1.0	MR-2	4,126	33.33%	32	4,094
	0.5	10.0	1.0	MR-4	4,126	50%	18	4,108
	1.0	10.0	1.0	MR-9	4,126	66.67%	64	4,062
	5.0	10.0	1.0	MR-6	4,126	66.67%	88	4,038
	10.0	10.0	1.0	MR-4	4,126	50%	75	4,051
	0.1	10.0	5.0	MR-2	4,072	33.33%	32	4,040
	0.5	10.0	5.0	MR-4	4,072	50%	18	4,054
	1.0	10.0	5.0	MR-9	4,072	66.67%	63	4,009
	5.0	10.0	5.0	MR-6	4,072	66.67%	87	3,985
	10.0	10.0	5.0	MR-4	4,072	50%	74	3,998
	0.1	10.0	10.0	MR-2	4,029	33.33%	32	3,997
	0.5	10.0	10.0	MR-4	4,029	50%	18	4,011
	1.0	10.0	10.0	MR-9	4,029	66.67%	62	3,967
	5.0	10.0	10.0	MR-6	4,029	66.67%	86	3,943
	10.0	10.0	10.0	MR-4	4,029	50%	73	3,956



(a)

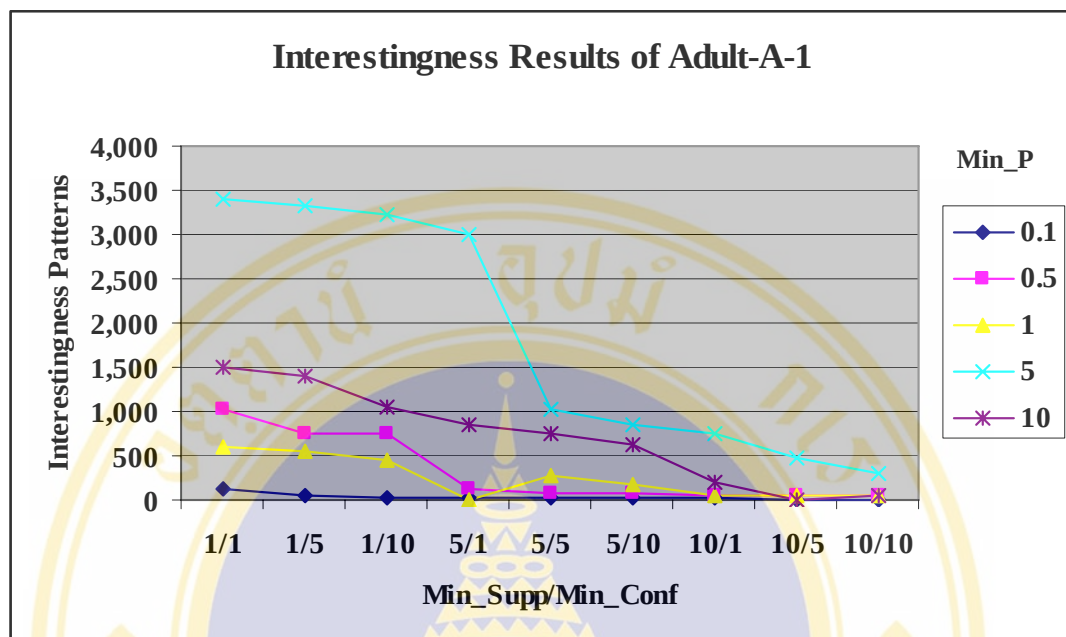


(b)

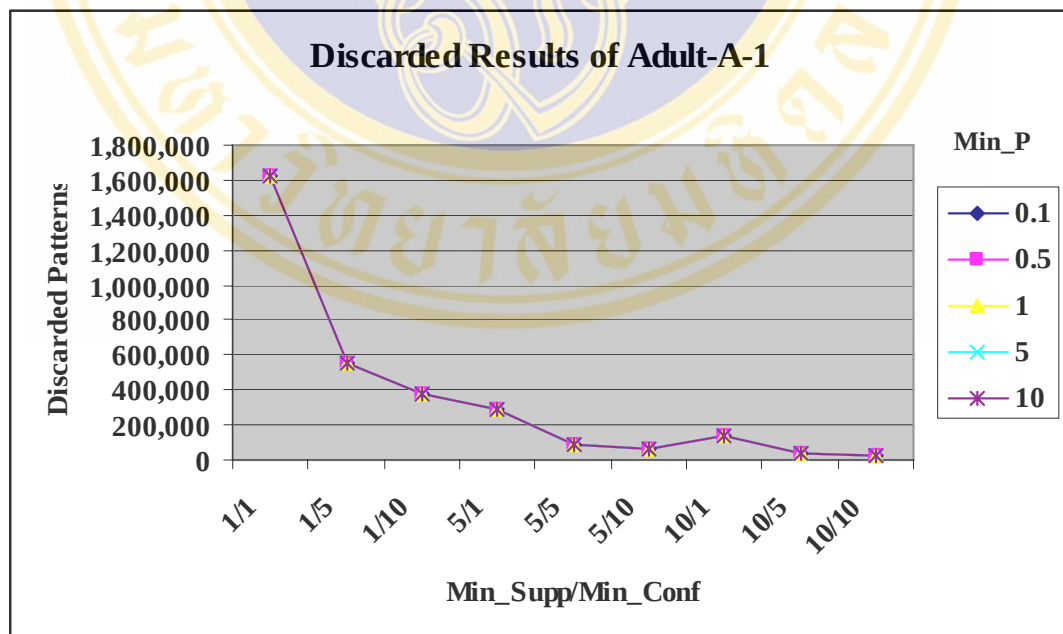
Figure 6-16: Comparative Interestingness/Discarded Results of Nursery-N-3 Domain

Table 6-12: The Detail Results (Adult-A-1 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Adult-A-1	0.1	1.0	1.0	MR-1	1,623,767	100%	126	1,623,641
	0.5	1.0	1.0	MR-13	1,623,767	100%	1,014	1,622,753
	1.0	1.0	1.0	MR-8	1,623,767	100%	597	1,623,170
	5.0	1.0	1.0	MR-5	1,623,767	100%	3,405	1,620,362
	10.0	1.0	1.0	MR-10	1,623,767	100%	1,507	1,622,260
	0.1	1.0	5.0	MR-1	558,770	100%	61	558,709
	0.5	1.0	5.0	MR-13	558,770	100%	744	558,026
	1.0	1.0	5.0	MR-8	558,770	100%	539	558,231
	5.0	1.0	5.0	MR-10	558,770	100%	3,325	555,445
	10.0	1.0	5.0	MR-5	558,770	100%	1,412	557,358
	0.1	1.0	10.0	MR-1	379,587	100%	33	379,554
	0.5	1.0	10.0	MR-13	379,587	100%	741	378,846
	1.0	1.0	10.0	MR-8	379,587	100%	457	379,130
	5.0	1.0	10.0	MR-10	379,587	100%	3,215	276,372
	10.0	1.0	10.0	MR-5	379,587	100%	1,041	378,546
	0.1	5.0	1.0	MR-1	292,963	100%	25	292,938
	0.5	5.0	1.0	MR-13	292,963	100%	116	292,847
	1.0	5.0	1.0	MR-8	292,963	100%	5	292,958
	5.0	5.0	1.0	MR-10	292,963	100%	3,011	289,952
	10.0	5.0	1.0	MR-5	292,963	100%	842	292,121
	0.1	5.0	5.0	MR-1	92,426	100%	23	92,404
	0.5	5.0	5.0	MR-13	92,426	100%	85	92,341
	1.0	5.0	5.0	MR-8	92,426	100%	266	92,160
	5.0	5.0	5.0	MR-10	92,426	100%	1,025	91,401
	10.0	5.0	5.0	MR-5	92,426	100%	748	91,678
	0.1	5.0	10.0	MR-1	61,792	100%	22	61,770
	0.5	5.0	10.0	MR-13	61,792	100%	75	61,717
	1.0	5.0	10.0	MR-8	61,792	100%	178	61,614
	5.0	5.0	10.0	MR-10	61,792	100%	841	60,951
	10.0	5.0	10.0	MR-5	61,792	100%	621	61,171
	0.1	10.0	1.0	MR-1	133,523	100%	15	133,509
	0.5	10.0	1.0	MR-13	133,523	100%	51	133,472
	1.0	10.0	1.0	MR-8	133,523	100%	48	133,476
	5.0	10.0	1.0	MR-10	133,523	100%	742	132,781
	10.0	10.0	1.0	MR-5	133,523	100%	198	133,325
	0.1	10.0	5.0	MR-1	42,256	100%	3	42,253
	0.5	10.0	5.0	MR-13	42,256	100%	51	42,205
	1.0	10.0	5.0	MR-8	42,256	100%	47	42,209
	5.0	10.0	5.0	MR-10	42,256	100%	481	41,775
	10.0	10.0	5.0	MR-5	42,256	100%	13	42,243
	0.1	10.0	10.0	MR-1	28,361	100%	3	28,358
	0.5	10.0	10.0	MR-13	28,361	100%	51	28,310
	1.0	10.0	10.0	MR-8	28,361	100%	44	28,317
	5.0	10.0	10.0	MR-10	28,361	100%	307	28,054
	10.0	10.0	10.0	MR-5	28,361	100%	43	28,318



(a)

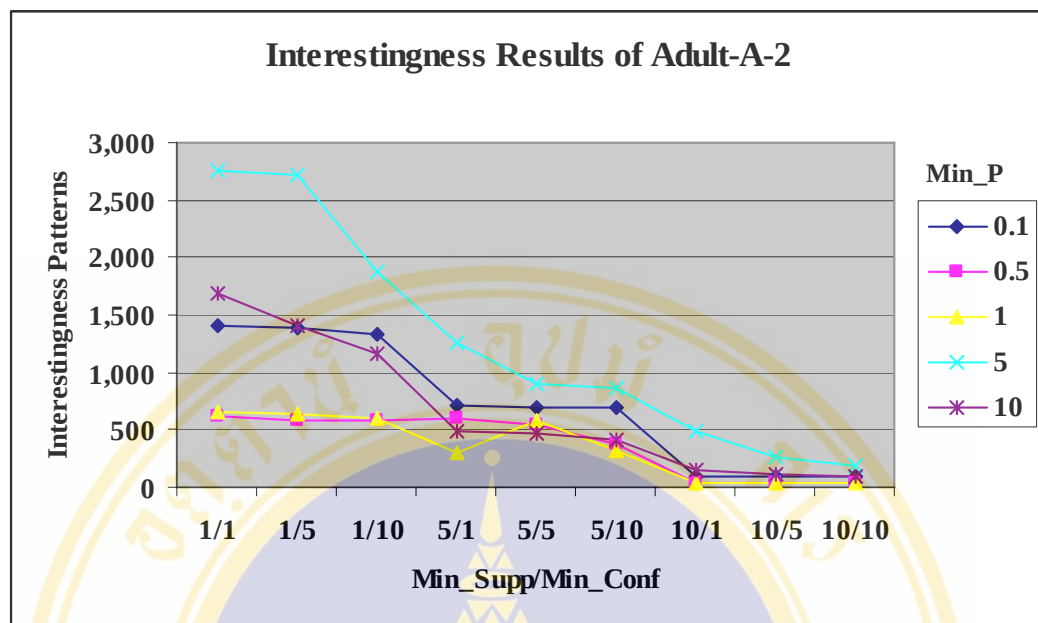


(b)

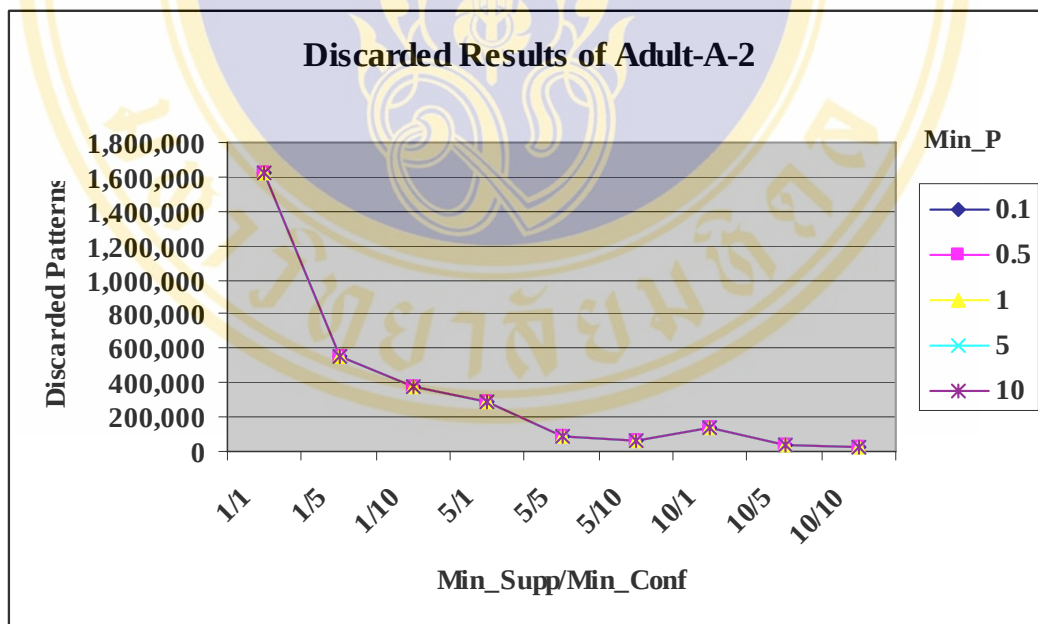
Figure 6-17: Comparative Interestingness/Discarded Results of Adult-A-1 Domain

Table 6-13: The Detail Results (Adult-A-2 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Adult-A-2	0.1	1.0	1.0	MR-1	1,623,767	50%	1,405	1,622,363
	0.5	1.0	1.0	MR-4	1,623,767	100%	613	1,623,154
	1.0	1.0	1.0	MR-6	1,623,767	100%	656	1,623,111
	5.0	1.0	1.0	MR-15	1,623,767	100%	2,765	1,621,002
	10.0	1.0	1.0	MR-9	1,623,767	33.33%	1,688	1,622,079
	0.1	1.0	5.0	MR-1	558,770	50%	1,387	557,383
	0.5	1.0	5.0	MR-4	558,770	100%	589	558,181
	1.0	1.0	5.0	MR-6	558,770	100%	641	558,129
	5.0	1.0	5.0	MR-15	558,770	100%	2,715	556,055
	10.0	1.0	5.0	MR-9	558,770	33.33%	1,400	557,370
	0.1	1.0	10.0	MR-1	379,587	50%	1,332	378,255
	0.5	1.0	10.0	MR-4	379,587	100%	576	379,011
	1.0	1.0	10.0	MR-6	379,587	100%	604	378,983
	5.0	1.0	10.0	MR-15	379,587	100%	1,867	377,720
	10.0	1.0	10.0	MR-9	379,587	33.33%	1,160	378,427
	0.1	5.0	1.0	MR-1	292,963	50%	213	292,750
	0.5	5.0	1.0	MR-4	292,963	100%	592	292,371
	1.0	5.0	1.0	MR-6	292,963	100%	102	292,861
	5.0	5.0	1.0	MR-15	292,963	100%	1,998	290,965
	10.0	5.0	1.0	MR-9	292,963	33.33%	489	292,474
	0.1	5.0	5.0	MR-1	92,426	50%	1,435	90,991
	0.5	5.0	5.0	MR-4	92,426	100%	549	91,877
	1.0	5.0	5.0	MR-6	92,426	100%	574	91,852
	5.0	5.0	5.0	MR-15	92,426	100%	92	92,334
	10.0	5.0	5.0	MR-9	92,426	33.33%	473	91,953
	0.1	5.0	10.0	MR-1	61,792	50%	929	60,863
	0.5	5.0	10.0	MR-4	61,792	100%	377	61,415
	1.0	5.0	10.0	MR-6	61,792	100%	317	61,475
	5.0	5.0	10.0	MR-15	61,792	100%	91	61,701
	10.0	5.0	10.0	MR-9	61,792	33.33%	412	61,380
	0.1	10.0	1.0	MR-1	133,523	50%	94	133,429
	0.5	10.0	1.0	MR-4	133,523	100%	43	133,480
	1.0	10.0	1.0	MR-6	133,523	100%	46	133,478
	5.0	10.0	1.0	MR-15	133,523	100%	960	132,563
	10.0	10.0	1.0	MR-9	133,523	33.33%	141	133,382
	0.1	10.0	5.0	MR-1	42,256	50%	92	42,164
	0.5	10.0	5.0	MR-4	42,256	100%	43	42,213
	1.0	10.0	5.0	MR-6	42,256	100%	45	42,211
	5.0	10.0	5.0	MR-15	42,256	100%	265	41,991
	10.0	10.0	5.0	MR-9	42,256	33.33%	114	42,142
	0.1	10.0	10.0	MR-1	28,361	50%	91	28,270
	0.5	10.0	10.0	MR-4	28,361	100%	43	28,318
	1.0	10.0	10.0	MR-6	28,361	100%	43	28,318
	5.0	10.0	10.0	MR-15	28,361	100%	181	28,180
	10.0	10.0	10.0	MR-9	28,361	33.33%	98	28,263



(a)

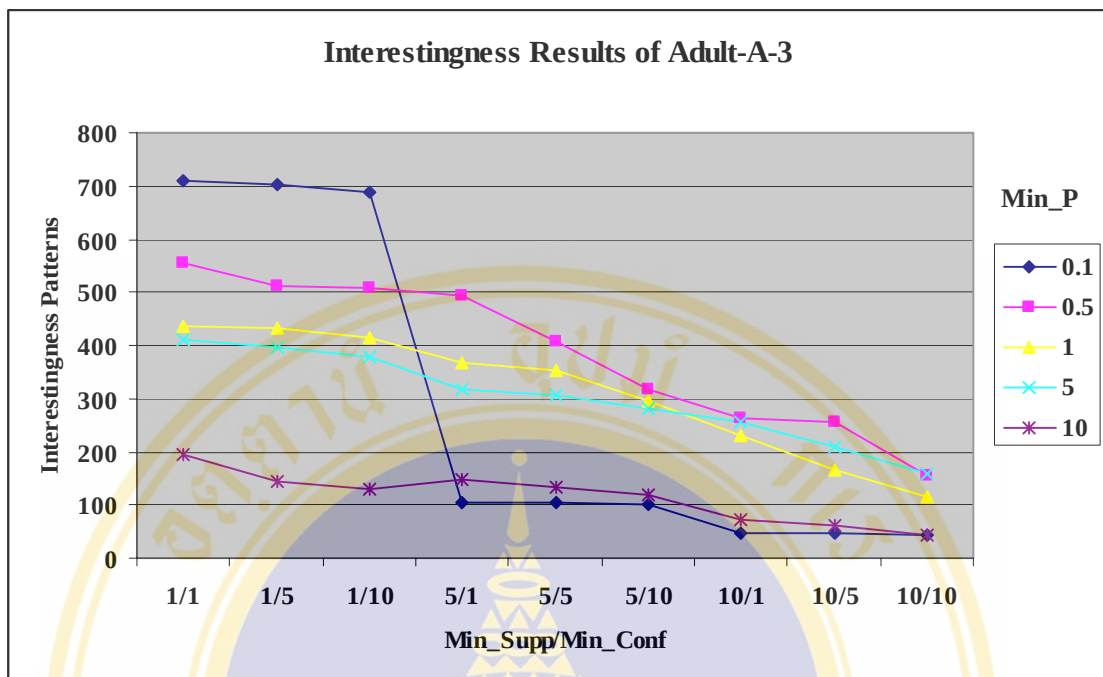


(b)

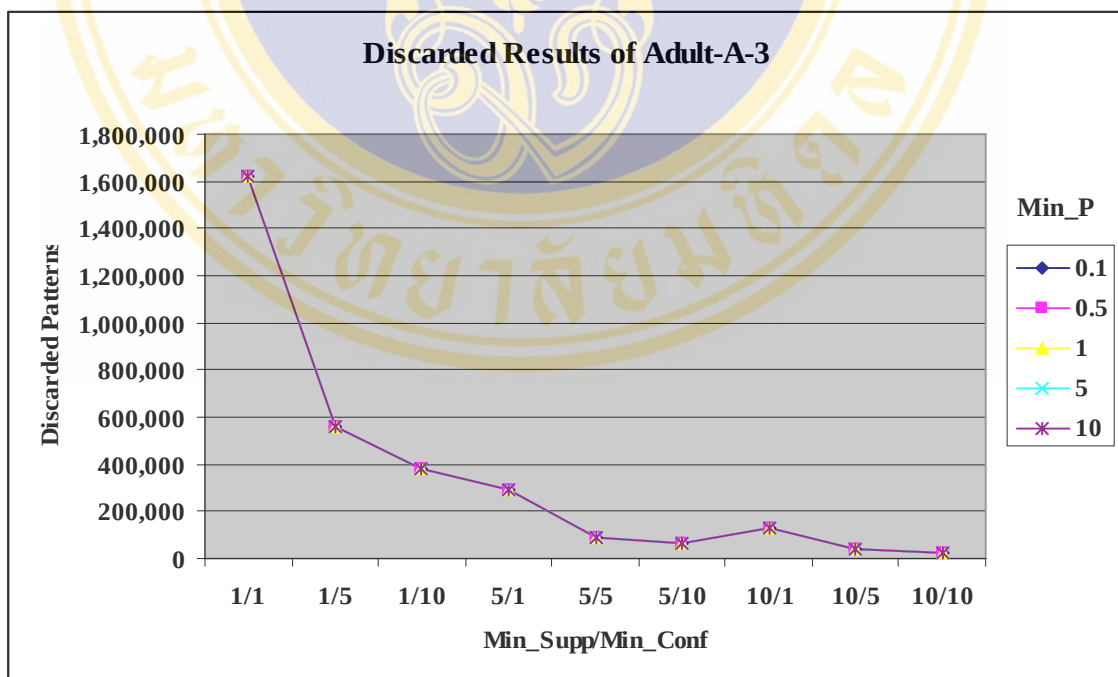
Figure 6-18: Comparative Discarded Results of Adult-A-2 Domain

Table 6-14: The Detail Results (Adult-A-3 Domain) of the Interestingness Testing

Domain	Min P	Min Supp	Min Conf	No. of Selected MR	Total of Discovered Patterns	Max Degree of Match	Total of Interestingness Patterns	Total of discarded Patterns
Adult-A-3	0.1	1.0	1.0	MR-1	1,623,767	100%	710	1,623,057
	0.5	1.0	1.0	MR-4	1,623,767	100%	556	1,623,211
	1.0	1.0	1.0	MR-6	1,623,767	100%	435	1,623,332
	5.0	1.0	1.0	MR-15	1,623,767	100%	412	1,623,356
	10.0	1.0	1.0	MR-9	1,623,767	100%	196	1,623,572
	0.1	1.0	5.0	MR-1	558,770	100%	702	558,068
	0.5	1.0	5.0	MR-4	558,770	100%	512	558,258
	1.0	1.0	5.0	MR-6	558,770	100%	431	558,339
	5.0	1.0	5.0	MR-15	558,770	100%	395	558,375
	10.0	1.0	5.0	MR-9	558,770	100%	144	558,626
	0.1	1.0	10.0	MR-1	379,587	100%	689	378,898
	0.5	1.0	10.0	MR-4	379,587	100%	507	379,080
	1.0	1.0	10.0	MR-6	379,587	100%	415	379,172
	5.0	1.0	10.0	MR-15	379,587	100%	377	379,210
	10.0	1.0	10.0	MR-9	379,587	100%	128	379,459
	0.1	5.0	1.0	MR-1	292,963	100%	106	292,857
	0.5	5.0	1.0	MR-4	292,963	100%	494	292,470
	1.0	5.0	1.0	MR-6	292,963	100%	366	292,597
	5.0	5.0	1.0	MR-15	292,963	100%	316	292,647
	10.0	5.0	1.0	MR-9	292,963	100%	149	292,814
	0.1	5.0	5.0	MR-1	92,426	100%	104	92,322
	0.5	5.0	5.0	MR-4	92,426	100%	409	92,018
	1.0	5.0	5.0	MR-6	92,426	100%	352	92,074
	5.0	5.0	5.0	MR-15	92,426	100%	305	92,121
	10.0	5.0	5.0	MR-9	92,426	100%	132	92,294
	0.1	5.0	10.0	MR-1	61,792	100%	100	61,692
	0.5	5.0	10.0	MR-4	61,792	100%	316	61,476
	1.0	5.0	10.0	MR-6	61,792	100%	295	61,498
	5.0	5.0	10.0	MR-15	61,792	100%	281	61,511
	10.0	5.0	10.0	MR-9	61,792	100%	118	61,674
	0.1	10.0	1.0	MR-1	133,523	100%	47	133,476
	0.5	10.0	1.0	MR-4	133,523	100%	262	133,261
	1.0	10.0	1.0	MR-6	133,523	100%	232	133,291
	5.0	10.0	1.0	MR-15	133,523	100%	255	133,268
	10.0	10.0	1.0	MR-9	133,523	100%	73	133,450
	0.1	10.0	5.0	MR-1	42,256	100%	47	42,209
	0.5	10.0	5.0	MR-4	42,256	100%	256	42,001
	1.0	10.0	5.0	MR-6	42,256	100%	165	42,091
	5.0	10.0	5.0	MR-15	42,256	100%	210	42,046
	10.0	10.0	5.0	MR-9	42,256	100%	62	42,195
	0.1	10.0	10.0	MR-1	28,361	100%	44	28,317
	0.5	10.0	10.0	MR-4	28,361	100%	155	28,206
	1.0	10.0	10.0	MR-6	28,361	100%	117	28,245
	5.0	10.0	10.0	MR-15	28,361	100%	160	28,201
	10.0	10.0	10.0	MR-9	28,361	100%	44	28,317



(a)



(b)

Figure 6-19: Comparative Interestingness/Discarded Results of Adult-A-3 Domain

In summary, we can ensure that the results of the interestingness testing yield the small set of the interestingness patterns. If the minimum support/confidence values are increased then the interestingness patterns are reduced. This implies that we can reduce the spurious set of patterns and represent only target results to user.

In addition, the results of the ontology testing and the interestingness testing of the above test cases indicate the volume of the interestingness results prepared for the subjective evaluation by the user's participation. The details are described in section 6.4.2 and 6.4.3.

6.4 FCA2MR Functional Evaluation

To evaluate the functional modules of FCA2MR, the system needs to be evaluated in three parts, namely, the functional comparison of ontology generator, the interestingness evaluation, and the user participation evaluation.

6.4.1 The Functional Comparison of Three Ontology Generators

First, the functional comparison of the ontology generator is conducted by comparing the functional feature of three ontology generators. We compare the FCA2MR prototype tool with two other programs: Protégé2000 and ConImp. The comparisons are revealed as follows:

Table 6-15: The Functional Comparison of Three Ontology Generators

Program	Require User Knowledge	Automatic Level	Representation Style	Ontology Structure	Based Program
Protégé 2000	Yes	Manual	GUI.	Hierarchical Directory	Windows Based
ConImp	Yes	Automatic	Text	Tree	Dos Based
FCA2MR	Yes	Automatic	GUI.	Lattice / Taxonomy / Hierarchical Directory	Windows Based

Moreover, we summarize on the advantages of the FCA2MR prototype tool, especially, in term of the ontology generator. The details are described below.

1. Applicability

The ontology construction process is a prominent task of the FCA2MR system because it is responsible for generating the Meta-Rule patterns for the discovery of subjective interestingness. As the ontology process is constructed by the Formal Concept Analysis, a proven and mathematical technique supporting the conceptual data analysis and knowledge processing, the correctness of the ontology feature is thus practical and effective. Obviously, the benefits of using ontology to generate the Meta-Rules are:

- Flexible and efficient representation of the results.
- Clear expression of the ontology concepts, and ontology relationship.
- Ease of expressing the domain knowledge, in terms of ontology concepts.

2. Visualization

To support the FCA2MR system to be more flexible and actionable, the Formal Concept Analysis (FCA) theory is adopted to form the ontology in the graphical way. The additional benefits of the visualized ontology representation are:

- Systematic representation scheme of domain knowledge by means of lattice.
- Multiple views of the ontology.
- Interactive representation with range of ontology concepts, ontology relationship, size, and scale.
- On the fly process of the ontology generator.

3. Stepwise Process

The ontology and meta-rules generation process is designed to support the step by step generation process. Therefore it is easy to trace the activity of any generation process as well as to realize of how the particular domain knowledge is constructed.

4. Scalability

Since the ontology construction process is independent to the target source of data and the lattice scheme supports the complex structure of data, the system is scalable no matter the size of the database is.

6.4.2 Interestingness Evaluation

The objective of this part is to evaluate the interestingness results. The evaluation is performed by means of the comparative summary of the results with Magnum Opus Demo program [45, 46, 47].

Magnum Opus (MO) delivers the rules that are most interesting according to the measure that users selected. A search is conducted, within the restrictions specified by the user. MO finds the rules with the highest values on the specified measure of interestingness, such as, coverage, RHS coverage, strength, lift, leverage, and support.

The methodology that MO used in the post analysis is close to our methodology. MO provides the way to select the attribute from the database to cope the search space. Identically, our prototype system provides the meta-rules mechanism to bind the interestingness results.

With the 1000 tuples limitation of the MO, we need to scope the size of datasets to 1000 tuples as well.

The comparisons of the interestingness results are shown in Table 6-16.

Table 6-16: The Summary Results of the Interestingness Comparisons with Magnum Opus

Selected Target MR	Interestingness Results			UnComparable Results
	FCA2MR Max Level Degree of Match	FCA2MR Comparable Results	Magnum Opus Results	
Car_C_1 Domain (Min_Supp=1, Min_Conf=1) MR1: ({More}, {}, 100) MR2: ({Med}, {}, 100) MR3: ({High, Med, Low}, {}, 100) MR4: ({High, Med, Low}, {Med}, 100) MR5: ({Med}, {High, Med, Low}, 21.63)	36 485 265 75 34	14 100 182 44 14	14 100 182 44 14	0 0 0 0 0
Car_C_2 Domain (Min_Supp=1, Min_Conf=5) MR6: ({More}, {}, 100) MR7: ({Med}, {}, 100) MR8: ({VHigh, Med}, {}, 100) MR9: ({Med}, {VHigh, Med}, 64.04) MR10: ({VHigh, Med}, {Med}, 100)	195 395 513 360 412	13 44 56 40 51	13 44 56 40 51	0 0 0 0 0
Car_C_3 Domain (Min_Supp=1, Min_Conf=10) MR11: ({More}, {}, 100) MR12: ({Med}, {}, 100) MR13: ({High, VHigh, Med, Low}, {}, 100) MR14: ({High, Low}, {Med, Low}, 51.68) MR15: ({Med}, {Med, Low}, 47.17)	24 74 186 21 12	13 32 100 11 9	13 32 100 11 9	0 0 0 0 0

Table 6-16: The Summary Results of the Interestingness Comparisons with Magnum Opus (Cont.)

Selected Target MR	Interestingness Results			UnComparable Results
	FCA2MR Max Level Degree of Match	FCA2MR Comparable Results	Magnum Opus Results	
Nursery_N_1 Domain (Min_Supp=1, Min_Conf=1) MR1: ({Convenient}, {}, 100) MR2: ({Problematic}, {}, 100) MR3: ({Incomplete}, {}, 100) MR4: ({Priority, Recommended}, {}, 100) MR5: ({Convenient, Priority, Recommended, Problematic, Usually, Incomplete}, {}, 100)	35	24	24	0
	21	15	15	0
	19	12	12	0
	34	24	24	0
	312	268	268	0
Nursery_2 Domain (Min_Supp=1, Min_Conf=5) MR6: ({Convenient }, {}, 100) MR7: ({Convenient, Critical}, {}, 100) MR8: ({Convenient}, {Convenient, Critical}, 24) MR9: ({More}, {Convenient, Priority, Critical, Problematic, Usual, More}, 3.70) MR10: ({Convenient, Priority, Critical, Problematic, Usual, More }, {Usual}, 100)	30	24	24	0
	32	27	27	0
	54	39	39	0
	29	13	13	0
	184	92	92	0
Nursery_3 Domain (Min_Supp=1, Min_Conf=10) MR11: ({Convenient }, {}, 100) MR12: ({Convenient }, {Convenient, Inconv}, 100) MR13: ({NoProb}, {}, 100) MR14: ({Completed}, {}, 100) MR15: ({Convenient}, {Convenient}, 100)	295	24	24	0
	190	24	24	0
	284	56	56	0
	354	78	78	0
	134	31	31	0

Table 6-16: The Summary Results of the Interestingness Comparisons with Magnum Opus (Cont.)

Selected Target MR	Interestingness Results			UnComparable Results
	FCA2MR Max Level Degree of Match	FCA2MR Comparable Results	Magnum Opus Results	
Adult_A_1 Domain (Min_Supp=1, Min_Conf=1) MR1: ({Bachelor}, {}, 100) MR2: ({Male}, {}, 100) MR3: ({Bachelor, Male}, {}, 100) MR4: ({Bachelor, Male, 17}, {}, 100) MR5: ({Bachelor, Male, 17}, {Bachelor}, 100)	3,576 9,120 1,980 2,830 4,216	136 15 34 16 37	136 15 34 16 37	0 0 0 0 0
Adult_A_2 Domain (Min_Supp=1, Min_Conf=5) MR6: ({Husband}, {}, 100) MR7: ({Never-Married}, {}, 100) MR8: ({Husband, Never-Married}, {}, 100) MR9: ({Husband}, {Never-Married }, 100) MR10: ({Some-College}, {Never-Married}, 100)	2,683 5,632 682 1,238 9,462	24 19 6 13 68	24 19 6 13 68	0 0 0 0 0
Adult_A_3 Domain (Min_Supp=1, Min_Conf=10) MR11: ({Male}, {}, 100) MR12: ({Female, Never-married}, {}, 100) MR13: ({Male}, {Master }, 100) MR14: ({Male}, {Some-College }, 100) MR15: ({Male, Never-Married }, {Some-College}, 100)	3,412 5,482 7,321 6,343 5,158	24 15 41 12 31	24 15 41 12 31	0 0 0 0 0

The interestingness evaluation shows the comparison results between the FCA2MR prototype system and the magnum opus program. This comparison results show that our interestingness results are comparable to the results of the commercial tool in all cases. In fact, the set of MR-patterns generated from our system are greater than the magnum opus because our system uses the combination element of MR-patterns to match the patterns while the magnum opus uses only single element match.

Moreover, our technique deals with the subjective interestingness, we designed the last evaluation method to carry out a number of questionnaires to check whether the ranking results do reflect people's intuitions of subjective interestingness. The details are described in the next section.

6.4.3. User Participation Evaluation

The last section is to evaluate the user participation with the results of the FCA2MR prototype system. We consider five user participation factors as follow:

- **System Interestingness (SI)** is the number of the interestingness results of the FCA2MR prototype system.
- **User Interestingness (UI)** is the number of the interestingness results that are classified as “Interestingness” and are actually interested by the user.
- **Relevance Interestingness (RI)** is the number of the interestingness results that are classified automatically from our prototype system and are actually interested by the user.
- **Recall of Interestingness (UI/RI)** is the proportion of relevant patterns that are actually retrieved in answering to a search request. The recall of interestingness can answer the user question [1], i.e., have user found all relevant information?

The equation for calculation the recall value is shown below:

$$\begin{aligned} \text{Recall("Interestingness")} &= \frac{\text{\#Results that are classified as Interestingness and are actually Interestingness}}{\text{\#Results that are classified automatically and are actually Interestingness}} \\ \text{Or} &= \frac{\text{User Interestingness (UI)}}{\text{Relevance Interestingness (RI)}} \end{aligned}$$

- **Precision of Interestingness (UI/SI)** is the proportion of retrieved patterns that are actually relevant to user needs. The precision of interestingness can answer the user question, i.e., how many interestingness results are really relevant to user needs?

The equation for calculation the precision value is shown below:

$$\begin{aligned} \text{Precision ("Interestingness")} &= \frac{\text{\#Results that are classified as Interestingness and are actually Interestingness}}{\text{\#Results that are classified as Interestingness}} \\ \text{Or} &= \frac{\text{User Interestingness (UI)}}{\text{System Interestingness (SI)}} \end{aligned}$$

We design the questionnaire to get the users interestingness. The questionnaires are composed of three data domains, namely, library domain, student domain, and short course training domain. Each data domain contains two major themes, and each theme contains ten questions. The details are described below.

6.4.3.1 Library Domain

We use the MUCC library data as the input data set of the first questionnaire. The total number of 1,165 records is collected from the active database currently used at the MUCC library. The questions of this questionnaire are described below.

THEME #1: Library_Q1 Domain

Assumption: Assume that you are the specialist in the computer science area. If you would like to learn the new technology that relevant to your career, which categories of book you will borrow from the MUCC library?

- Question#1-1** If your position is System Analyst (Position=SA), which book from the Database category you will borrow?
- Question#1-2** If your position is System Analyst (Position=SA), which book from the Programming category you will borrow?
- Question#1-3** If your position is Developer (Position=Developer), which book from the Database category you will borrow?
- Question#1-4** If your position is Developer (Position=Developer), which book from the Programming category you will borrow?
- Question#1-5** If your position is Staff (Position=Staff), which book from the WebDevelop category you will borrow?
- Question#1-6** If your position is Staff (Position=Staff), which book from the Database category you will borrow?
- Question#1-7** If your position is Staff (Position=Staff), which book from the Programming category you will borrow?
- Question#1-8** If your position is Staff (Position=Staff), which book from the MultiMedia category you will borrow?
- Question#1-9** If your position is Student (Position=Student), which book from the Computer Science category you will borrow?
- Question#1-10** If your position is Student (Position=Student), which book from the WebDevelop category you will borrow?

THEME #2: Library_Q2 Domain

Assumption: Assume that you have ever borrowed any kind of book from the MUCC library and you would like to search the relevant topic in this area continually. What kind of book you will borrow?

Question#2-1 If you have ever been borrowed the book from the Multimedia category (Subject=Multimedia), which book from the WebDevelop category you will borrow?

Question#2-2 If you have ever been borrowed the book from the Database category (Subject=DB), which book from the Programming category you will borrow?

Question#2-3 If you have ever been borrowed the book from the Database category (Subject=DB), which book from the Database category you will borrow?

Question#2-4 If you have ever been borrowed the book from the Database category (Subject=DB), which book from the Programming category you will borrow?

Question#2-5 If you have ever been borrowed the book from the Security category (Subject=Security), which book from the Computer Science category you will borrow?

Question#2-6 If you have ever been borrowed the book from the Database category (Subject=DB), which book from the Computer Science category you will borrow?

Question#2-7 If you have ever been borrowed the ASP book from the Programming category (Programming=ASP), which book from the Database category you will borrow?

Question#2-8 If you have ever been borrowed the Photoshop book from the Multimedia category (Multimedia=Photoshop), which book from the Programming category you will borrow?

Question#2-9 If you have ever been borrowed the SQL book from the Database category (Database=SQL), which book from the Programming category you will borrow?

Question#2-10 If you have ever been borrowed the Dreamweaver book from the MultiMedia category (Multimedia=Dreamweaver), which book from the MultiMedia category you will borrow?

The detail results of the first questionnaire are presented in Table 6-17 and Table 6-18.



Table 6-17: The Summary Results of the User Participation Evaluation of the Library_Q1 Domain

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-1					
IR1: Position=SA → Database=DBMS	10	10	10	100.00%	100.00%
IR2: Position=SA → Database=Access	58	56	58	96.55%	96.55%
IR3: Position=SA → Database=SQL	38	37	38	97.37%	97.37%
Total	106	103	106		
Question #1-2					
IR4: Position=SA → Programming=VB	28	28	28	100.00%	100.00%
IR5: Position=SA → Programming=ASP	29	29	29	100.00%	100.00%
IR6: Position=SA → Programming=Script	22	22	22	100.00%	100.00%
IR7: Position=SA → Programming=SQL	68	65	68	95.59%	95.59%
Total	147	144	147		
Question #1-3					
IR8: Position=Developer → Database=DBMS	10	10	10	100.00%	100.00%
IR9: Position=Developer → Database=Access	36	35	36	97.22%	97.22%
IR10: Position=Developer → Database=SQL	34	33	34	97.06%	97.06%
Total	80	78	80		
Question #1-4					
IR11: Position=Developer → Programming=VB	12	12	12	100.00%	100.00%
IR12: Position=Developer → Programming=ASP	8	8	8	100.00%	100.00%
IR13: Position=Developer → Programming=Script	12	12	12	100.00%	100.00%
IR14: Position=Developer → Programming=SQL	50	49	50	98.00%	98.00%
Total	82	81	82		

Table 6-17: The Summary Results of the User Participation Evaluation of the Library_Q1 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-5					
IR15: Position=Developer → WebDevelop=ASP	7	7	7	100.00%	100.00%
IR16: Position=Developer → WebDevelop=PHP	9	9	9	100.00%	100.00%
Total	16	16	16		
Question #1-6					
IR17: Position=Staff → Database=DBMS	6	6	6	100.00%	100.00%
IR18: Position=Staff → Database=Access	26	25	26	96.15%	96.15%
IR19: Position=Staff → Database=SQL	16	16	16	100.00%	100.00%
Total	48	47	48		
Question #1-7					
IR20: Position=Staff → Programming=VB	10	10	10	100.00%	100.00%
IR21: Position=Staff → Programming=ASP	10	10	10	100.00%	100.00%
IR22: Position=Staff → Programming=Script	3	2	3	66.67%	66.67%
IR23: Position=Staff → Programming=SQL	31	29	31	93.55%	93.55%
Total	54	51	54		
Question #1-8					
IR24: Position=Developer → Programming=VB	17	17	17	100.00%	100.00%
IR25: Position=Developer → Programming=ASP	29	29	29	100.00%	100.00%
IR26: Position=Developer → Programming=Script	15	15	15	100.00%	100.00%
Total	61	61	61		

Table 6-17: The Summary Results of the User Participation Evaluation of the Library_Q1 Domain (Cont.)

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-9					
IR27: Position=Student → ComScience=DB	12	12	12	100.00%	100.00%
IR28: Position=Student → ComScience=Network	10	10	10	100.00%	100.00%
IR29: Position=Student → ComScience=Security	22	22	22	100.00%	100.00%
IR30: Position=Student → ComScience=DataStruct	5	5	5	100.00%	100.00%
Total	49	49	49		
Question #1-10					
IR31: Position=Student → WebDevelop=ASP	23	23	23	100.00%	100.00%
IR32: Position=Student → WebDevelop=PHP	11	10	11	90.91%	
Total	34	33	34		

Table 6-18: The Summary Results of the User Participation Evaluation of the Library_Q2 DomainTotal number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #2-1 IR1: Subject=MultiMedia → WebDevelop=ASP IR2: Subject=MultiMedia → WebDevelop=PHP Total	35 26 61	35 26 61	35 26 61	100.00% 100.00%	100.00% 100.00%
Question #2-2 IR3: Subject=DB → Programming=VB IR4: Subject=DB → Programming=ASP IR5: Subject=DB → Programming=Script IR5: Subject=DB → Programming=SQL Total	29 23 21 88 161	29 23 21 85 158	29 23 21 88 161	100.00% 100.00% 100.00% 96.59%	100.00% 100.00% 100.00% 96.59%
Question #2-3 IR6: Subject=DB → Database=DBMS IR7: Subject=DB → Database=Access IR8: Subject=DB → Database=SQL Total	15 83 58 156	15 81 57 153	15 83 58 156	100.00% 97.59% 98.28%	100.00% 97.59% 98.28%
Question #2-4 IR9: Subject=Security → Programming=VB IR10: Subject=Security → Programming=ASP IR11: Subject=Security → Programming=Script IR12: Subject=Security → Programming=SQL Total	2 6 20 5 33	2 6 19 4 31	2 6 21 6 35	100.00% 100.00% 90.48% 66.67%	100.00% 100.00% 95.00% 80.00%

Table 6-18: The Summary Results of the User Participation Evaluation of the Library_Q2 Domain (Cont.)Total number of questionnaire ($N = 100$, $Min_Supp=0.1$, $Min_Conf=0.1$)

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
Question #2-5						
IR13: Subject=Security → ComScience=DB	2	2	2		100.00%	100.00%
IR14: Subject=Security → ComScience=Network	5	5	5		100.00%	100.00%
IR15: Subject=Security → ComScience=Security	24	23	24		95.83%	95.83%
IR16: Subject=Security → ComScience=DataStruct	12	12	12		100.00%	100.00%
Total	43	42	43			
Question #2-6						
IR17: Subject=DB → ComScience=DB	51	48	51		94.12%	94.12%
IR18: Subject=DB → ComScience=Network	9	9	9		100.00%	100.00%
IR19: Subject=DB → ComScience=Security	42	41	42		97.62%	97.62%
IR20: Subject=DB → ComScience=DataStruct	10	9	10		90.00%	90.00%
Total	112	107	112			
Question #2-7						
IR21: Programming=ASP → Database=DBMS	21	20	23		86.96%	95.24%
IR22: Programming=ASP → Database=Access	36	35	36		97.22%	97.22%
IR23: Programming=ASP → Database=SQL	19	19	19		100.00%	100.00%
Total	76	74	79			
Question #2-8						
IR24: MultiMedia=Photoshop → Programming=VB	8	8	8		100.00%	100.00%
IR25: MultiMedia=Photoshop → Programming=ASP	21	21	21		100.00%	100.00%
IR26: MultiMedia=Photoshop → Programming=Script	15	15	15		100.00%	100.00%
IR27: MultiMedia=Photoshop → Programming=SQL	42	42	42		100.00%	100.00%
Total	86	86	86			

Table 6-18: The Summary Results of the User Participation Evaluation of the Library_Q2 Domain (Cont.)Total of questionnaire $N = 100$, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #2-9					
IR28: Database=SQL → Programming=VB	37	37	37	100.00%	100.00%
IR29: Database=SQL → Programming=ASP	19	19	19	100.00%	100.00%
IR30: Database=SQL → Programming=Script	10	10	10	100.00%	100.00%
IR31: Database=SQL → Programming=SQL	74	73	74	98.65%	98.65%
Total	140	139	140		
Question #2-10					
IR32: MultiMedia=Dream → MultiMedia=Photoshop	21	21	21	100.00%	100.00%
IR33: MultiMedia=Dream → MultiMedia=Flash	32	32	32	100.00%	100.00%
IR34: MultiMedia=Dream → MultiMedia=Dream	3	3	3	100.00%	100.00%
Total	56	56	56		

6.4.3.2 Student Domain

For the second questionnaire, we use the grade report of the ICT student at the Mahidol University as the input data set. The total number of 70 data records is collected from the grade report of the second year ICT student. List of questions of this questionnaire are presented below.

THEME #1: Student_Q1 Domain

Assumption: Assume that after you finished up the examination of this semester, how about the grade result that you will earn?

- Question#1-1** If you can make the maximum score of the Midterm examination (Midterm=Max), what is the level of final examination that you can make?
- Question#1-2** If you can make the medium score of the Midterm examination (Midterm=Mid), what is the level of final examination that you can make?
- Question#1-3** If you can make the minimum score of the Midterm examination (Midterm=Min), what is the level of final examination that you can make?
- Question#1-4** If you can get grade A (Grade=A), what is the level of final examination that you can make?
- Question#1-5** If you can get grade B (Grade=B), what is the level of final examination that you can make?
- Question#1-6** If you can get grade C (Grade=C), what is the level of final examination that you can make?
- Question#1-7** If you can get grade D (Grade=D), what is the level of final examination that you can make?
- Question#1-8** If you can make the Maximum score of the Final examination (Final=Max), what is the grade that you can get from this course?
- Question#1-9** If you can make the Medium score of the Final examination (Final=Mid), what is the grade that you can get from this course?
- Question#1-10** If you can make the Minimum score of the Final examination (Final=Min), what is the grade that you can get from this course?

THEME #2: Student_Q2 Domain

Assumption: Assume that if you would like to get a particular grade defined in the given condition, what is the score you should earn in each proportion?

- Question#2-1** If you would like to get grade A (Grade=A), what is the level of Final examination that you can make?
- Question#2-2** If you would like to get grade A (Grade=A), what is the level of Midterm examination that you can make?
- Question#2-3** If you would like to get grade B (Grade=B), what is the level of Final examination that you can make?
- Question#2-4** If you would like to get grade B (Grade=B), what is the level of Midterm examination that you can make?
- Question#2-5** If you would like to get grade A (Grade=A) but you can make the Minimum Score of the Midterm examination (Midterm=Min), what is the level of Final examination that you can make?
- Question#2-6** If you would like to get grade A (Grade=A) but you can make the Medium Score of the Term-Project (Project=Mid), what is the level of Final examination that you can make?
- Question#2-7** If you would like to get grade A (Grade=A) but your Attendant level is average (Attendant=Average), what is the level of Final examination that you can make?
- Question#2-8** If you would like to get grade A (Grade=A) but you can make the Medium Score of the Midterm examination (Midterm=Mid) and Attendant level is Average (Attendant=Average), what is the level of Final examination that you can make?
- Question#2-9** If you would like to get grade B (Grade=B) but you can make the Maximum Score of the Midterm examination (Midterm=Max) and Attendant level is Full (Attendant=Full), what is the level of Final examination that you can make?

Question#2-10 If you would like to get grade A (Grade=A) but you can make the Maximum Score of the Midterm examination (Midterm=Max) and Term-Project is Minimum (Project=Min), what is the level of Final examination that you can make?

The detail results of the student questionnaire are presented in Table 6-19 and Table 6-20.



Table 6-19: The Summary Results of the User Participation Evaluation of the Student_Q1 Domain

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-1 IR1: Midterm=Max → Final=Max IR2: Midterm=Max → Final=Middle IR3: Midterm=Max → Final=Min	18	18	18	100.00%	100.00%
	2	2	2	100.00%	100.00%
	8	8	8	100.00%	100.00%
	28	28	28		
Question #1-2 IR4: Midterm=Middle → Final=Max IR5: Midterm=Middle → Final=Middle IR6: Midterm=Middle → Final=Min	26	26	26	100.00%	100.00%
	24	24	24	100.00%	100.00%
	18	18	18	100.00%	100.00%
	68	68	68		
Question #1-3 IR7: Midterm=Min → Final=Max IR8: Midterm=Min → Final=Mid IR9: Midterm=Min → Final=Min	8	8	8	100.00%	100.00%
	32	32	32	100.00%	100.00%
	34	34	34	100.00%	100.00%
	74	74	74		
Question #1-4 IR10: Grade=A → Final=Max IR11: Grade=A → Final=Middle IR12: Grade=A → Final=Min	34	34	34	100.00%	100.00%
	5	5	5	100.00%	100.00%
	0	0	0	0%	0%
	39	39	39		

Table 6-19: The Summary Results of the User Participation Evaluation of the Student_Q1 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
Question #1-5 IR10: Grade=B → Final=Max IR11: Grade=B → Final=Middle IR12: Grade=B → Final=Min	12 34 27 73	12 34 27 73	12 34 27 73		100.00% 100.00% 100.00%	
Total	73	73	73			
Question #1-6 IR13: Grade=C → Final=Max IR14: Grade=C → Final=Middle IR15: Grade=C → Final=Min	0 20 24 44	0 20 24 44	0 20 24 44		0% 100.00% 100.00%	
Total	44	44	44			
Question #1-7 IR16: Grade=D → Final=Max IR17: Grade=D → Final=Middle IR18: Grade=D → Final=Min	0 0 12 12	0 0 12 12	0 0 12 12		0% 0% 100.00%	
Total	12	12	12			
Question #1-8 IR19: Final=Max → Grade=A IR20: Final=Max → Grade=B IR21: Final=Max → Grade=C IR22: Final=Max → Grade=D	34 12 0 0 46	34 12 0 0 46	34 12 0 0 46		100.00% 100.00% 0% 0%	
Total	46	46	46			

Table 6-19: The Summary Results of the User Participation Evaluation of the Student_Q1 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
Question #1-9 IR23: Final=Middle → Grade=A IR24: Final=Middle → Grade=B IR25: Final=Middle → Grade=C IR26: Final=Middle → Grade=D	0	0	0	0	0%	0%
	34	34	34		100.00%	100.00%
	20	20	20		100.00%	100.00%
	0	0	0		0%	0%
	Total 54	54	54			
Question #1-10 IR27: Final=Min → Grade=A IR28: Final=Min → Grade=B IR29: Final=Min → Grade=C IR30: Final=Min → Grade=D	0	0	0	0	0%	0%
	27	27	27		100.00%	100.00%
	24	24	24		100.00%	100.00%
	12	12	12		100.00%	100.00%
	Total 51	51	51			

Table 6-20: The Summary Results of the User Participation Evaluation of the Student_Q2 Domain

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #2-1 IR1: Grade=A → Final=Max IR2: Grade=A → Final=Middle IR3: Grade=A → Final=Min	34 2 0 36	34 2 0 36	34 2 0 36	100.00% 100.00% 0%	100.00% 100.00% 0%
Total	36	36	36		
Question #2-2 IR4: Grade=A → Midterm=Max IR5: Grade=A → Midterm =Middle IR6: Grade=A → Midterm =Min	18 20 8 46	18 20 8 46	18 20 8 46	100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	46	46	46		
Question #2-3 IR7: Grade=B → Final=Max IR8: Grade=B → Final=Middle IR9: Grade=B → Final=Min	12 34 27 73	12 34 27 73	12 34 27 73	100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	73	73	73		
Question #2-4 IR10: Grade=B → Midterm =Max IR11: Grade=B → Midterm =Middle IR12: Grade=B → Midterm =Min	9 39 24 72	9 39 24 72	9 39 24 72	100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	72	72	72		

Table 6-20: The Summary Results of the User Participation Evaluation of the Student_Q2 Domain (Cont.)

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
Question #2-5 IR13: Midterm=Min → Final=Max IR14: Midterm=Min → Final=Middle IR15: Midterm=Min → Final=Min	8 32 31 71	8 32 31 71	8 32 31 71		100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	71	71	71			
Question #2-6 IR16: Project=Mid → Final=Max IR17: Project=Mid → Final=Middle IR18: Project=Mid → Final=Min	26 24 17 67	26 24 17 67	26 24 17 67		100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	67	67	67			
Question #2-7 IR19: Attendant=Average → Final=Max IR20: Attendant=Average → Final=Middle IR21: Attendant=Average → Final=Min	18 20 14 52	18 20 14 52	18 20 14 52		100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	52	52	52			
Question #2-8 IR22: Midterm=Mid and Attendant=Average → Final =Max IR23: Midterm=Mid and Attendant=Average → Final =Middle IR24: Midterm=Mid and Attendant=Average → Final =Min	45 44 32 121	45 44 32 121	45 44 32 121		100.00% 100.00% 100.00%	100.00% 100.00% 100.00%
Total	121	121	121			

Table 6-20: The Summary Results of the User Participation Evaluation of the Student_Q2 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #2-9 IR25: Midterm=Max and Attendant=Full → Final =Max IR26: Midterm=Max and Attendant=Full → Final =Middle IR27: Midterm=Max and Attendant=Full → Final =Min	46	46	46	100.00%	100.00%
	32	32	32	100.00%	100.00%
	43	43	43	100.00%	100.00%
	121	121	121		
Total					
Question #2-10 IR28: Midterm=Max and Project=Min → Final =Max IR29: Midterm=Max and Project=Min → Final =Middle IR30: Midterm=Max and Project=Min → Final =Min	34	34	34	100.00%	100.00%
	24	24	24	100.00%	100.00%
	30	30	30	100.00%	100.00%
	88	88	88		
Total					

6.4.3.3 Short Course Training Domain

For the last questionnaire, we use the short course training data at the Mahidol University Computing Center (MUCC) as the input data set. The total number of 315 data records is collected from the user who enrolled the short course training in the year 2005-2006. The questions of this questionnaire are shown below.

THEME #1: ShortCourse_Q1 Domain

Assumption: Assume that if you would like to enroll the short course training defined in the given condition, what is the course you would like to enroll in each proportion?

- Question#1-1** If you have ever enrolled the Flash MX course (MultiMedia=FlashMX), what is the MultiMedia course you would like to enroll continually?
- Question#1-2** If you have ever enrolled the Flash MX course (MultiMedia=FlashMX), what is the Web Publishing course you would like to enroll continually?
- Question#1-3** If you have ever enrolled the ASP course (WebPublishing=ASP), what is the MultiMedia course you would like to enroll continually?
- Question#1-4** If you have ever enrolled the ASP course (WebPublishing=ASP), what is the Web Publishing course you would like to enroll continually?
- Question#1-5** If you have ever enrolled the MS-Access course (MS-Office=Access), what is the Web Publishing course you would like to enroll continually?
- Question#1-6** If you have ever enrolled the MS-Access course (MS-Office=Access), what is the Programming course you would like to enroll continually?
- Question#1-7** If you have ever enrolled the Photoshop course (MultiMedia=Photoshop), what is the Web Publishing course you would like to enroll continually?
- Question#1-8** If you have ever enrolled the VB course (Programming=VB), what is the Web Publishing course you would like to enroll continually?
- Question#1-9** If you have ever enrolled the VB course (Programming=VB), what is the MS-Office course you would like to enroll continually?
- Question#1-10** If you have ever enrolled the Frontpage course (WebPublishing=Frontpage), what is the Programming course you would like to enroll continually?

THEME #2: ShortCourse_Q2 Domain

Assumption: Assume that if you finished up the course defined in the given condition, what is the course you would like to enroll in each proportion?

- Question#2-1** If you have ever enrolled the MultiMedia course, what is the course you would like to enroll continually?
- Question#2-2** If you have ever enrolled the Web Publishing course, what is the course you would like to enroll continually?
- Question#2-3** If you have ever enrolled the Programming course, what is the course you would like to enroll continually?
- Question#2-4** If you have ever enrolled the MS-Office course, what is the course you would like to enroll continually?
- Question#2-5** If you have ever enrolled the Photoshop (Multimedia=Photoshop) and Programming course, what is the course you would like to enroll continually?
- Question#2-6** If you have ever enrolled the MS-Access (MS-Office=Access) and Programming course, what is the course you would like to enroll continually?
- Question#2-7** If you have ever enrolled the FlashMX (Multimedia=FlashMX) and Programming course, what is the course you would like to enroll continually?
- Question#2-8** If you have ever enrolled the Photoshop (Multimedia=Photoshop) and Web Publishing course, what is the course you would like to enroll continually?
- Question#2-9** If you have ever enrolled the Courseware (Multimedia=Courseware) and Programming course, what is the course you would like to enroll continually?
- Question#2-10** If you have ever enrolled the VB (Programming=VB) and MS-Access course (MS-Office=Access), what is the course you would like to enroll continually?

The detail results of the short course questionnaire are presented in Table 6-21 and Table 6-22.



Table 6-21: The Summary Results of the User Participation Evaluation of the ShortCourse_C1 Domain

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-1					
IR1: Multimedia=FlashMX → Multimedia=Photoshop	11	11	11	100.00%	100.00%
IR2: Multimedia=FlashMX → Multimedia=Courseware	14	14	14	100.00%	100.00%
IR3: Multimedia=FlashMX → Multimedia=FlashMX	0	0	0	0%	0%
Total	25	25	25		
Question #1-2					
IR4: Multimedia=FlashMX → WebPublishing=Frontpage	0	0	0	0%	0%
IR5: Multimedia=FlashMX → WebPublishing=Dreamweaver	0	0	0	0%	0%
IR6: Multimedia=FlashMX → WebPublishing=ASP	10	10	10	100.00%	100.00%
Total	16	16	16		
Question #1-3					
IR7: WebPublishing=ASP → Multimedia=Photoshop	8	8	8	100.00%	100.00%
IR8: WebPublishing=ASP → Multimedia=Courseware	6	6	6	100.00%	100.00%
IR9: WebPublishing=ASP → Multimedia=FlashMX	0	0	0	0%	0%
Total	14	14	14		
Question #1-4					
IR7: WebPublishing=ASP → WebPublishing=Frontpage	4	4	4	100.00%	100.00%
IR8: WebPublishing=ASP → WebPublishing=Dreamweaver	0	0	0	0%	0%
IR9: WebPublishing=ASP → WebPublishing=ASP	0	0	0	0%	0%
Total	16	16	16		
Question #1-5					
IR10: MS-Office=Access → WebPublishing=Frontpage	7	7	7	100.00%	100.00%
IR11: MS-Office=Access → WebPublishing=Dreamweaver	0	0	0	0%	0%
IR12: MS-Office=Access → WebPublishing=ASP	11	11	11	100.00%	100.00%
Total	18	18	18		

Table 6-21: The Summary Results of the User Participation Evaluation of the ShortCourse_C1 Domain (Cont.)

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns		Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
		System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #1-6						
IR13: MS-Office=Access → Programming=VB		16	16	16	100.00%	100.00%
IR14: MS-Office=Access → Programming=ASP		11	11	11	100.00%	100.00%
Total		27	27	27		
Question #1-7						
IR15: Multimedia=Photoshop → WebPublishing=Frontpage		0	0	0	0%	0%
IR16: Multimedia=Photoshop → WebPublishing=Dreamweaver		0	0	0	0%	0%
IR17: Multimedia=Photoshop → WebPublishing=ASP		8	8	8	100.00%	100.00%
Total		8	8	8		
Question #1-8						
IR18: Programming=VB → WebPublishing=Frontpage		6	6	6	100.00%	100.00%
IR19: Programming=VB → WebPublishing=Dreamweaver		0	0	0	0%	0%
IR20: Programming=VB → WebPublishing=ASP		12	12	12	100.00%	100.00%
Total		18	18	18		
Question #1-9						
IR21: Programming=VB → MS-Office=Excel		0	0	0	0%	0%
IR22: Programming=VB → MS-Office=Powerpoint		4	4	4	100.00%	100.00%
IR23: Programming=VB → MS-Office=Access		12	12	12	100.00%	100.00%
Total		16	16	16		
Question #1-10						
IR24: WebPublishing=Frontpage → Programming=VB		4	4	4	100.00%	100.00%
IR25: WebPublishing=Frontpage → Programming=ASP		12	12	12	100.00%	100.00%
Total		16	16	16		

Table 6-22: The Summary Results of the User Participation Evaluation of the ShortCourse_C2 Domain

Total number of questionnaire (N) = 100, Min_Supp=0.1, Min_Conf=0.1

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
Question #2-1						
IR1: Multimedia → WebPublishing=Frontpage	6	6	6		100.00%	100.00%
IR2: Multimedia → WebPublishing=Dreamweaver	0	0	0		0%	0%
IR3: Multimedia → WebPublishing=ASP	8	8	8		100.00%	100.00%
Total	14	14	14			
Question #2-2						
IR4: WebPublishing → Multimedia=Photoshop	8	8	8		100.00%	100.00%
IR5: WebPublishing → Multimedia=Courseware	0	0	0		0%	0%
IR6: WebPublishing → Multimedia=FlashMX	6	6	6		100.00%	100.00%
Total	14	14	14			
Question #2-3						
IR7: Programming → WebPublishing=Frontpage	6	6	6		100.00%	100.00%
IR8: Programming → WebPublishing=Dreamweaver	0	0	0		0%	0%
IR9: Programming → WebPublishing=ASP	12	12	12		100.00%	100.00%
Total	18	18	18			
Question #2-4						
IR10: MS-Office → Programming=VB	16	16	16		100.00%	100.00%
IR11: MS-Office → Programming=ASP	11	11	11		100.00%	100.00%
Total	27	27	27			
Question #2-5						
IR12: Multimedia=Photoshop and Programming → WebPublishing=Frontpage	6	6	6		100.00%	100.00%
IR13: Multimedia=Photoshop and Programming → WebPublishing=Dreamweaver	0	0	0		0%	0%
IR14: Multimedia=Photoshop and Programming → WebPublishing=ASP	20	20	20		100.00%	100.00%
Total	26	26	26			

Table 6-22: The Summary Results of the User Participation Evaluation of the ShortCourse_C2 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results			Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)		
Question #2-6 IR15: MS-Office=Access and Programming → MS-Office=Excel IR16: MS-Office=Access and Programming → MS-Office=Powerpoint IR17: MS-Office=Access and Programming → MS-Office=Access	0 4 12 16	0 4 12 16	0 4 12 16	0% 100.00% 100.00%	0% 100.00% 100.00%
Question #2-7 IR18: Multimedia=FlashMX and Programming → WebPublishing=Frontpage IR19: Multimedia=FlashMX and Programming → WebPublishing=Dreamweaver IR20: Multimedia=FlashMX and Programming → WebPublishing=ASP	6 0 12 18	6 0 12 18	6 0 12 18	100.00% 0% 100.00%	100.00% 0% 100.00%
Question #2-8 IR21: Multimedia=Photoshop and WebPublishing → Programming=VB IR22: Multimedia=Photoshop and WebPublishing → Programming=ASP	18 12 30	18 12 30	18 12 30	100.00% 100.00%	100.00% 100.00%
Question #2-9 IR23: Multimedia=Courseware and Programming → WebPublishing=Frontpage IR24: Multimedia=Courseware and Programming → WebPublishing=Dreamweaver	8 0	8 0	8 0	100.00% 0%	100.00% 0%

Table 6-22: The Summary Results of the User Participation Evaluation of the ShortCourse_C2 Domain (Cont.)

Total number of questionnaire (N) = 100, $Min_Supp=0.1$, $Min_Conf=0.1$

Interestingness Patterns	Interestingness Results				Recall (UI/RI)	Precision (UI/SI)
	System Int. (SI)	User Int. (UI)	Relevance Int. (RI)			
IR25: Multimedia=Courseware and Programming → WebPublishing=ASP	12	12	12		100.00%	100.00%
Total	20	20	20			
Question #2-10						
IR26: Programming=VB and MS-Office=Access → WebPublishing=Frontpage	6	6	6		100.00%	100.00%
IR27: Programming=VB and MS-Office=Access → WebPublishing=Dreamweaver	0	0	0		0%	0%
IR28: Programming=VB and MS-Office=Access → WebPublishing=ASP	23	23	23		100.00%	100.00%
Total	29	29	29			

In the last set of the experiment, three data domains of the questionnaires were evaluated, namely, library domain, ICT student domain and the short course training domain. Each data domain was collected from the 100 questionnaires. The summary results are displayed in Table 6-17 to 6-22.

In practical, the statistical measurement is a sound and standard technique for satisfying and determining the quantitative and qualitative measurement of the data. In this work, the recall and precision measurements are used to indicate the degree of interestingness results generated from the system. The high recall values indicate that the user is satisfied with the high relevant information. While the high precision values show the degree of accuracy of the interestingness results which are really relevant to user's needs.

In the library domain, some test cases give low recall/precision values where the min_supp is 0.1% and min_conf is 0.1% because the library dataset contains low proportion to support the implication process. However, the results show an average performance indicating the recall and precision criteria are acceptable by the users.

In the student domain and short course domain, 29 test cases give zero value of the recall/precision because there are no patterns contained in this dataset. Therefore, most of the patterns give the highest degree of recall/precision. This implies that the system gives more meaningful to obtain the high accuracy value.

In summary, the results of the user participation evaluation are reasonable enough to conclude that the proposed system gives the high accuracy and the interestingness results are relevant to user's needs.

CHAPTER 7

DISCUSSION AND CONCLUSION

7.1 Discussion and Conclusion

The interestingness approach of data mining is arisen to support the users in the real-life usage, not only to reduce the complexity of the post processing of the data mining application, but also to support the user convenience in analyzing the final results of the system. There are two most applicable ways to define the interestingness patterns of data mining, namely, the objective interestingness, and the subjective interestingness.

Our approach focuses on the subjective interestingness, especially the actionable criteria. This implies that the proposed approach is used to support the user to purify the discovered patterns of the data mining, in the way that user's needs as the main criterion. We propose a novel framework to construct the meta-rules guided mining by applying the ontology. The proposed framework is an "Extended Version" of the traditional data mining. The work focused on the post analysis of the data mining. The proposed framework consists of two main parts, namely, the specialized rules discovery part, and the generalized rules discovery part.

- o First, the specialized rules discovery part is basically designed for data mining process in order to generate the discovered patterns. In contrast to the traditional data mining application, this part is generated once for discovering all the patterns that might be seen in the database.
- o Second, the generalized rules discovery part is used to discover the "Generalized Rules" defined in the generalization task of the proposed framework which is used to construct the meta-rules patterns. We design the mechanism to produce the meta-rules patterns based on the ontology. In this research, we adopt the Formal Concept Analysis (FCA) technique to construct the ontology and represent in the graphical representation style.

For the implementation part, we developed the prototype system called FCA2MR (Formal Concept Analysis to Meta-Rules) to serve as an easy-to-use and step-by-step process generation tool. The prototype system was developed to test and prove the proposed approach as summarized in the chapter of experiments and results. From the experiments, three benchmark datasets from the UCI machine learning repository were used to test the prototype system and three active datasets from the Mahidol University were used to evaluate the interestingness results. The test and evaluation of experimental results indicate that our approach is feasible to implement and serve as a reasonable solution for subjective interestingness data mining tool comparable to the commercial tools and acceptable from the users.

In experimental evaluation, as total of 45 experiments from nine data models for the interestingness evaluation and 300 questionnaires from three problem domains for the user participation evaluation have been evaluated and shown in Session 6.4.2 and 6.4.3. We analyzed these results to identify the accuracy of the proposed system as described below:

❑ The Interestingness Comparison between FCA2MR and the Commercial Tool (Magnum Opus Program)

The interestingness evaluation was done by comparing the results between the FCA2MR and Magnum Opus program. This comparison results reveal that our interestingness results are comparable to the results of the commercial tool in all cases.

❑ The User Participation Evaluation

The user participation evaluation was done by checking the user's opinion collected from the questionnaires to determine whether the ranking results reflect the subjective interestingness of the users. The results indicate that our proposed system provides satisfactory performance in both recall and precision criteria.

First, in order to determine the recall criterion of the FCA2MR, the high recall values indicate that user can find all relevant information corresponding to user's needs.

Second, the precision criterion is used to evaluate the accuracy of the FCA2MR system to obtain the interestingness results which are relevant to user's needs.

In the library domain, it gives low recall/precision values for a few test cases because the library dataset contains low proportion to support the implication process. However, the results show an average performance indicating the recall/precision criteria are acceptable by the users.

In the student and short course domain, 29 test cases give zero value of the recall/precision because there are no patterns contained in this dataset. Therefore, most of the patterns give the highest degree of recall/precision. This implies that the system gives more meaningful to obtain the high accuracy value.

In conclusion, we raise the following issues as the main contributions of our proposed approach:

1. The proposed framework, a unified framework of Meta-Rules guided mining based on ontology is a generic framework that is applicable for various application domains.
2. Our approach stimulates the use of new technology in the data mining research area by applying the Ontology and the Formal Concept Analysis (FCA) technique. These techniques are found to be applicable and efficient for knowledge representation in data mining. Moreover, the meta-rules guided mining is the recent technique to produce the interestingness of data mining.
3. FCA2MR provides the user friendly interface that allows user to analyze the data mining process as the step-by-step fashion. Moreover, the user can preview the ontology result in the graphical representation style. As a result, the prototype of FCA2MR serves as the realization of our framework.
4. The benefits of the outstanding feature of the FCA technique can serve the process of ontology construction automatically. This feature also supports the loss-less of information because every binary relationship between certain objects always has the relation with certain attributes. Furthermore, for every binary relation, a lattice is constructed.

7.2 Future Works

The following are interesting suggestions for future research.

1. The improvement of the ontology generator performance

The FCA2MR ontology generator is the process that adopts the mathematical technique to generate the ontology, it handles the objects/attributes extracted from the database and generates the cross-table which is used to represent the formal context. Thereafter, the formal concept will be generated to form the set of concepts of the ontology. Aftermost, the formal lattice is generated and represented in terms of the taxonomy. Obviously, the performance of the process depends on the number of objects/attributes. The processing time is $2^n * 2^n$. Hence, the algorithm for ontology generator should be modified to support the parallel processing more efficiently.

2. The representation technique of the interestingness result

The visualization technique is more interactive and attractive method to support the users to analyze the interestingness patterns.

3. The extension of our representation style of the meta-rules patterns

The integration of various knowledge representation techniques, such as, frame based, model based, case-based method, etc., are the real challenge for FCA2MR. At the practical level, the system should incorporate various knowledge representation techniques so that the system will be more flexible and applicable.

REFERENCES

- [1] T.Aydin and H.A.Guvenir, “**Interactive Rule Interestingness Learning**”, Technical Report, Faculty of Engineering, Bilkent University, BU-CE-040, 2004, pages 1-24.
- [2] C.L.Blake, and C.J. Merz., “**UCI Repository of Machine Learning Database**”, Irvine, CA: University of California, Department of Information and Computer Science, [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]
- [3] C. Borgelt, and R. Kruse, “**Induction of Association Rules: Apriori Implementation**”, Technical Report, Department of Knowledge Processing and Language Engineering, School of Computer Science, Otto-von-Guericke-University of Magdeburg, D-39106, Germany, pages 1-6.
- [4] P.Cabena et al. “**Discovering Data Mining from Concept to Implementation**”, Prentice Hall, NJ, pp.42, 1998.
- [5] J.Cerquides, R.L.Mantaras, “**Fuzzy Metaqueries for Guiding the Discovery Process in KDD**”, Proceedings of the sixth IEEE International Conference on Fuzzy Systems, 1997, Vol.3, 1-5 July 1997, pp.1555-1559.
- [6] G. Chen, “**Influence and Conditional Influence-New Interestingness Measures in Association Rule Mining**”, IEEE International Fuzzy Systems Conference, 2001, Page(s) 1440-1443.
- [7] Ming-Syan Chen, Jiawei Han, Philip S. Yu, “**Data Mining : An Overview from a Database Perspective**”, IEEE Transactions on Knowledge and Data Engineering, Vol.8, No.6, December 1996, pp.866-883.
- [8] R. Cole, “**Automated Layout of Concept Lattices Using Layered Diagrams and Additive Diagrams**”, IEEE Computer Society Press, Los Alamitos, CA, USA, 2001, pp.47-53.
- [9] Darmstadt University of Technology, Available on <http://www.mathematik.tu-darmstadt.de/>

- [10] U.M. Fayyad, G. Piatetsky-Shapiro, P.Smyth, "**Knowledge Discovery and Data Mining : Towards a Unifying Framework**", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining(KDD-96), Portland, Oregon, August 2-4, 1996, AAAI Press.
- [11] M. Fernandez, A. Gomez-Perez, and N. Juristo, "**Building a Chemical Ontology using Methontology and the Ontology Design Environment**", IEEE journal, January/February 1999, pp. 37-46.
- [12] N.Guarino, "**Formal Ontology and Information System**", Proc. Of FOIS'98, Ternto, Italy, 1998, pp.3-15.
- [13] I.Jurisica, J.Mylopoulos & E.Yu, "**Using Ontologies for Knowledge Management: An Information Systems Perspective**", Annual Conference of the American Society for Information Science, Washington, D.C., November 1-4, 1999.
- [14] M. Kamber, J. Han, J.Y.Chiang, "**Metarule-Guided Mining of Multi-Dimensional Association Rules Using Data Cubes**", Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA., 1997, pp.207-210.
- [15] M. Kamber, J. Han, J.Y.Chiang, "**Using Data Cubes for Meta-Rules Mining of Multi-Dimensional Association Rules**", Technical Report CMPT-TR-97-10, School of Computing Science, Simon Fraser University, May 1997.
- [16] R.E. Kent, and J. Brady, "**Formal Concept Analysis with Many-sorted Attributes**", IEEE, 1993, pp. 349-353.
- [17] L.Khan and F.Luo, "**Ontology Construction for Information Selection**", Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02), 2002
- [18] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo, "**Finding Interesting Rules from Large Sets of Discovered Association Rules**", Proceedings of the Third International Conference on Information and Knowledge Management, December 1994, pp. 970-974

- [19] J. Li and G. Chen, "**A SAR-Based Interesting Rule Mining Algorithm**", Proceedings of Fuzzy Information Processing Society, NAFIPS, 2002. Annual Meeting of the North American, 2002, Page(s) 178-183.
- [20] Bing Liu et al., "**Analyzing the Subjective Interestingness of Association Rules**", IEEE Journal, Vol.15, Issue 5, September/October 2000, pp. 47-55.
- [21] Bing Liu, et al., "**Finding Interesting Patterns using User Expectations**", IEEE Transactions on Knowledge and Data Engineering, Vol.11, No.6, November/December 1999, pp.817-832.
- [22] Bing Liu, et al., "**Identifying Interesting Missing Patterns**", DISCS Technical Report, the National University of Singapore, TRA8/96, August 1996, pp. 1-12.
- [23] H. Mannila, "**Methods and Problems in Data Mining**", Proceedings of International Conference on Database Theory, Delphi, Greece, January 1997.
- [24] S.Mitra, S.K.Pal, P.Mitra, "**Data Mining in Soft Computing Framework**", IEEE Transactions on Neural Networks, Vol.13, No.1, January 2002, pp.3-14
- [25] J.L.Mitrpanont and P.Tangworakitthaworn, "**Towards A Unified Framework of Meta-Rules Guided Mining Based on Ontology**", Proceedings of the Seven National Computer Science and Engineering Conferences (NCSEC2003), Chonburi, Thailand, October 28-30 , 2003, pp.157-162.
- [26] J.L.Mitrpanont and P.Tangworakitthaworn, "**The Development of a System for Mining the Subjective Interestingness Patterns**", Joint Conference on Computer Science and Software Engineering (JCSSE2005), Chonburi, Thailand, November 17-18 , 2005, pp.211-216.
- [27] Amit A. Nanavati et al., "**Mining Generalized Disjunction Association Rules**", Proceeding of CIKM'01, November 5-10, 2001, Atlanta, Georgia, USA. Pp.482-489.
- [28] J. Phillips, B.G. Buchanan, "**Ontology-Guided Knowledge Discovery in Database**", First International Conference on Knowledge Capture (K-CAP'01), Victoria, BC, Canada; 21-23 October 2001.

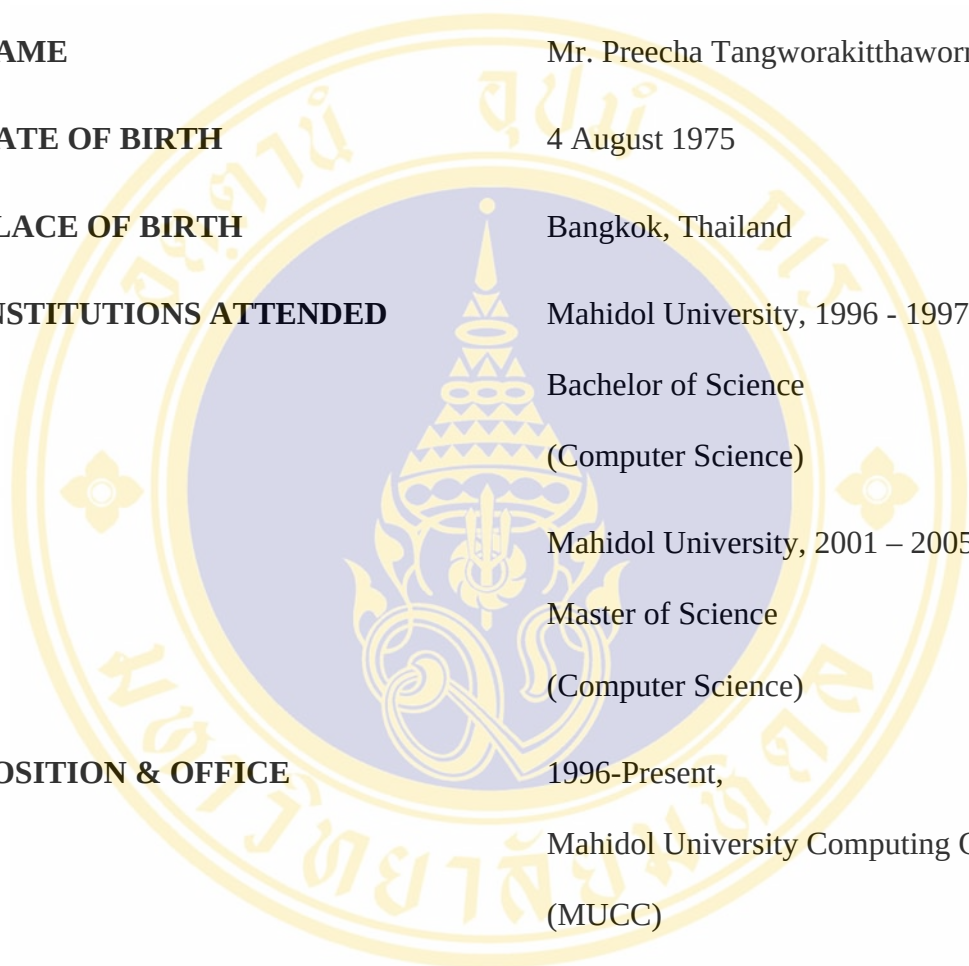
- [29] K.A. Ross, and C.R.B. Wright, “**Discrete Mathematics: Second Edition**”, Prentice-Hall International INC., Englewood Cliffs, New Jersey, 1988.
- [30] S. Sahar, “**Interestingness Via What is not Interesting**”, ACM, 1999
- [31] W.M.Shen and B.Leng, “**A Metapattern-Based Automated Discovery Loop for Integrated Data Mining – Unsupervised Learning of Relational Patterns**”, IEEE Transactions on Knowledge and Data Engineering, Vol.8, No.6, December 1996, pp.898-910.
- [32] M. Siff, and T. Reps, “**Identifying Modules Via Concept Analysis**”, Proceedings of the International Conference on Software Maintenance, 1997, pp.170-179.
- [33] A. Silberschatz, A.Tuzhilin, “**On Subjective Measures of Interestingness in Knowledge Discovery**”, Proceedings of the First International Conference on Knowledge Discovery and Data Mining, 1995, pp.275-281.
- [34] A. Silberschatz, A. Tuzhilin, “**What Makes Patterns Interesting in Knowledge Discovery Systems**”, IEEE Transactions on Knowledge and Data Engineering, Vol.8, No.6, December 1996, pp.970-974
- [35] Barry Smith, “**Ontology**”,
Available on <http://ontology.buffalo.edu/smith/articles/ontologies.htm>
- [36] G. Snelting, “**Understanding Class Hierarchies Using Concept Analysis**”, ACM Transactions on Programming Languages and Systems, Vol.22, No.3, May 2000, pp. 540-582.
- [37] J. F. Sowa, “**Knowledge Representation : Logical, Philosophical, and Computational Foundations**”, Brooks/Cole Publishing Co., Pacific Grove, CA, 2002, pp. 51-123, 492-512.
- [38] R. Srikant, R.Agrawal, “**Mining Generalized Association Rules**”, in Proceeding of the 21st VLDB Conference, Zurich, Switzerland, 1995.
- [39] Standard Upper Ontology (SUO) Working Group, available on <http://suo.ieee.org>
- [40] Stanford Medical Informatics, Standford University School of Medicine,

Available on <http://protege.stanford.edu/>.

- [41] G. Stumme, and A. Maedche, “***FCA-Merge: Bottom-Up Merging of Ontologies***”, In 7th International Conference on Artificial Intelligence (IJCAI’01), Seattle, WA, 2001, pp.255-230.
- [42] H. Suryanto, and P. Compton, “***Discovery of Ontologies from Knowledge Bases***”, First International Conference on Knowledge Capture (K-CAP’01), Victoria, BC, Canada; 21-23 October 2001, pp.171-177.
- [43] V. Tamma, T. Bench-Capon, “***An Enriched Knowledge Model for Formal Ontological Analysis***”, Proceedings of FOIS’01, IOS Press, 2001.
- [44] P.E. van der Vet, and N. J. Mars, “***Bottom-Up Construction of Ontologies***”, IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 4, July/August, 1998. pp.513 – 526.
- [45] B.B. Wang, R.I. Mckay, H.A.Abbass, and M.Barlow, “***A Comparative Study for Domain Ontology Guided Feature***”, In 25th Australian Computer Science Conference (ACSC’2003, Adelaide, Australia, 2003.
- [46] G.I. Webb, “***Beyond Association Rules: Generalized Rule Discover***”, Technical Report, Kluwer Academic Publishers, Printed in the Netherland, 2003.
- [47] G.I. Webb, “***Efficient Search for Association Rules.***” In *The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 99-107. ACM Press, 2000.
- [48] G.I. Webb, S. Butler, and D. Newland. “***On Detecting Differences Between Groups.***” In Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’03), pages 256-265. ACM, 2003.
- [49] P.C. Weinstein, “***Ontology-Based Metadata: Transforming the MARC Legacy***”, ACM, 1998, pp. 254-263.
- [50] Y.Yao, Y.Chen, X.Yang, “***A Measurement-Theoretic Foundation of Rule Interestingness Evaluation***”, Lin, T.Y., Ohsuga, S., Liau, C.J. and Hu, X. (Eds), Springer-Verlag, 2005

- [51] M.J. Zaki, and M. Ogihara, “***Theoretical Foundations of Association Rules***”, In Proceedings of 3rd SIGMOD’98 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD’98), Seattle, Washington, USA, June 1998, pp. 1-8.
- [52] Zhengxin Chen., “***Data Mining and Uncertain Reasoning: An Integrated Approach***”, Wiley, New York, 2001, pp.178-199.



BIOGRAPHYThe background of the table features a large, faint, circular seal of Mahidol University. The seal is gold-colored with a blue center. It contains a central emblem of a tiered umbrella (parasol) and a sword, surrounded by Thai script in the outer ring.

NAME	Mr. Preecha Tangworakitthaworn
DATE OF BIRTH	4 August 1975
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Mahidol University, 1996 - 1997: Bachelor of Science (Computer Science) Mahidol University, 2001 – 2005: Master of Science (Computer Science)
POSITION & OFFICE	1996-Present, Mahidol University Computing Center (MUCC) Position: Computer Scientist