

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

PREECHA TANGWORAKITTHAWORN

**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**

การค้นหารูปแบบที่มีความน่าสนใจตามอรรถวิสัยด้วยเทคนิคเมต้า-รูล ไรด์ ไม่นิ่งบนพื้นฐานออนโทโลยี
(FINDING SUBJECTIVE INTERESTINGNESS PATTERNS BY USING META-RULES GUIDED MINING BASED ON ONTOLOGY)

ปริษา ตั้งวรกิจถาวร 4437187 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

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

บทคัดย่อ

งานวิจัยนี้ได้นำเสนอแนวทางการออกแบบสถาปัตยกรรมรูปแบบใหม่ในการสร้างเมต้า-รูล ไรด์ ไม่นิ่งด้วยการประยุกต์ใช้ออนโทโลยี เพื่ออธิบายถึง "โมเดลเชิงแนวคิด" ของวัตถุในโดเมนที่สนใจ หลักการของงานวิจัยที่นำเสนอนี้ เป็นส่วนขยายต่อยอดของระบบการสืบค้นข้อมูลแบบพื้นฐาน โดยมุ่งเน้นในส่วนของการวิเคราะห์ขั้นตอนสุดท้ายของระบบ งานวิจัยนี้ประกอบด้วยการทำงาน 2 ส่วนหลัก ได้แก่ 1. ส่วนของการสืบค้นรูปแบบลักษณะเฉพาะ ทำหน้าที่ในการสืบค้นรูปแบบที่เกิดขึ้นได้ทั้งหมดของข้อมูลด้วยอัลกอริทึม Apriori และ 2. ส่วนของการสืบค้นรูปแบบลักษณะทั่วไป ซึ่งทำหน้าที่ในการสืบค้น "รูปแบบลักษณะทั่วไป" ของชุดข้อมูลซึ่งเรียกว่า "เมต้า-รูล" เราได้ทำการออกแบบกระบวนการในการสร้าง เมต้า-รูล 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 โดยผลลัพธ์จากการทดลองชี้ให้เห็นว่าแนวคิดที่เรานำเสนอนั้น สามารถให้ผลลัพธ์ที่มีประสิทธิภาพ ประสิทธิผลและสมเหตุสมผลสำหรับการสืบค้นรูปแบบที่มีความน่าสนใจในเชิงอรรถวิสัย โดยสามารถเปรียบเทียบผลลัพธ์ได้เท่ากับเครื่องมือเพื่อการค้า นอกจากนี้แล้วระบบจำลองที่เราได้พัฒนาขึ้น ยังให้ความถูกต้องแม่นยำที่สูงและให้ผลลัพธ์ที่มีความน่าสนใจ ที่สอดคล้องกับความต้องการ ของผู้ใช้ได้เป็นอย่างดี

**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)