การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

นางสาวประภัสสร สิเนหะวัฒนะ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2555

TRANSFORMATION OF RELATIONAL DATABASE TO A KNOWLEDGE REPOSITORY IN DEDUCTIVE DATABASE

Prapatsorn Sinahawattana

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Engineering in Computer Engineering

Suranaree University of Technology

Academic Year 2012

การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

_____

(รศ. ดร.กิตติศักดิ์ เกิดประสพ)
ประธานกรรมการ

_____

(รศ. ดร.นิตยา เกิดประสพ)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

_____

(อ. ดร.ธรา อั่งสกุล)
กรรมการ

มหาวิทยาลัยเทคโนโลยีสุรนารี

_____

(ศ. ดร.ชูกิจ ลิมปีจำนงค์)
รองอธิการบดีฝ่ายวิชาการ

_____

(รศ. ร.อ. ดร.กนต์ธร ชำนิประศาสน์)
คณบดีสำนักวิชาวิศวกรรมศาสตร์

ประภัสสร สิเนหะวัฒนะ : การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย (TRANSFORMATION OF RELATIONAL DATABASE TO A KNOWLEDGE REPOSITORY IN DEDUCTIVE DATABASE) อาจารย์ที่ปรึกษา: รองศาสตราจารย์ ดร. นิตยา เกิดประสพ, 91หน้า.

ในปัจจุบันไม่ว่าจะเป็นองค์กรที่มีขนาดเล็กหรือขนาดใหญ่ ต่างก็มีการเก็บข้อมูลในรูปแบบของระบบฐานข้อมูล ฐานข้อมูลที่นิยมใช้กันมากคือ ฐานข้อมูลเชิงสัมพันธ์ ซึ่งหากข้อมูลที่เก็บในฐานข้อมูลมีจำนวนเพิ่มมากขึ้นเรื่อย ๆ จะส่งผลให้การค้นหาข้อมูล ทำได้ยากมากขึ้นเนื่องจากว่าข้อมูลมีจำนวนมาก มีความซับซ้อน ดังนั้นเพื่อที่จะลดเวลาในการค้นหาข้อมูลและลดเนื้อที่ในการเก็บข้อมูล จึงเกิดแนวคิดในการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานข้อมูลนิรนัย ซึ่งฐานข้อมูลประเภทนี้จะประกอบด้วย 2 ส่วน คือส่วนที่เป็นข้อมูลเรียกว่า Extensional Database (EDB) และส่วนของกฎเรียกว่า Intensional Database (IDB) ส่วน EDB แปลงได้โดยตรงจากตารางเชิงสัมพันธ์ ในส่วนของ IDB นั้น ในงานวิจัยนี้มีการนำกฎซึ่งเป็นรูปแบบหรือโมเดลของข้อมูลที่ได้จากการทำเหมืองข้อมูลชนิดการสังเคราะห์ต้นไม้ตัดสินใจนำมาเพิ่มเข้าไปในส่วนของ IDB กฎที่ได้นี้จะมีส่วนช่วยในกระบวนการค้นหาข้อมูล ส่งผลให้แหล่งความรู้ที่ได้นั้นมีประสิทธิภาพเพิ่มขึ้นเมื่อทำการเปรียบเทียบกับแหล่งความรู้ปกติที่ประกอบด้วย EDB และ IDBในฐานข้อมูลนิรนัย

ผลลัพธ์ที่ได้จากงานวิจัยนี้พบว่า ฐานข้อมูลที่ประกอบไปด้วยกฎที่ได้จากการหารูปแบบข้อมูลจากการทำเหมืองข้อมูล มีส่วนช่วยในการค้นหาข้อมูลให้รวดเร็วขึ้นในกรณีที่ค้นหาข้อมูลที่มีอยู่หรือตรงตามรูปแบบของข้อมูล แต่ในกรณีที่ค้นหาข้อมูลที่ไม่มีอยู่หรือไม่ตรงกับรูปแบบของข้อมูล เวลาที่ใช้ค้นหาจะนานกว่าฐานข้อมูลปกติที่ประกอบด้วยข้อมูลเพียงอย่างเดียว

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2555 ลายมือชื่ออาจารย์ที่ปรึกษา_____

PRAPATSORN SINAHAWATTANA : TRANSFORMATION OF RELATIONAL DATABASE TO A KNOWLEDGE REPOSITORY IN DEDUCTIVE DATABASE. THESIS ADVISOR : ASSOC. PROF. NITTAYA KERDPRASOP, Ph.D., 91 PP.

DEDUCTIVE DATABASE/ PROLOG/ DATALOG/DECISION TREE (ID3)

Presently, in any small or large organization there have been a collection of data that are stored in a database. The most common form of database is relational database. The steadily increasing of stored data can more or less slow down the search for specific data items. We thus propose the strategy to reduce search time by transforming relational data to the knowledge repository of deductive database. This kind of database consists of two main parts : extensional database (EDB) and intensional database (IDB). The EDB part can be directly transformed from the data tables, whereas the IDB part contains rules that are constructed from the decision-tree induction method. It is the IDB part that can speedup the search process.

The experimental results showed that if the data items to be searched exist in the database, the proposed deductive database scheme with induced rules can find data faster than the conventional database that contains only facts. However, if the search items do not exist in the database, the deductive database scheme with induced rules takes longer time than database with ground facts.

School of Computer Engineering Student's Signature_____

Academic Year 2012 Advisor's Signature_____

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดี เนื่องจากได้รับความช่วยเหลืออย่างดียิ่ง ทั้งด้านวิชาการและด้านการดำเนินงานวิจัย จากบุคคลดังต่อไปนี้

รองศาสตราจารย์ ดร. นิตยา เกิดประสพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำแนะนำปรึกษา ช่วยแก้ไขปัญหาและให้กำลังใจแก่ผู้วิจัยตลอดมา รวมทั้งช่วยตรวจทาน และแก้ไขวิทยานิพนธ์เล่มนี้จนเสร็จสมบรูณ์

รองศาสตราจารย์ ดร. กิตติศักดิ์ เกิดประสพ หัวหน้าสาขาวิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี ที่ให้โอกาสในด้านการศึกษา คำปรึกษาด้านวิชาการ และกำลังใจแก่ผู้วิจัยมาโดยตลอด

ขอขอบคุณเพื่อนๆ บัณฑิตศึกษาสาขาวิศวกรรมคอมพิวเตอร์ทุกคนที่เป็นกำลังใจมาโดยตลอด

นอกจากนี้ ขอขอบคุณครู อาจารย์ทั้งในอดีตและปัจจุบันที่ให้ความรู้แก่ผู้วิจัยจนประสบความสำเร็จในชีวิต

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์เล่มนี้ ผู้วิจัยขอมอบให้กับบิดา มารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่าน ที่ได้ประสิทธิ์ประสาทความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ผู้วิจัยตลอดมา จนทำให้ประสบความสำเร็จในชีวิตตลอดมา

ประภัสสร สิเนหะวัฒนะ

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย) ..... ก

บทคัดย่อ (ภาษาอังกฤษ) ..... ข

กิตติกรรมประกาศ ..... ค

สารบัญ ..... ง

สารบัญตาราง ..... ซ

สารบัญรูป ..... ฌ

บทที่

1 บทนำ 1

1.1 ความสำคัญและที่มาของปัญหาการวิจัย 1

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 ขอบเขตของการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับ 3

2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 4

2.1 ฐานข้อมูล (Database) 4

2.1.1 ฐานข้อมูลเชิงสัมพันธ์ (Relational Database) 4

2.1.2 ฐานข้อมูลนิรนัย (Deductive Database) 6

2.2 การทำเหมืองข้อมูล (Data Mining) 9

2.2.1 วิวัฒนาการของเทคโนโลยีฐานข้อมูล 9

2.2.2 ประเภทของข้อมูลที่จะใช้ในการทำเหมืองข้อมูล 11

2.2.3 ลักษณะเฉพาะของข้อมูลที่สามารถทำเหมืองข้อมูล 12

2.2.4 ขั้นตอนการทำเหมืองข้อมูล 12

สารบัญ (ต่อ)

หน้า

2.2.5 เทคนิคการสังเคราะห์ต้นไม้ตัดสินใจ 13
2.2.6 วิธีการสร้างต้นไม้ตัดสินใจแบบ ID3 15
2.3 ภาษา Prolog และ Datalog 16
2.3.1 ภาษาโปรล็อก (Prolog) 16
2.3.2 องค์ประกอบพื้นฐานของภาษาโปรล็อก 16
2.3.3 ตัวอย่างการดำเนินการในภาษาโปรล็อก 17
2.3.4 ภาษาดาต้าล็อก (Datalog) 18
2.3.5 โปรแกรม SWI-Prolog 20
2.3.6 โปรแกรม DES 21
2.4 งานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย 27
3 วิธีดำเนินการวิจัย 31
3.1 วิธีการวิจัย 31
3.1.1 ศึกษาและวิเคราะห์ถึงปัญหาในการลดเวลาของการค้นหาข้อมูลหรือสอบถามข้อมูล 31
3.1.2 ออกแบบและพัฒนากรอบแนวคิดของการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย 31
3.1.3 ทดสอบการทำงาน ประเมินผลและปรับปรุง 36
3.2 เครื่องมือที่ใช้ในงานวิจัย 36
3.2.1 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม 36
3.2.2 เครื่องมือที่ใช้ในการประเมินประสิทธิภาพ 36
3.3 วิธีการแปลงฐานข้อมูลไปเป็นแหล่งความรู้ในฐานข้อมูลนิรนัย 37
3.3.1 อัลกอริทึมในการทำเหมืองข้อมูล 37

สารบัญ (ต่อ)

หน้า

3.3.2 อัลกอริทึมในการแปลงข้อมูล 43

3.3.3 ลักษณะที่ใช้ในการทดสอบแหล่งความรู้ 46

3.3.4 ขั้นตอนการทดสอบ 47

3.4 การเปรียบเทียบประสิทธิภาพของแหล่งความรู้ 49

4 ผลการวิเคราะห์ข้อมูลและอภิปรายผล 50

4.1 ฐานข้อมูลที่ใช้ในการแปลงเป็นแหล่งความรู้ 50

4.2 แหล่งความรู้ที่ได้จากการแปลงฐานข้อมูล 53

4.3 ผลการทดสอบประสิทธิภาพของแหล่งความรู้ 57

4.3.1 การทดสอบกับข้อมูลในฐานข้อมูลเชิงสัมพันธ์ 57

4.3.2 การทดสอบประสิทธิภาพของข้อมูลชุดที่ 1 60

4.3.3 การทดสอบประสิทธิภาพของข้อมูลชุดที่ 2 63

4.3.4 การตรวจสอบความถูกต้องของผลลัพธ์ที่ได้ 67

4.4 อภิปรายสรุปผลการเปรียบเทียบการทดสอบแหล่งความรู้ 69

5 บทสรุป 70

5.1 สรุปผลการวิจัย 70

5.2 ประโยชน์ของการแปลงฐานข้อมูลเป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย 71

5.3 ข้อจำกัดของการแปลงข้อมูลเป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย 71

5.4 แนวทางในการพัฒนาต่อ 71

รายการอ้างอิง 72

ภาคผนวก

ภาคผนวก ก. รหัสต้นฉบับโปรแกรม 74

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ (ต่อ)

หน้า

ภาคผนวก ข. บทความทางวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา 81
ประวัติผู้เขียน 91

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ **หน้า**

2.1 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย 29

4.1 แสดงรูปแบบการสอบถามข้อมูล Car 57

4.2 แสดงรูปแบบการสอบถามข้อมูล Nursery 58

4.3 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Car 59

4.4 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Nursery 59

4.5 แสดงรูปแบบการสอบถามข้อมูล Car ที่ตรงตามเงื่อนไข 60

4.6 แสดงรูปแบบการสอบถามข้อมูล Car ที่ไม่ตรงตามเงื่อนไข 60

4.7 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Car 61

4.8 แสดงรูปแบบการสอบถามข้อมูล Nursery ที่ตรงตามเงื่อนไข 64

4.9 แสดงรูปแบบการสอบถามข้อมูล Nursery ที่ไม่ตรงตามเงื่อนไข 64

4.10 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Nursery 65

ข.1 ตารางแสดงเวลาที่ใช้ในการสอบถามข้อมูล 89

สารบัญรูป

รูปที่ **หน้า**

1.1 ตัวอย่าง EDB และ IDB ของฐานข้อมูลนิรนัย 2

2.1 ส่วนประกอบภายในตารางของฐานข้อมูลเชิงสัมพันธ์ 5

2.2 ตัวอย่างของส่วนที่เป็น Extensional Database (EDB) 7

2.3 ตัวอย่างของส่วนที่เป็น Intensional Database (IDB) 8

2.4 ตัวอย่างฐานข้อมูลนิรนัย ประกอบด้วย EDB IDB และ IC 9

2.5 วิวัฒนาการของการทำเหมืองข้อมูล 11

2.6 แสดงส่วนประกอบของต้นไม้ตัดสินใจ 13

2.7 ตัวอย่างโปรแกรม SWI-Prolog 20

2.8 ตัวอย่างแสดงความพร้อมของโปรแกรม 21

2.9 ตัวอย่างโปรแกรม DES ในระบบวินโดว์ แบบ Command line 22

2.10 ตัวอย่างโปรแกรม DES+ACIDE ในระบบวินโดว์ 22

2.11 ตัวอย่างโปรแกรม DES ในระบบลีนุกซ์ 23

2.12 ตัวอย่างโปรแกรม DES ในระบบ OS MAC 23

2.13 ตัวอย่างโปรแกรม DES โหมด Datalog 24

2.14 ตัวอย่างของคำสั่งในการเปลี่ยนโหมดเป็น Prolog 25

2.15 ตัวอย่างของคำสั่งในการเปลี่ยนโหมดเป็น SQL 26

3.1 กรอบแนวคิดของการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานข้อมูลนิรนัย 32

3.2 แสดงแผนผังการทำงานในการแปลงข้อมูลเป็นฐานข้อมูลนิรนัย 33

3.3 แสดงแผนผังการทำงานในเพิ่มส่วนที่เป็นกฎที่ได้จากการทำเหมืองข้อมูล 34

3.4 แสดงแผนผังการทำงานของการทำเหมืองข้อมูล 35

สารบัญรูป (ต่อ)

รูปที่ หน้า

3.5 ตัวอย่างตารางข้อมูล 37

3.6 แสดงตัวอย่างต้นไม้ตัดสินใจในขั้นแรก 40

3.7 แสดงตัวอย่างต้นไม้ตัดสินใจ 42

3.8 ตัวอย่างข้อมูลในรูปแบบเอสคิวแอล 44

3.9 ข้อมูลที่อยู่ในรูปแบบ Fact 45

3.10 รูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล 46

3.11 รูปแบบของข้อมูลในภาษาดาต้าล็อก 46

3.12 ตัวอย่างรูปแบบของข้อมูลที่ใช้ในการทำเหมืองข้อมูล 47

3.13 ตัวอย่างรูปแบบของข้อมูลเอสคิวแอล 48

4.1 ตัวอย่างตารางข้อมูลของชุดข้อมูลเกี่ยวกับรถ (Car) 51

4.2 ตัวอย่างตารางข้อมูลของชุดข้อมูลเกี่ยวกับสถานที่รับเลี้ยงเด็ก (Nursery) 52

4.3 ฐานความรู้แบบที่ 1 ของข้อมูล Car 53

4.4 ฐานความรู้แบบที่ 2 ของข้อมูล Car 54

4.5 ฐานความรู้แบบที่ 1 ของข้อมูล Nursery 55

4.6 ฐานความรู้แบบที่ 2 ของข้อมูล Nursery 56

4.7 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Car ที่ตรงกับเงื่อนไข 62

4.8 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Car ที่ไม่ตรงกับเงื่อนไข 63

4.9 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูลที่ Nursery ตรงกับเงื่อนไข 66

4.10 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Nursery ที่ไม่ตรงกับเงื่อนไข 67

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.11 แสดงผลลัพธ์จากการสอบถามข้อมูลของฐานข้อมูลเชิงสัมพันธ์ 68
4.12 แสดงผลลัพธ์จากการสอบถามข้อมูลของแหล่งความรู้แบบที่ 1 68
ข.1 ตัวอย่างตารางข้อมูล 85
ข.2 ตัวอย่างของภาษาดาต้าล็อกที่ใช้ในการสร้างฐานความรู้ 85
ข.3 ตัวอย่างของข้อมูลในฐานความรู้ที่ได้ 86
ข.4 รูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล 87
ข.5 รูปแบบของข้อมูลในภาษาดาต้าล็อก 87
ข.6 ตัวอย่างฐานความรู้ของงานวิจัย 88
ข.7 ฐานความรู้ที่นำมาเปรียบเทียบ 88
ข.8 การเปรียบเทียบเวลาที่ใช้ในการสอบถามข้อมูลแบบตรงกับรูปแบบข้อมูล (Rule) 89
ข.9 การเปรียบเทียบเวลาที่ใช้ในการสอบถามข้อมูลแบบไม่ตรงกับรูปแบบข้อมูล (Rule) 89

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

ในปัจจุบันเทคโนโลยีฐานข้อมูลได้เข้ามามีบทบาทมากขึ้นในการดำเนินงาน ในองค์กรต่าง ๆ ไม่ว่าจะเป็นบริษัทขนาดเล็กหรือขนาดใหญ่ต่างก็มีการเก็บรวบรวมข้อมูลไว้จำนวนมาก ในอดีตนั้นการจัดเก็บข้อมูลจะทำการจัดเก็บข้อมูลต่าง ๆ ในรูปแบบของแฟ้มเอกสาร ซึ่งรูปแบบของการจัดเก็บข้อมูลแบบแฟ้มนั้นมีข้อเสียหลายประการ ไม่ว่าจะเป็นการทำให้ข้อมูลมีจำนวนมาก ข้อมูลมีความซับซ้อน ยากต่อการค้นหาและแก้ไขข้อมูล ทำให้เมื่อผู้ใช้ต้องการที่จะค้นหาข้อมูลจะต้องทำการค้นหาข้อมูลไปทีละแฟ้มเอกสารทำให้เสียเวลามาก ต่อมาเมื่อมีการนำเอาคอมพิวเตอร์มาใช้งานมากขึ้น การจัดเก็บข้อมูลจึงเริ่มปรับเปลี่ยนจากการเก็บในรูปแบบของแฟ้มเอกสารมาเป็นการจัดเก็บข้อมูลไว้ในคอมพิวเตอร์ที่เราเรียกว่า ระบบฐานข้อมูล โดยรูปแบบในการจัดเก็บข้อมูลแบบนี้กำลังเป็นที่นิยมในเกือบทุกหน่วยงานที่มีการใช้งานระบบสารสนเทศ เนื่องจากว่าระบบฐานข้อมูลจะช่วยให้สามารถจัดเก็บข้อมูลได้อย่างเป็นระบบ ทำให้ข้อมูลไม่เกิดการซ้ำซ้อนและง่ายต่อการค้นหาข้อมูล

การทำเหมืองข้อมูล หรือ การค้นหาความรู้ในฐานข้อมูล เป็นวิธีการหารูปแบบหรือโมเดลของข้อมูลที่ซ่อนอยู่ในฐานข้อมูลที่มีจำนวนมาก โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่องและหลักการทางคณิตศาสตร์ ซึ่งประโยชน์ของการทำเหมืองข้อมูลนั้นจะได้รูปแบบที่ช่วยในการตัดสินใจ การวางแผนและการดำเนินงานบางอย่างได้ (วิทยา พรพัชรพงษ์, 2549)

จากที่กล่าวมาจะเห็นได้ว่าปัจจุบันการจัดเก็บข้อมูลมีมากขึ้น ส่งผลให้ข้อมูลที่ถูกเก็บไว้เพิ่มขึ้นเป็นจำนวนมาก ซึ่งข้อมูลที่มีจำนวนมากนี้ บางข้อมูลก็ไม่ได้ถูกนำกลับมาใช้ประโยชน์อีก และข้อมูลที่มีจำนวนมากนี้ยังส่งผลให้ใช้ระยะเวลาในการค้นหาข้อมูลที่ต้องการ หรือการนำข้อมูลมาวิเคราะห์ เกิดความซับซ้อน ยุ่งยาก และใช้เวลานานมากขึ้น ผู้วิจัยจึงได้เสนอแนวคิดในการที่จะแปลงฐานข้อมูลเชิงสัมพันธ์ ที่เป็นรูปแบบของฐานข้อมูลที่เป็นที่นิยมนี้ให้ไปเป็นแหล่งความรู้ที่จะนำไปใช้ในฐานข้อมูลนิรนัยที่มีส่วนของ Rules ที่ช่วยให้การจัดเก็บข้อมูลใช้เนื้อที่ลดลงได้ ฐานข้อมูลนิรนัยจะแบ่งออกเป็นส่วน Extensional Database (EDB) ประกอบไปด้วย Fact ของข้อมูล และ Intensional Database (IDB) ประกอบด้วย Rule ที่ได้จากการนำ Fact มาสร้างเงื่อนไขดังรูปที่

1.1 จะแสดงถึงลักษณะของ EDB และ IDB ในฐานข้อมูลนิรนัย เมื่อเปรียบเทียบกับฐานข้อมูลเชิงสัมพันธ์

Parant	Child	Parant_sex
adam	able	male
adam	cain	male
john	angie	male
john	sally	male

Father	Child
adam	able
adam	cain
john	angie
john	sally

(a) ฐานข้อมูลเชิงสัมพันธ์

```
parant(adam,able).
parant(adam,cain).

parant(john,angie).
parant(john,sally).
male(adam).
male(john).
```

Facts comprising the "extensional database"

```
father(X,Y) :- parant(X,Y),male(X).
```

Rules comprising the "intensional databse"

(b) ฐานข้อมูลนิรนัย

รูปที่ 1.1 ตัวอย่าง EDB และ IDB ของฐานข้อมูลนิรนัย

ฐานข้อมูลนิรนัยในงานวิจัยนี้เป็นฐานข้อมูลที่เพิ่มในส่วนของการทำเหมืองข้อมูลเข้ามาช่วยเพื่อสร้างข้อมูลในส่วน IDB ทำให้ฐานข้อมูลนิรนัยที่ได้ต่างจากฐานข้อมูลทั่วไป การทำเหมืองข้อมูลจะทำให้เราได้รูปแบบหรือแพทเทิร์นของข้อมูลขึ้นมา รูปแบบของข้อมูลนี้จะช่วยทำให้การค้นหาข้อมูลและการวิเคราะห์ข้อมูลเป็นไปได้ง่ายขึ้น พร้อมทั้งยังช่วยลดเวลาในการค้นหาอีกด้วย

1.2 วัตถุประสงค์ของการวิจัย

จากแนวคิดในการทำงานวิจัย ผู้วิจัยได้ตั้งวัตถุประสงค์ในการวิจัยไว้ดังนี้

1.2.1 เพื่อทำการแปลงข้อมูลจาก Relational Database ไปเป็น แหล่งความรู้ทั้งส่วน EDB และ IDB ใน Deductive Database

1.2.2 เพื่อช่วยลดเวลาในการสอบถามข้อมูล (Query) กับข้อมูลที่มีจำนวนมาก ๆ ได้

1.3 ขอบเขตของการวิจัย

จากการศึกษา ค้นคว้าข้อมูลที่เกี่ยวข้องกับงานวิจัยนี้ ผู้วิจัยได้กำหนดขอบเขตของงานวิจัย การเปลี่ยนฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย คือ การใช้ภาษาดาต้าล็อกและโปรล็อกในการแปลงข้อมูล ซึ่งข้อมูลที่ใช้ในการแปลงจะเป็นฐานข้อมูลเชิงสัมพันธ์เท่านั้น และแหล่งความรู้ที่ได้จะแบ่งออกเป็น 2 แบบ คือ แบบที่ 1 จะประกอบไปด้วยส่วนของข้อมูล (Fact) เพียงอย่างเดียวเท่านั้น และแบบที่ 2 จะประกอบด้วยทั้งส่วนของข้อมูล (Fact) รวมกับส่วนของกฎที่ได้จากการทำเหมืองข้อมูล (Rule) การทำเหมืองข้อมูลใช้เทคนิคการสร้างต้นไม้ตัดสินใจ (Decision Tree Induction) งานวิจัยนี้จะทำการเปรียบเทียบประสิทธิภาพและเวลาในการค้นหาข้อมูลของแหล่งความรู้เพื่อวัดผลของการวิจัย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

จากการศึกษาเกี่ยวกับงานวิจัยนี้ผู้วิจัยคาดว่าจะเกิดประโยชน์ดังนี้

1.4.1 ช่วยให้ได้แนวทางและอัลกอริทึมที่ง่ายต่อการแปลงข้อมูลจาก Relational Database ไปเป็น Deductive Database

1.4.2 ช่วยให้ได้รูปแบบที่ง่ายต่อการสอบถามข้อมูล (Query) ในฐานข้อมูล

1.4.3 ช่วยให้ลดเวลาในการสอบถามข้อมูล (Query)

บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในบทที่ 2 นี้กล่าวถึง การทบทวนวรรณกรรม ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการวิจัยเรื่อง การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ประกอบไปด้วย ความหมายของฐานข้อมูล การทำเหมืองข้อมูล ภาษาโปรล็อก ภาษาดาต้าล็อก และงานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

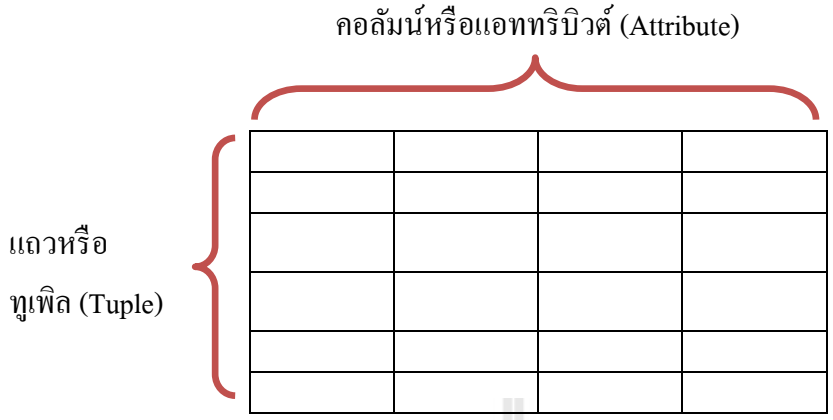
2.1 ฐานข้อมูล (Database)

ฐานข้อมูล เป็นองค์ประกอบสำคัญทั้งในงานธุรกิจ หรืองานในสาขาอื่น ๆ ที่ต้องการจัดเก็บข้อมูลเป็นจำนวนมาก ความหมายของฐานข้อมูล คือ กลุ่มของข้อมูลที่มีความสัมพันธ์กัน นำมาเก็บรวบรวมไว้ด้วยกันอย่างเป็นระบบ เปรียบเสมือนกับตู้เก็บเอกสารอิเล็กทรอนิกส์ที่รวมแฟ้มข้อมูลต่าง ๆไว้ รูปแบบของฐานข้อมูลที่นิยม คือ ฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ฐานข้อมูลเชิงวัตถุ (Object-Oriented Database) ฐานข้อมูลนิรนัย (Deductive Database) และฐานข้อมูลเว็บ (Web Database)

รูปแบบของฐานข้อมูล (ชญามน บุญประสิทธิ์, 2550) คือ รูปแบบที่ใช้อธิบายลักษณะของความสัมพันธ์ระหว่างข้อมูล วิธีการจัดการกับข้อมูลรวมไปถึงเงื่อนไขและการกำหนดเกี่ยวกับข้อมูล ในส่วนของงานวิจัยนี้ จะเน้นไปที่ฐานข้อมูล 2 ประเภท คือ ฐานข้อมูลเชิงสัมพันธ์ และฐานข้อมูลนิรนัย

2.1.1 ฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

ฐานข้อมูลเชิงสัมพันธ์ เป็นรูปแบบของฐานข้อมูลที่ได้รับความนิยมมากที่สุด (โอฬาริก สุรินต๊ะ, 2552) ฐานข้อมูลประเภทนี้จะมีลักษณะของการจัดเก็บข้อมูลไว้ในรูปแบบของตาราง (Table) ซึ่งแทนความสัมพันธ์ (Relation) และภายในตารางก็จะประกอบไปด้วย แถวแทนเรคคอร์ดของแต่ละรายการข้อมูลซึ่ง เรียกว่า ทูเพิล (Tuple) และ คอลัมน์แทนลักษณะของข้อมูล เรียกว่า แอททริบิวต์ (Attribute) ดังรูปที่ 2.1



รูปที่ 2.1 ส่วนประกอบภายในตารางของฐานข้อมูลเชิงสัมพันธ์

โครงสร้างข้อมูลเชิงสัมพันธ์ (Relational Data Structure)

โครงสร้างของฐานข้อมูล จะใช้โครงสร้างข้อมูลในลักษณะ คือ ตารางความสัมพันธ์หรือรีเลชัน (Relation) โดยรีเลชันจะถูกมองเห็นในลักษณะของตารางที่มีคุณสมบัติดังนี้ (ศุภชัย จิวะรังสินี และขจรศักดิ์ สังข์เจริญ, 2549)

- ข้อมูลในแต่ละแถว จะใช้แทน Tuple ใน Relation
- จะต้องไม่มีแถวคู่ใดในตารางเดียวกัน ที่มีข้อมูลซ้ำกัน
- ลำดับของแถวใด ๆ ในตาราง จะไม่มีความสำคัญ นั่นคือ กำหนดให้แถวใดปรากฏก่อนหรือหลังจะไม่ส่งผลต่อการใช้งานหรือจัดการกับข้อมูล
- ลำดับก่อนหลังของคอลัมน์ใด ๆ ในตาราง จะไม่มีความสำคัญ เช่นเดียวกับแถว

ชนิดของ Relations

ในระบบจัดการฐานข้อมูลทั่วๆ ไปจำแนกรีเลชันได้ 2 ประเภท (แพรตะวัน จารุตัน, 2554) คือ

- รีเลชันหลัก (Base Relation) เป็นรีเลชันที่ถูกกำหนดขึ้นเพื่อเก็บข้อมูลและเพื่อนำข้อมูลไปใช้เมื่อมีการสร้างรีเลชันโดยใช้ภาษาเฉพาะที่เรียกว่า Data Definition Language
- วิว (View) หรือรีเลชันสมมติ (Virtual Relation) เป็นรีเลชันที่ถูกสร้างขึ้นตามความต้องการใช้ข้อมูลของผู้ใช้แต่ละคน เนื่องจากผู้ใช้แต่ละคนอาจต้องการข้อมูลในลักษณะที่แตกต่างกัน จึงทำการกำหนดวิวของตัวเองขึ้นมาจากรีเลชันหลัก เพื่อความสะดวกในการใช้ข้อมูล

ข้อดีของฐานข้อมูลเชิงสัมพันธ์

ฐานข้อมูลเชิงสัมพันธ์เป็นรูปแบบฐานข้อมูลที่ง่ายสำหรับผู้ใช้งานทั่วไป โดยเฉพาะกับผู้ที่ไม่ใช่นักวิเคราะห์หรือนักออกแบบ จึงทำให้ฐานข้อมูลนี้เป็นที่นิยมใช้กันมากกว่าฐานข้อมูลแบบอื่น ๆ เนื่องจากมีข้อดีดังนี้ (จารุณี ซามาตย์, 2551)

- มีโครงสร้างข้อมูลที่ง่าย ทำให้ง่ายต่อการใช้งาน
- ภาษาที่ใช้ เป็นภาษาระดับสูงที่เรียกว่า Relational Complete Language เช่น พวกภาษา SQL, QBE เป็นต้น
- มีความยืดหยุ่นเนื่องจากสามารถที่จะเพิ่ม ลบ แก้ไขข้อมูลได้โดยไม่ต้องเปลี่ยนแปลง Application

2.1.2 ฐานข้อมูลนิรนัย (Deductive Database)

ฐานข้อมูลนิรนัย คือฐานข้อมูลที่ประกอบด้วยข้อมูล (Fact) และกฎ (Rules) ที่นำไปใช้อนุมานเพื่อสร้างข้อมูลเพิ่มเติมจากส่วนที่ประกาศไว้เป็นข้อเท็จจริง นอกจากงานด้านบันทึกและประมวลผลแล้ว ฐานข้อมูลนิรนัยยังถูกนำไปใช้ทางด้านปัญญาประดิษฐ์

โครงสร้างของฐานข้อมูลนิรนัย

ฐานข้อมูลนิรนัย จะประกอบไปด้วย 2 ส่วนที่สำคัญ คือ Extensional Database (EDB) และ Intensional Database (IDB) ดังนี้ (Tomasic, 1993)

1. Extensional Database (EDB) จะประกอบไปด้วยเซตของข้อมูลที่อยู่ในรูปแบบ Fact ซึ่ง Fact ในที่นี้จะเปรียบเสมือนกับค่าของข้อมูลทุกแอททริบิวต์แต่ละแถวในตารางข้อมูลทั่วไป หรือเรียกว่าข้อมูลในแต่ละเรคคอร์ดของตาราง ดังรูปที่ 2.2 จะแสดงตัวอย่างของข้อมูลที่อยู่ในรูปแบบของ EDB

```
parent(adam, able).

parent(adam, may).

parent(ann, tom).

male(adam).

female(may).
```

รูปที่ 2.2 ตัวอย่างของส่วนที่เป็น Extensional Database (EDB)

2. Intensional Database (IDB) จะเป็นส่วนที่ประกอบด้วยเซตของกฎหรือเงื่อนไขที่มาจากการนำส่วนของ Fact (EDB) มาเขียนเป็นเงื่อนไข ซึ่งจะเรียกว่า Rule ส่วน IDB อาจจะมองแล้วคล้ายกับส่วนที่เป็นวิวของฐานข้อมูลทั่วไป แต่จะแตกต่างกันส่วนของ IDB สามารถที่จะทำการเรียกซ้ำได้ สามารถเพิ่มข้อมูลจากเดิมที่มีอยู่จริงให้เพิ่มขึ้นตามเงื่อนไขที่ได้กำหนดเพียงแค่ไม่กี่เงื่อนไข ดังรูปที่ 2.3 จะเป็นการแสดงตัวอย่างลักษณะของ IDB ที่กำหนดกฎเกี่ยวกับ father, mother, sibling, และ cousin โดยกฎเหล่านี้สร้างจาก Fact ที่ประกาศข้อเท็จจริงเกี่ยวกับ parent, male และ female

```
father(X,Y) :- parent (X,Y) ,male(X).

mother(X,Y) :- parent (X,Y) ,female(X).

sibling(X,Y) :- parent(X,Z), parent(Y,Z), X <> Y.

cousin(X,Y) :- parent(X,Xp), parent(Y,Yp),
                    sibling(Xp,Yp).
```

รูปที่ 2.3 ตัวอย่างของส่วนที่เป็น Intensional Database (IDB)

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่างานวิจัยบางงานจะกล่าวว่าฐานข้อมูลนิรนัยจะประกอบไปด้วย 3 ส่วน (Lee, 1996) คือส่วนของ Extensional Database (EDB), Intensional Database (IDB) และ Integrity Constraints (IC) ส่วนของ IDB และ IC จะมีลักษณะเป็นเงื่อนไขในฐานข้อมูล ซึ่งส่วนของ IDB และ IC เป็นกฎที่จะแสดงถึงคำสั่งหรือเงื่อนไขตรรกะเบื้องต้นใช้ Fact ประกอบการเขียนให้อยู่ในรูปแบบของประโยคของฐานข้อมูล ดังรูปที่ 2.4

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
EDB :
female(X)  male(X)
mother(X,Y) father(X,Y)
IDB:
parent(X,Y) :- mother(X,Y)
parent(X,Y) :- father(X,Y)
sibling(Y,Z) :-parent(X,Y),parent(X,Z)
sister(X,Y) :- sibling(X,Y) ,female(X)
IC:
false :- male(X),female(X)
false :- mother(X,Y),not female(X)
false :- father(X,Y),not male(X)
```

รูปที่ 2.4 ตัวอย่างฐานข้อมูลที่มีทั้ง EDB, IDB และ IC

2.2 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล หมายถึง การค้นหาความสัมพันธ์หรือรูปแบบ (Pattern) ทั้งหมดของข้อมูลจากฐานข้อมูล ซึ่งรูปแบบและความสัมพันธ์ที่ได้จะอยู่ในรูปของกฎ (Rule) ผลลัพธ์ที่ได้จากการทำเหมืองข้อมูล จะนำไปใช้ประโยชน์ให้กับองค์กรในส่วนของการวางแผน การดำเนินงานหรือช่วยตัดสินใจดำเนินการบางอย่างได้ (วิทยา พรพัชรพงศ์, 2549)

2.2.1 วิวัฒนาการของเทคโนโลยีฐานข้อมูล

ในช่วงปี ค.ศ. 1960 เทคโนโลยีฐานข้อมูลได้เริ่มพัฒนามาจากการประมวลผลและแฟ้มข้อมูลพื้นฐาน การค้นคว้าและพัฒนาระบบฐานข้อมูลได้มีการปรับปรุงมาโดยลำดับ

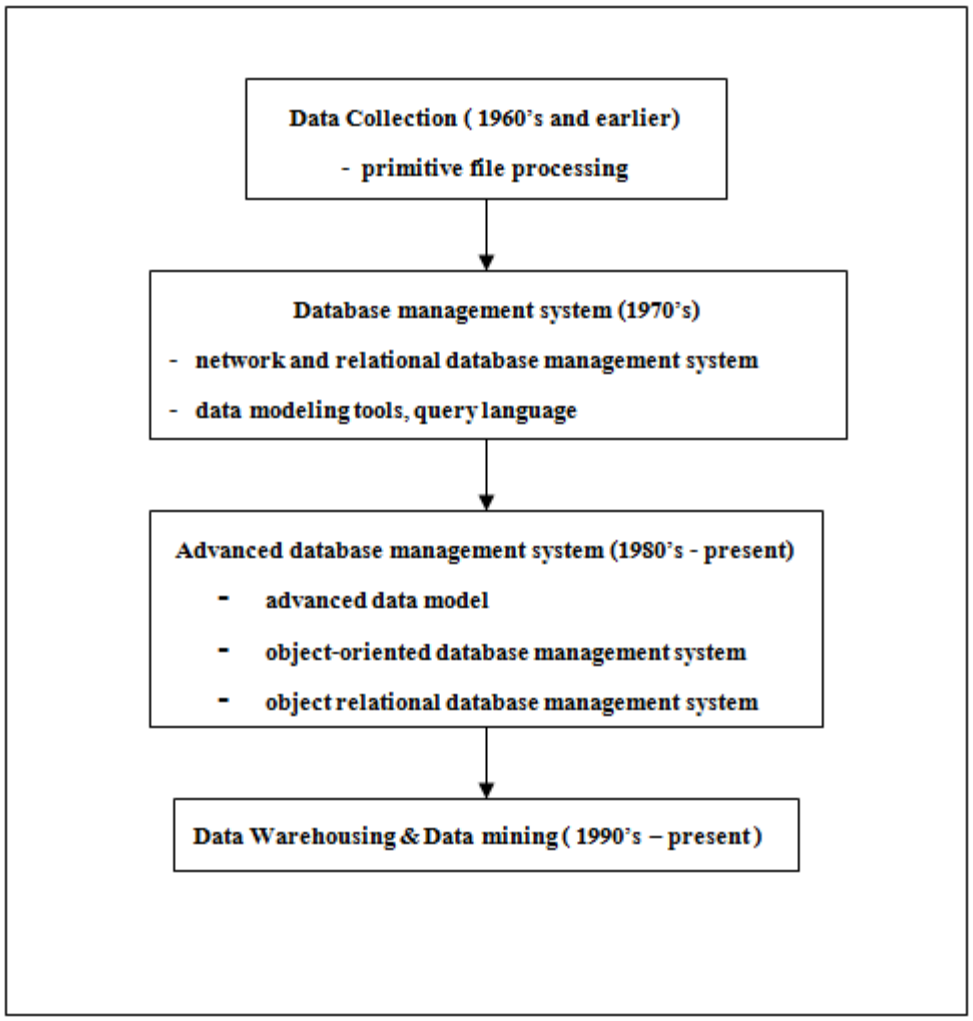
ช่วงปี ค.ศ. 1970 ได้นำไปสู่การพัฒนาระบบการเก็บข้อมูลในรูปแบบตาราง (Relational Database System) มีเครื่องมือจัดการโมเดลข้อมูล และมีเทคนิคการใช้อินเด็กซ์และการ

บริหารข้อมูล นอกจากนี้ผู้ใช้ยังได้รับความสะดวกในการเข้าถึงข้อมูลโดยการใช้ภาษาในการเรียกข้อมูล (Query Language)

ช่วงปี ค.ศ. 1980 เทคโนโลยีฐานข้อมูลได้เริ่มมีการปรับปรุงและพัฒนาในการหาระบบจัดการที่มีศักยภาพมากขึ้น ความก้าวหน้าในเทคโนโลยีฮาร์ดแวร์ ใน 30 ปีที่ผ่านมา ได้นำไปสู่การจัดเก็บข้อมูลจำนวนมากที่มีความซับซ้อนได้อย่างมีประสิทธิภาพเพิ่มขึ้น

ช่วงปี ค.ศ. 1990 - ปัจจุบัน สามารถจัดเก็บข้อมูลได้ในหลายรูปแบบ แตกต่างกันทั้งระบบปฏิบัติการ หรือการจัดเก็บฐานข้อมูล ซึ่งการนำข้อมูลทั้งหมดมารวมและจัดเก็บไว้ในรูปแบบเดียวกันเรียกว่า Data Warehouse เพื่อความสะดวกในการจัดการต่อไป ซึ่งเทคโนโลยี Data Warehouse รวมไปถึง Data Cleansing, Data Integration และ On-Line Analytical Processing (OLAP) เป็นเทคนิคการวิเคราะห์ข้อมูลในหลาย ๆ มิติรวมถึงการวิเคราะห์ข้อมูลในลักษณะการทำเหมืองข้อมูลได้เกิดขึ้นมาตามลำดับ (รูปที่ 2.5)

มหาวิทยาลัยเทคโนโลยีสุรนารี



วิทยาลัยเทคโนโลยีสุร

รูปที่ 2.5 วิวัฒนาการของการทำเหมืองข้อมูล (พูนศักดิ์ ประคำทอง, 2551)

2.2.2 ประเภทของข้อมูลที่จะใช้ในการทำเหมืองข้อมูล

ข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูลนั้นจะต้องเป็นข้อมูลที่มีจำนวนมาก เพื่อที่จะนำมาค้นหาความสัมพันธ์ ประเภทของข้อมูลมีดังนี้ (อดุลย์ ยิ้มงาม, 2552)

1. ข้อมูลที่มาจากฐานข้อมูลเชิงสัมพันธ์ (Relational Databases) เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity-Relationship (ER) Model

2. ข้อมูลจากคลังข้อมูล (Data Warehouses) เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่เดียวกัน

3. ข้อมูลจากฐานข้อมูลรายการเปลี่ยนแปลง (Transactional Databases) ประกอบด้วยข้อมูลที่แต่ละทรานแซคชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลที่ประกอบด้วย ชื่อลูกค้าและรายการสินค้าที่ลูกค้ารายนั้นซื้อ เป็นต้น

4. จากฐานข้อมูลพิเศษหรือที่เก็บข่าวสารพิเศษ (Advanced Databases) ซึ่งได้แก่

- ฐานข้อมูลเชิงวัตถุ (Object-Oriented Database)
- ฐานข้อมูลเกี่ยวกับเวลา (Temporal Database)
- ฐานข้อมูลข้อความ (Text Databases) และฐานข้อมูลมัลติมีเดีย
- ฐานข้อมูลแบบเก่าในอดีตหรือข้อมูลที่มาจากต่างฐานข้อมูลกัน
- ฐานข้อมูลจากแหล่ง WWW

2.2.3 ลักษณะเฉพาะของข้อมูลที่สามารถทำเหมืองข้อมูล

ข้อมูลที่สามารถนำมาทำเหมืองข้อมูลนั้น จะต้องมีลักษณะเฉพาะของข้อมูล ดังนี้

1. ข้อมูลขนาดใหญ่มากจนยากที่จะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือโดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล

2. ข้อมูลที่มาจากแหล่งต่าง ๆ และได้จากการรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle, DB2, MS SQL, MS Access เป็นต้น

3. ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการ Mining หากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำ Mining แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำ Mining สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลาจึงต้องทำ Mining ใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม

4. ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้สามารถนำมาทำ Mining ได้เช่นกันแต่ต้องใช้เทคนิคการทำ Data Mining ขั้นสูง

2.2.4 ขั้นตอนการทำเหมืองข้อมูล

ขั้นตอนในการทำเหมืองข้อมูลประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้ (Zaiane, 1999)

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไปและแก้ไขข้อมูลบกพร่อง
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มาจากหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน

– Data Selection เป็นขั้นตอนการดึงข้อมูลเพียงบางส่วนจากแหล่งที่บันทึกไว้มาใช้สำหรับการวิเคราะห์

– Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้มีรูปแบบที่เหมาะสมสำหรับการทำเหมืองข้อมูล

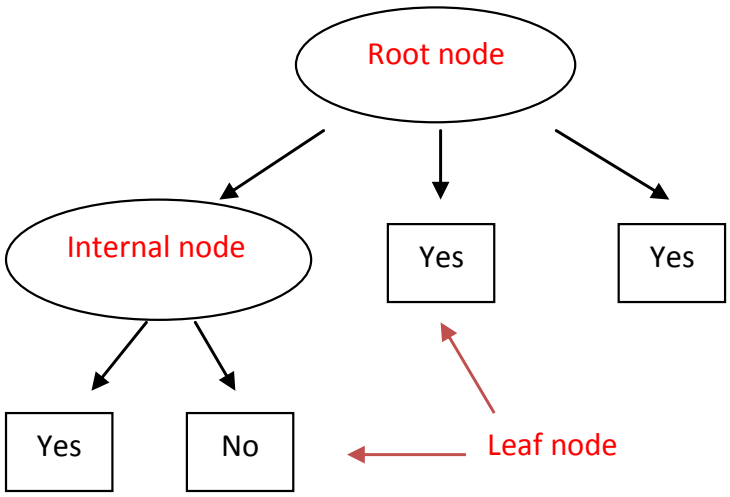
– Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่

– Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

– Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจได้ง่าย

2.2.5 เทคนิคการสังเคราะห์ต้นไม้ตัดสินใจ

การสังเคราะห์ต้นไม้ตัดสินใจ (Decision Tree Induction) เป็นการเรียนรู้โดยการจำแนกประเภทของข้อมูล (Classification) ออกเป็นกลุ่ม (Class) ต่างๆ โดยใช้คุณสมบัติ (Attribute) ของข้อมูลในการจำแนก (Quinlan, 1986) ต้นไม้ตัดสินใจ (Decision Tree) จะมีลักษณะเป็น Flow Chart เหมืองโครงสร้างของต้นไม้ ที่แต่ละโหนดก็จะแสดงถึงลักษณะของข้อมูล ดังรูปที่ 2.5 จะเป็นการแสดงส่วนประกอบของต้นไม้ตัดสินใจ



รูปที่ 2.6 แสดงส่วนประกอบของต้นไม้ตัดสินใจ

(1) ลักษณะของต้นไม้ตัดสินใจ

– ผลที่ได้จากการเรียนรู้จะแสดงอยู่ในรูปต้นไม้ ซึ่งประกอบไปด้วยโหนดแรกสุดที่เรียกว่าโหนดราก (Root Node) จาก Root Node ก็จะแตกออกเป็นโหนดลูกหรือโหนดภายใน (Internal Node) และโหนดลูกก็จะมีลูกของตัวเองซึ่งโหนดในระดับสุดท้ายจะเรียกว่าโหนดใบ (Leaf Node)

– แต่ละเส้นทางจากโหนดเริ่มต้นไปจนถึงโหนดสุดท้ายสามารถแสดงให้อยู่ในรูปกฎ IF-THEN ได้

– มีความทนทานต่อข้อมูลที่มีสัญญาณรบกวน (Noisy Data) เช่น แอททริบิวต์ที่ไม่เกี่ยวข้องและค่าของแอททริบิวต์ที่ผิดพลาดหรือขาดหายไป

– การเรียนรู้วิธีการนี้มีความรวดเร็วกว่าอัลกอริทึมอื่นที่ใช้ในการจำแนกข้อมูล

(2) วิธีการของต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจจะเป็นกระบวนการค้นหาแอททริบิวต์ที่เหมาะสมจากบนลงล่าง โดยเริ่มจากการเลือกแอททริบิวต์ที่ดีที่สุดมาเป็นโหนดราก เมื่อข้อมูลผ่านการแบ่งที่โหนดรากแล้ว ก็จะหาแอททริบิวต์ที่ดีที่สุดของข้อมูลที่ผ่านการแบ่งนั้นมาสร้างโหนดลูกของโหนดรากนั้นต่อไป และจะวนสร้างโหนดลูกและต้นไม้ย่อยของแต่ละกิ่งไปเรื่อย ๆ จนข้อมูลนั้นถูกจัดอยู่ในกลุ่มเดียวกัน วิธีการที่นำมาใช้ในการสร้างต้นไม้ตัดสินใจนั้น มีหลายวิธี ไม่ว่าจะเป็น C4.5 (Quinlan, 1992), ID3 (Quinlan, 1979), CART (Breiman et al, 1984) และ Hunt's algorithm (Quinlan, 1979) ในงานวิจัยนี้จะมุ่งเน้นไปยังการสร้างต้นไม้ตัดสินใจโดยใช้วิธี ID3

(3) ข้อดีของต้นไม้ตัดสินใจ

ข้อดีของการสร้างต้นไม้ตัดสินใจ ที่ทำให้เป็นวิธีการจำแนกกลุ่มของข้อมูลที่ได้รับความนิยมในการใช้งานมากที่สุด

– เป็นวิธีการที่ง่าย และให้ผลลัพธ์เร็วกว่าวิธีการอื่น

– ผลลัพธ์ที่ได้สามารถแปลงเป็นกฎได้ และสามารถนำไปใช้ได้จริง

– สามารถนำไปประยุกต์ใช้งานกับการค้นหาข้อมูลได้

– ให้ผลลัพธ์ที่มีความถูกต้องและแม่นยำ

2.2.6 วิธีการสร้างต้นไม้ตัดสินใจแบบ ID3

ในการสร้างต้นไม้ตัดสินแบบ ID3 จะใช้ค่าเกนของสารสนเทศ (Information Gain) ในการตัดสินใจที่จะเลือกแอททริบิวต์ที่จะใช้เป็นโหนดรากหรือโหนดในต้นไม้ โดยดูจากค่าเกนสูงที่สุด เป็นโหนดเริ่มต้น (นิตยา เกิดประสพ, 2547)

ค่าเกนนี้จะคำนวณได้โดยใช้ความรู้จากทฤษฎีสารสนเทศ คือ ค่าสารสนเทศของข้อมูลขึ้นกับค่าความน่าจะเป็นของข้อมูล

ค่าสารสนเทศของข้อมูล = $-\log_2$(ความน่าจะเป็นของข้อมูล)

ถ้า M หมายถึงชุดข้อมูล ค่าสารสนเทศ (Information) ของชุดข้อมูล M นั้น หรือเรียกว่าค่าเอนโทรพี (Entropy) ของชุดข้อมูล M จะเขียนแทนด้วย I(M)

$$I(M) = \sum_{i}^{n} - P(m_i) \log_2 P(m_i)$$

ในการเลือกแอททริบิวต์ที่จะมาเป็นโหนดรากจะอาศัยค่าเกนของสารสนเทศ ซึ่งจะคำนวณได้จากค่าสารสนเทศทั้งหมดของชุดข้อมูลนั้นลบด้วยค่าสารสนเทศหลังจากเลือกแอททริบิวต์หนึ่งเป็นโหนดราก ค่าสารสนเทศหลังจากแบ่งตามคุณสมบัติที่เลือกแล้วจะคำนวณได้จาก ค่าผลรวมของผลคูณระหว่างค่าสารสนเทศของแต่ละโหนดกับอัตราส่วนของข้อมูลในแต่ละกิ่งต่อข้อมูลทั้งหมดที่โหนดนั้นๆ

ถ้าให้ข้อมูลคือ T และแอททริบิวต์ที่เป็นโหนด คือ X ซึ่งมีค่าที่เป็นไปได้ n ค่า โหนด T จะถูกแบ่งออกเป็น {t1, t2,..., tn} ตามค่าที่เป็นไปได้ของ X ดังนั้นการคำนวณหาค่าสารสนเทศหลังจากแบ่งตามแอททริบิวต์X จะเป็นดังนี้

$$I_x(T) = \sum_{i=1}^{n} \frac{|t_i|}{|T|} I(t_i)$$

ค่าเกนของสารสนเทศของแอททริบิวต์ X สามารถคำนวณได้จากการลบค่าสารสนเทศทั้งหมดที่โหนดนี้กับค่าสารสนเทศที่ได้หลังจากแบ่งด้วยแอททริบิวต์ X ดังนี้

$$\text{Gain}(X) = I(T) - I_x(T)$$

ในงานวิจัยนี้จะใช้เทคนิคของการเรียนรู้ต้นไม้ตัดสินใจในการสร้างรูปแบบของข้อมูลในฐานข้อมูล แล้วนำรูปแบบที่ได้มาแปลงเป็นกฎให้อยู่ในรูปแบบของแหล่งความรู้ เพื่อเพิ่มประสิทธิภาพให้กับแหล่งความรู้ ในส่วน IDB และ IC ของฐานข้อมูลนิรนัย

2.3 ภาษา Prolog และ Datalog

2.3.1 ภาษาโปรล็อก (Prolog)

เป็นภาษาสำหรับการเขียนโปรแกรมเชิงตรรกะ (Logic Programming) Prolog ย่อมาจาก PROgrammation en LOGique ถูกสร้างขึ้นโดย Alian Colmerauer ในช่วง ค.ศ. 1972 โดยมีวัตถุประสงค์เพื่อประมวลผลภาษาธรรมชาติ (ศศลักษณ์ ทองขาว, 2547)

ภาษาโปรล็อกเป็นภาษาที่มีความแตกต่างจากภาษาอื่น ตรงที่สามารถประมวลผลในเชิงสัญลักษณ์ (Symbolic Computation) ได้ดี จึงถูกนำไปใช้ในการเขียนโปรแกรมด้านปัญญาประดิษฐ์ ลักษณะของภาษานั้นมีไวยากรณ์และความหมายที่ง่ายและชัดเจน เป็นภาษาที่มีรากฐานจาก Predicate Logic ในทางคณิตศาสตร์

งานวิจัยจำนวนมากที่ทำให้เกิดการพัฒนาภาษาโปรล็อกในปัจจุบันนั้น เป็นผลมาจากโครงการระบบคอมพิวเตอร์ยุคที่ห้า (Fifth Generation Computer Systems Project - FGCS) ซึ่งเลือกรูปแบบหนึ่งของภาษาโปรล็อกเป็นภาษาแก่น (Kernel Language) ของระบบปฏิบัติการ

ตัวแปลภาษาของภาษาโปรล็อกนั้นมีมากมายไม่ว่าจะเป็นแบบ Public Domain และในเชิงพาณิชย์ สำหรับในงานวิจัยนี้ตัวแปลที่ใช้ จะใช้ SWI-Prolog (http://www.swi-prolog.org/) เป็นตัวแปลภาษา

2.3.2 องค์ประกอบพื้นฐานของภาษาโปรล็อก

องค์ประกอบที่สำคัญของโปรล็อก ประกอบด้วย ตัวแปร (Variables), เพรดิเคต (Predicates), และประโยค (Clause) ซึ่งแต่ละองค์ประกอบมีลักษณะดังนี้ (สุนิสา ริมเจริญ, 2554)

(1) ตัวแปร (Variables)

ตัวแปรในภาษาโปรล็อกจะประกอบไปด้วย ตัวอักษร, ตัวเลข และเครื่องหมาย Underscore (_) โดยกำหนดว่าตัวแปรทุกตัวจะต้องขึ้นต้นด้วยตัวอักษรตัวพิมพ์ใหญ่หรือเครื่องหมาย Underscore อย่างใดอย่างหนึ่ง เช่น X, Y, Output, Input และ _ เป็นต้น สำหรับตัวแปรที่เป็น _ นั้นจะหมายถึงตัวแปรที่ไม่สนใจในค่าของตัวแปรนั้น

(2) เพรดิเคต (Predicates)

Predicates หมายถึงความสัมพันธ์ (Relation) ซึ่งชื่อของ Predicates จะประกอบไปด้วยตัวอักษร, ตัวเลข และเครื่องหมาย Underscore (_) โดยกำหนดว่าชื่อที่ใช้ในการตั้งตัวแรกต้องเป็นตัวอักษรตัวพิมพ์เล็กเท่านั้น เช่น student ที่หมายถึงความสัมพันธ์ของนักเรียน หรือ parent ที่หมายถึงความสัมพันธ์ของผู้ปกครอง จำนวน Argument ของ Predicates จะเรียกว่า Arity ของความสัมพันธ์ โดยทั่วไป Predicate q ที่มี Arity n รูปแบบการเขียนจะเป็น q/n เช่น parent/2

ส่วนของนิพจน์ที่ประกอบด้วย Predicates โดยดำเนินการกับ Argument จะเรียกว่า Atomic Formula เป็นหน่วยพื้นฐานของภาษาโปรล็อก ตัวอย่างของ Atomic Formula เช่น

parent(adle, tom).

result(a, b, R).

student(susan, 15, female).

(3) Clauses

ภาษาโปรล็อกมีประสิทธิภาพในการประมวลผลแบบ Clauses ซึ่งหมายถึงประโยค หรือ Statement ที่อยู่ในรูปของ head หรือ head : - body เมื่อส่วนของ head ประกอบด้วยส่วนของ Atomic Formula หนึ่งชุดและส่วนของ body ประกอบไปด้วยส่วน Atomic Formula ตั้งแต่หนึ่งชุดเป็นต้นไป เช่น

parent(adle, tom). เป็น clauses ที่มีเฉพาะ head

student(susan, 15, female). เป็น clauses ที่มีเฉพาะ head

father(X,Y) : - parent(X,_) , male(X). เป็น clauses ที่มี head และส่วนของ body

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.3.3 ตัวอย่างการดำเนินการในภาษาโปรล็อก

(1) การเท่ากัน (Equality)

การเท่ากันของวัตถุในลักษณะ เช่น p(c) = p(c) หรือ a = a จะเรียกการเท่ากันลักษณะนี้ว่า การเท่ากันตามโครงสร้างของไวยากรณ์ การเท่ากันแบบนี้จะเขียนแทนด้วยสัญลักษณ์ = = เช่น

?- b = = b

?- 5 = = 5

?- p(a) = = p(a)

สำหรับการหารเท่ากันอีกกรณี คือ โครงสร้างไวยากรณ์ของนิพจน์ไม่เท่ากัน แต่มีค่าที่เท่ากัน หรือมีเนื้อหาเท่ากัน การเท่ากันแบบนี้จะเขียนแทนด้วยสัญลักษณ์ =:= เช่น

?- 2+1 =:= 3+0

?- 5 =:= 5

?- p(a) =:= p(a)

(2) การทำให้เป็นหน่วยเดียวกัน (Unification)

Unification เป็นการจับคู่นิพจน์สองชุด เพื่อตรวจสอบการเท่ากัน ถ้านิพจน์ประกอบด้วยตัวแปรจะมีการแทนค่าตัวแปรก่อนการตรวจสอบการ เมื่อนำมาดำเนินการมีผลทำให้นิพจน์ทั้งสองเท่ากันตามโครงสร้างของไวยากรณ์ การเท่ากันลักษณะนี้จะเขียนแทนด้วยสัญลักษณ์ = เช่น

?- p(X) = p(a) ผลลัพธ์ที่ได้ คือ X=a ซึ่งเป็นการแทนค่าตัวแปรที่ส่งผลให้นิพจน์ทั้งสองเท่ากัน

(3) การประมวลผล (Computation)

ภาษาโปรล็อกจะทำหน้าที่ประมวลผลด้วยการดำเนินการ กับเป้าหมาย เมื่อเป้าหมายคือ Atomic Formula ตั้งแต่หนึ่งตัวขึ้นไป โดยมีเครื่องหมาย “,” เป็นตัวคั่น Atomic Formula แต่ละชุดที่อยู่ในเป้าหมายเรียกว่าเป้าหมายย่อย เช่น ตัวอย่างเป้าหมายต่อไปนี้ประกอบด้วยเป้าหมายย่อย 2 ชุด

?- parent(X, james), parent(Y, X).

การดำเนินการกับเป้าหมาย จะทำการการค้นหา head ที่เหมาะสมกับเป้าหมายย่อยก่อน โดยโปรแกรมจะทำการค้นหาจากส่วนต้นของโปรแกรมไปจนถึงส่วนสุดท้ายของโปรแกรม และในกรณีที่มีเป้าหมาย 2 ตัวขึ้นไป จะเริ่มดำเนินการกับเป้าหมายจากซ้ายไปขวา

ภาษาโปรล็อกนี้ ผู้วิจัยจะนำมาใช้ในส่วนของการทำเหมืองข้อมูล คือใช้ในการเขียนอัลกอริทึมในการหารูปแบบของข้อมูล แล้วในส่วนของการแปลงรูปแบบที่ได้จากการทำเหมืองให้อยู่ในรูปแบบของกฎในฐานข้อมูลนิรนัย

2.3.4 ภาษาดาต้าล็อก (Datalog)

Datalog ย่อมาจากคำว่า Deductive Database Logic Programming Language ภาษา Datalog เป็นภาษาที่ใช้ในการสอบถาม (Query) และประมวลผลข้อมูลสำหรับฐานข้อมูลนิรนัย (Deductive Database) ภาษา Datalog มีลักษณะเหมือนภาษา Prolog แต่ในภาษา Datalog นั้นจะไม่ซับซ้อนเท่ากับ Prolog

โครงสร้างของ Datalog

โครงสร้างของ Datalog จะแบ่งออกเป็นส่วนที่เป็น Atom หรือ Fact และส่วนที่เป็นประโยคแบบมีเงื่อนไข หรือกฎ (McCarthy, 2011)

ในส่วนที่เป็น Atom หรือ Fact แสดงได้ดังนี้

Syntax : $P(T_1, \ldots, T_n)$

Semantics : predicate P is true of terms T_1 and ... and Tn

ไวยากรณ์ในส่วนที่เป็นกฎ แสดงได้ดังนี้

Syntax : $A_0 :- A_1, \ldots, A_k$

Semantics : atom A_0 is true if atoms A_1 and ... and Ak are true

ตัวอย่างภาษา Datalog

ภาษา Datalog จะประกอบไปด้วย Fact ซึ่งเป็นส่วนที่เก็บข้อมูลตาราง เช่น ถ้าตารางชื่อว่า parent และ john เป็นผู้ปกครองของ sam ข้อมูลใน fact ที่เก็บอยู่ในตารางจะเขียนได้ดังนี้

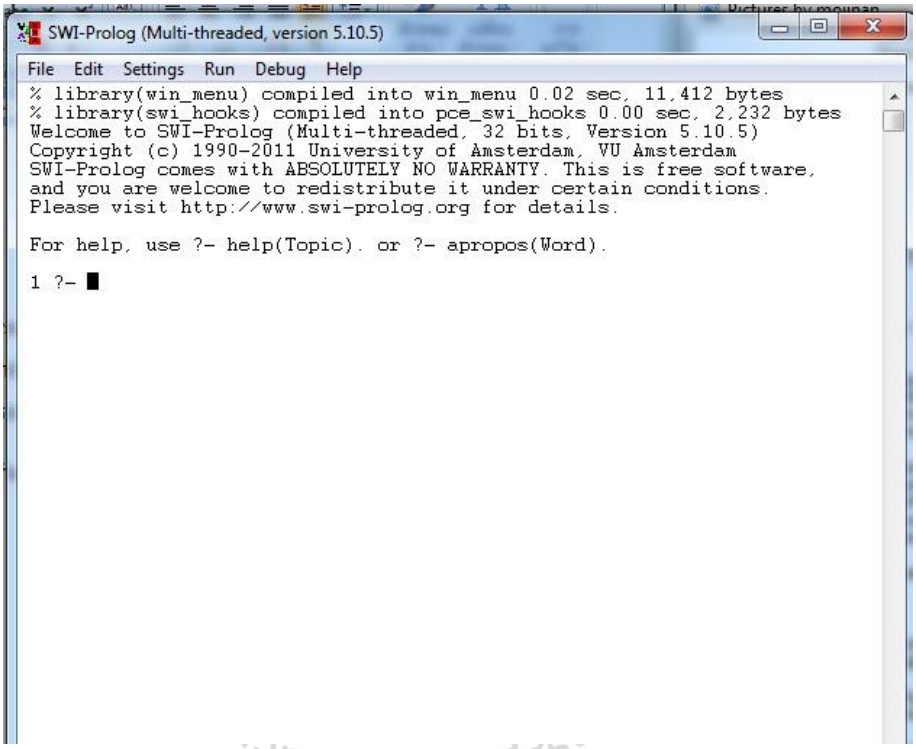
```
> parent(john, sam).
```

ในส่วนของฐานข้อมูลนิรนัย (Deductive Database) สามารถใช้กฎ (Rule) มาช่วยในการสร้าง Fact ขึ้นมาใหม่ เช่น กฎที่บอกถึงการเป็นบรรพบุรุษกัน ถ้า A เป็นผู้ปกครองของ C และ C เป็นผู้ปกครองของ B แล้วหมายความว่า A เป็นบรรพบุรุษของ B โดยเขียนเป็นกฎได้ดังนี้

```
ancestor(A, B) :- parent(A, C) , parent(C, B).
```

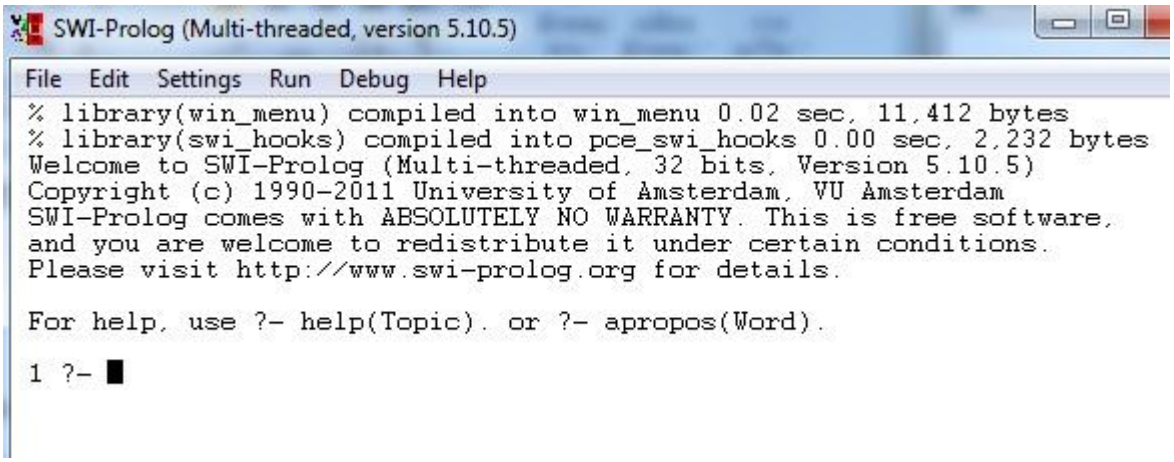
2.3.5 โปรแกรม SWI-Prolog

โปรแกรม SWI-Prolog เป็นโปรแกรมที่ใช้ในการแปลภาษาโปรล็อก ซึ่งตัวโปรแกรมจะมีให้เลือกใช้ทั้งระบบปฏิบัติการลีนุกซ์ และระบบปฏิบัติการวินโดว์ และยังมีให้เลือกใช้ทั้งแบบ Command Line และแบบ GUI ตัวโปรแกรมสามารถโหลดได้จาก http://www.swi-prolog.org/Download.html จากรูปที่ 2.7 จะเป็นตัวอย่างของโปรแกรม SWI-Prolog



รูปที่ 2.7 ตัวอย่างโปรแกรม SWI-Prolog

การใช้งานโปรแกรม SWI-Prolog สามารถทำได้โดยการเปิดโปรแกรมขึ้นมา แล้วจะพบข้อความที่ขึ้นในโปรแกรม ดังรูปที่ 2.8 และเมื่อเข้าโปรแกรมพร้อมใช้งานแล้วที่หน้าจอจะปรากฏเครื่องหมาย ?- ซึ่งเป็นการแสดงว่าตัวโปรแกรมพร้อมที่จะรับคำสั่งจากผู้ใช้งานแล้ว



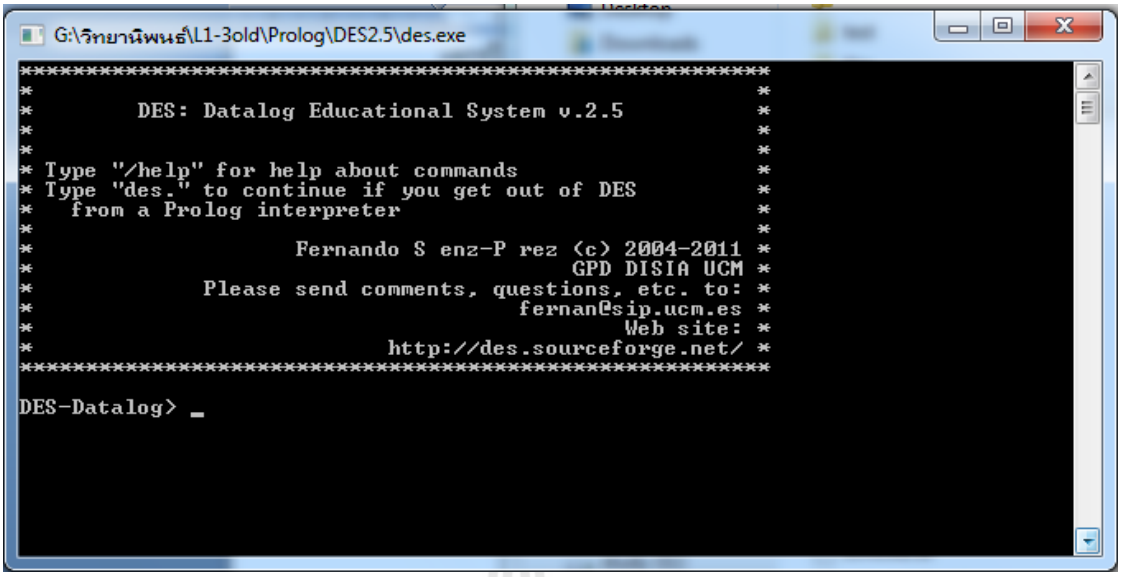
รูปที่ 2.8 ตัวอย่างแสดงความพร้อมของโปรแกรม

ในส่วนของงานวิจัยนี้จะใช้โปรแกรม SWI-Prolog ในการแปลภาษาในส่วนของการทำเหมืองข้อมูล และในส่วนของการแปลงรูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูลให้อยู่ในรูปแบบของกฎในแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย

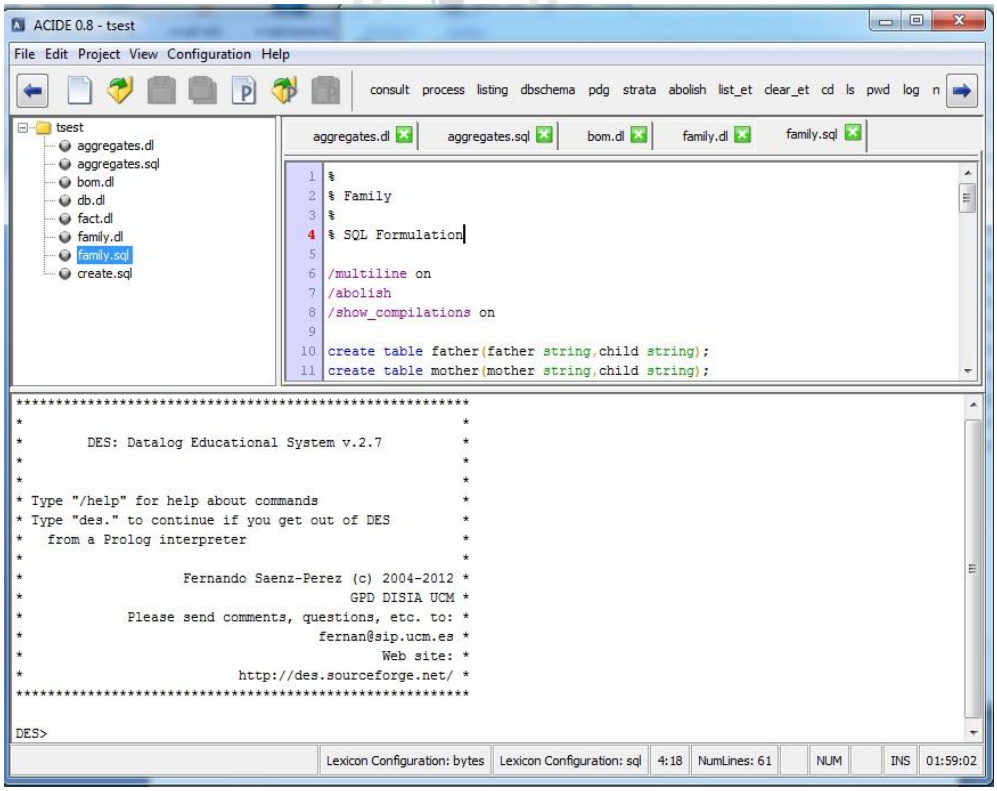
2.3.6 โปรแกรม DES

DES ย่อมาจากคำว่า Datalog Educational System โปรแกรม DES นี้เป็นโปรแกรมโอเพนซอร์ส สามารถดาวน์โหลดได้จาก http://www.fdi.ucm.es/profesor/fernan/des/

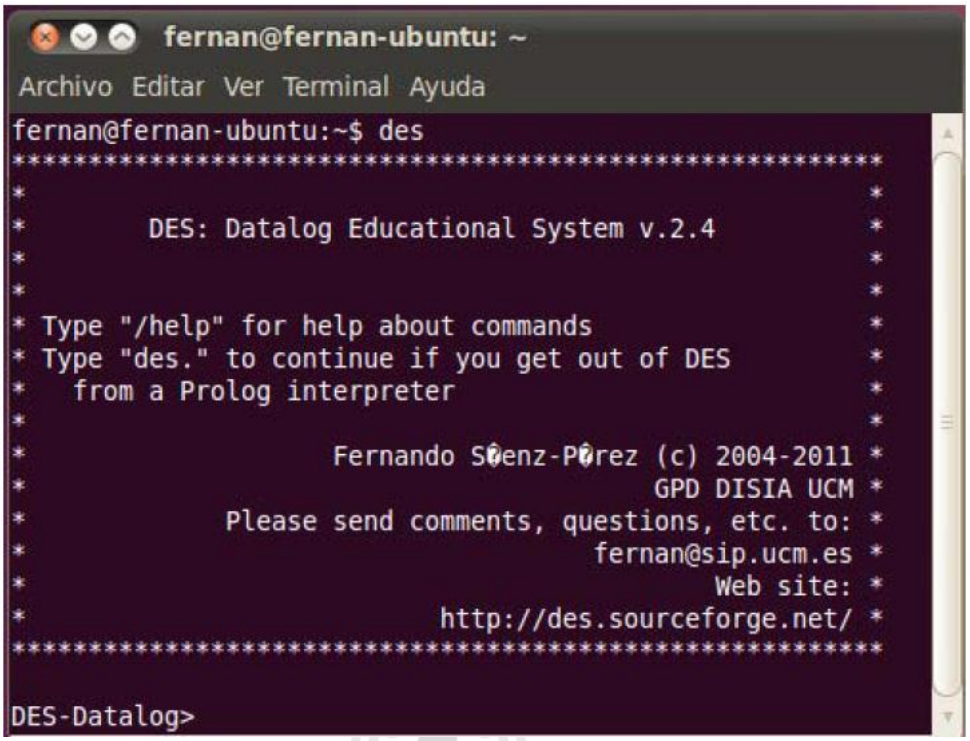
โปรแกรม DES สามารถรองรับทุกระบบปฏิบัติการ ไม่ว่าจะเป็น Windows Linux และ Mac จอภาพเริ่มต้นของ DES ในแต่ละระบบปฏิบัติการ แสดงได้ดังรูปที่ 2.9, 2.10, 2.11 และรูปที่ 2.12



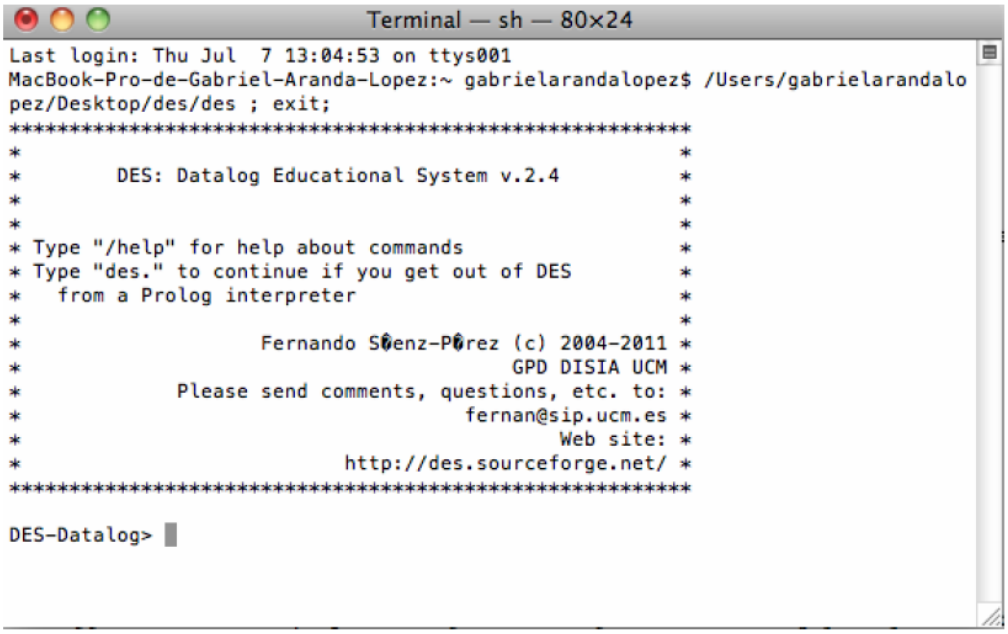
รูปที่ 2.9 ตัวอย่างโปรแกรม DES ในระบบวินโดว์ แบบ Command line



รูปที่ 2.10 ตัวอย่างโปรแกรม DES+ACIDE ในระบบวินโดว์



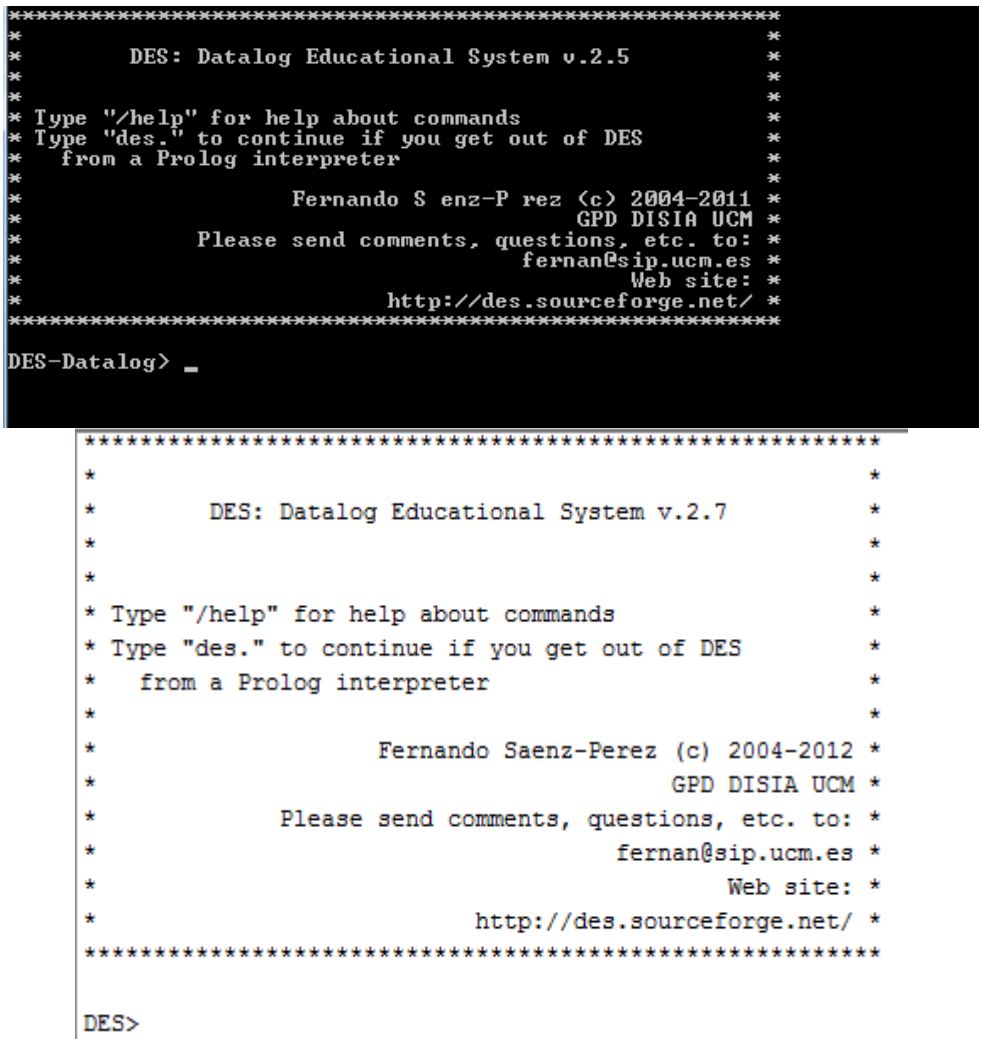
รูปที่ 2.11 ตัวอย่างโปรแกรม DES ในระบบลีนุกซ์



รูปที่ 2.12 ตัวอย่างโปรแกรม DES ในระบบ OS MAC

เริ่มต้นการใช้งานโปรแกรม DES

เมื่อทำการเปิดโปรแกรมขึ้นนั้น โปรแกรมจะอยู่ในโหมดของ Datalog เป็นค่าเริ่มต้น ดังรูปที่ 2.13 ที่แสดงโปรแกรมในโหมดเริ่มต้น



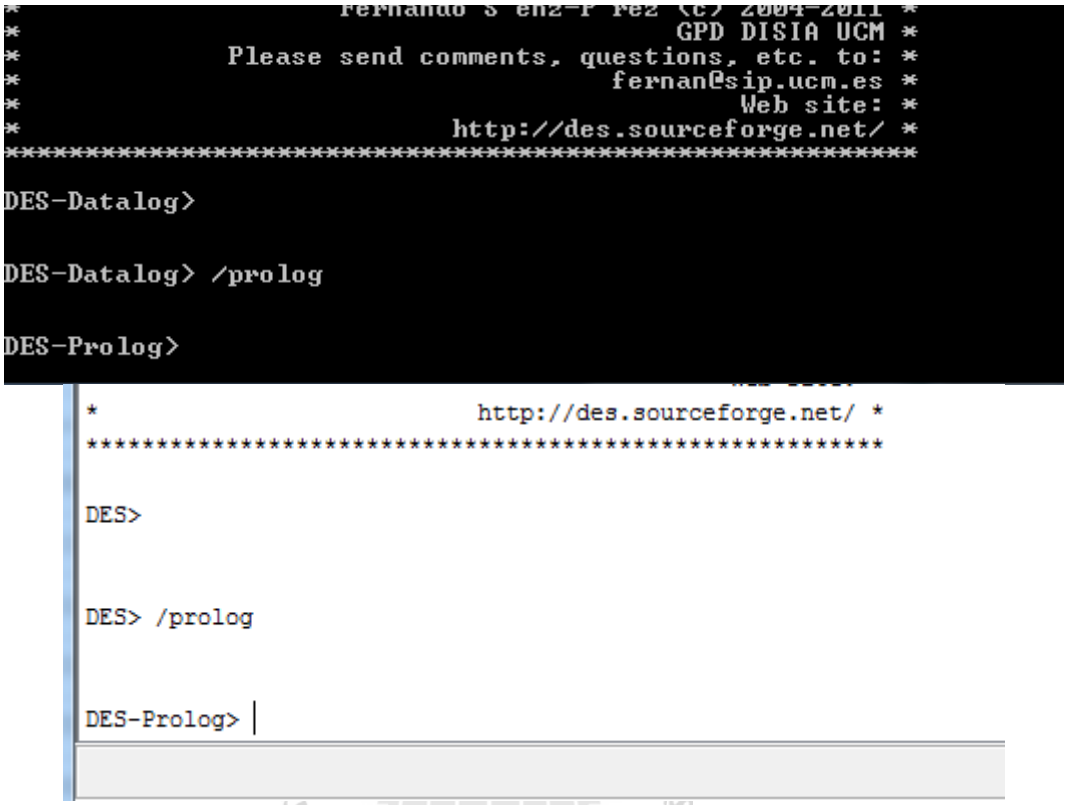
รูปที่ 2.13 ตัวอย่างโปรแกรม DES โหมด Datalog

การใช้งานโปรแกรม DES นั้นภายในโปรแกรมจะแบ่งออกเป็น 3 โหมดคือ โหมดของ Datalog โหมด Prolog และโหมด SQL ซึ่งในการเปลี่ยนโหมดทั้ง 3 ในโปรแกรม DES นั้นสามารถทำได้โดยการพิมพ์คำสั่ง ดังนี้

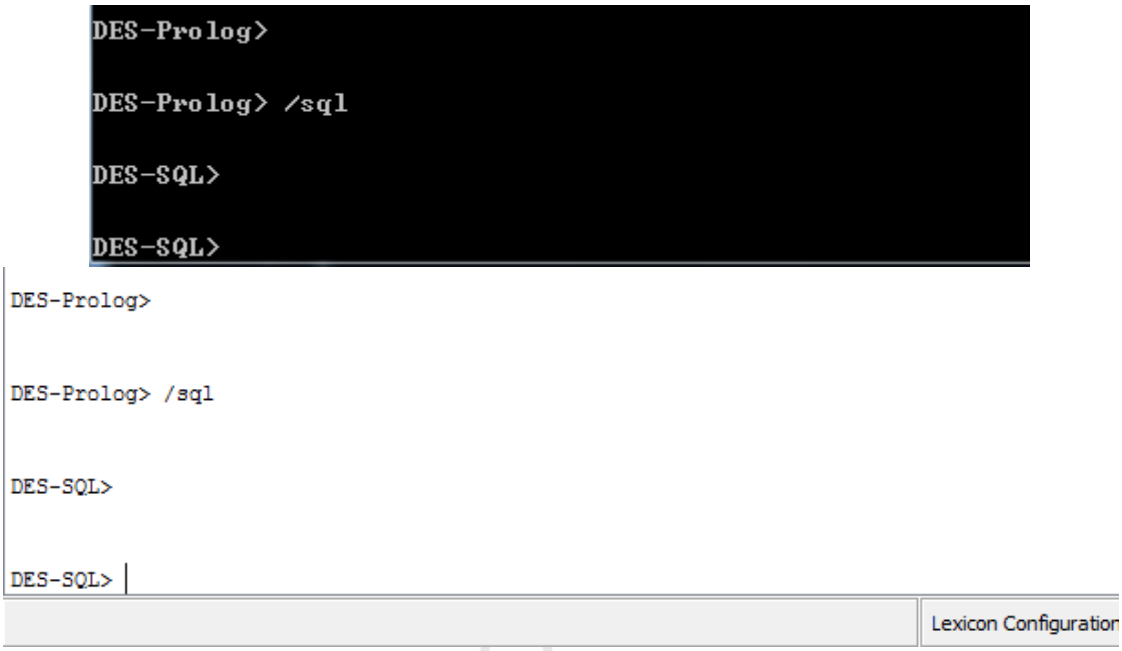
- โหมด Datalog ใช้คำสั่ง /datalog
- โหมด Prolog ใช้คำสั่ง /prolog

– โหมด SQL ใช้คำสั่ง /sql

จากรูปที่ 2. 14 และรูปที่ 2.15 จะเป็นการแสดงการเปลี่ยนโหมดของโปรแกรม จาก Datalog เป็น Prolog และจาก Prolog เป็น SQL ตามลำดับ



รูปที่ 2.14 ตัวอย่างของคำสั่งในการเปลี่ยนโหมดเป็น Prolog



รูปที่ 2.15 ตัวอย่างของคำสั่งในการเปลี่ยนโหมดเป็น SQL

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.4 งานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

ในการวิจัยเรื่องการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ผู้วิจัยได้ทำการศึกษาค้นคว้างานวิจัยในอดีตที่มีความเกี่ยวข้องกับการออกแบบแนวคิดในการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัยดังสรุปในตารางที่ 2.1 และมีรายละเอียดดังนี้

Sander Goos (2009) ได้นำเสนอเกี่ยวกับความถูกต้องของการรวมฐานข้อมูลในรูปแบบของโปรล็อก (Prolog) วิธีการของเขาจะทำการแปลงข้อมูลจากฐานข้อมูลปกติให้อยู่ในรูปแบบของภาษาโปรล็อกโดยการสร้างฟังก์ชันขึ้นมาในการแปลงแต่ละส่วน คือ ฟังก์ชันในการแปลงข้อมูลในฐานข้อมูล (T_D) ฟังก์ชันในการแปลงวิว (T_V) ฟังก์ชันในการแปลงการสอบถามข้อมูล (T_Q) และฟังก์ชันในการแปลงข้อมูลกลับ (T'_D) จากนั้นเขาได้ทำการตรวจสอบถึงความถูกต้องของการแปลงฐานข้อมูลนี้ พบว่ามีทั้งข้อดีและข้อเสีย ข้อดีคือในการแปลงของเขาจะมีฟังก์ชัน (T'_D) ในการตรวจสอบถึงความถูกต้องของฐานข้อมูลที่ได้ทำการแปลงแล้ว ส่วนข้อเสียคือ ข้อมูลที่ซ้ำกันนั้น ในภาษาโปรล็อกจะทำการสร้างขึ้นเพียงข้อมูลเดียว ส่งผลให้ความถูกต้องของข้อมูลในฐานข้อมูลนั้นไม่ถูกต้อง ดังนั้นวิธีการตรวจสอบความความถูกต้องของงานวิจัยนี้จึงต้องมีสองวิธีที่จะรองรับทั้งสองกรณีที่เป็นไปได้

Kazem Taghva และ Jayalakshmi Jayaraman (2008) ได้เสนอเกี่ยวกับการเพิ่มการวนซ้ำของการสอบถามข้อมูล (Query) สำหรับระบบสารสนเทศ ในงานวิจัยจะกล่าวถึงลักษณะของการสอบถามข้อมูล การกำหนดความสัมพันธ์ของข้อมูลในรูปแบบปกติและในรูปแบบของดาต้าล็อก เพื่อแสดงให้เห็นว่าในรูปแบบของดาต้าล็อกนี้การกำหนดความสัมพันธ์ที่ได้จะมีขนาดที่สั้นกว่าและช่วยลดเวลาในการประมวลผลทำให้ได้ผลลัพธ์เร็วขึ้น

Ignacis Blanco และคณะ (2009) ได้นำเสนอเกี่ยวกับการนำคุณสมบัติของภาษาดาต้าล็อกมาประยุกต์ใช้งานร่วมกับฐานข้อมูล GEFRED โมเดล GEFRED ถูกเสนอขึ้นมาโดย Medina และคณะ ซึ่งโมเดลนี้เสนอขึ้นมาเพื่อช่วยในการขยายความสามารถของโมเดล Relational ให้มีความสามารถในการจัดการและจัดเก็บข้อมูลที่มีความคลุมเครือและข้อมูลที่มีความไม่สมบรูณ์ จากการรับข้อมูลหรือจากความสัมพันธ์ของข้อมูล เพื่อที่จะทำให้ข้อมูลที่มีความคลุมเครือเหล่านี้สามารถถูกนำออกมาใช้งานได้ ในงานวิจัยเขาจะทำการพิจารณาจากโมเดล GEFRED และใช้วิธีการสร้างกฎของข้อมูลร่วมกับค่าความยืนหยุ่นของข้อมูล ในงานวิจัยจะกล่าวถึงลักษณะของรูปแบบ GEFRED และการใช้ภาษาดาต้าล็อกสร้างกฎของข้อมูลแล้วนำมาประยุกต์ให้เข้ากับ ฐานข้อมูลเชิงสัมพันธ์แบบคลุมเครือในรูปแบบของ GEFRED

Joseph Fong และคณะ (2001) ได้นำเสนอถึงการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นเอกสารในรูปแบบ XML โดยทำการออกแบบการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานข้อมูล XML เพื่อที่จะนำไปใช้งานผ่านเว็บ ขั้นตอนในการแปลงโครงสร้างนั้นจะใช้รูปแบบ EER ร่วมกับโครงสร้างของ XML แล้วนำไปแสดงในเอกสาร XML ในลักษณะของโครงสร้างต้นไม้

จากการทบทวนวรรณกรรมเกี่ยวข้องกับงานวิจัยนี้พบว่า งานวิจัยส่วนมากจะมีการนำภาษาโปรล็อกเข้ามาประยุกต์ใช้งานกับระบบฐานข้อมูล เนื่องจากคุณสมบัติของภาษามีส่วนช่วยในการลดขนาดของข้อมูลและเพิ่มความเร็วในการประมวลผลของการสอบถามข้อมูล หรือค้นหาข้อมูลในฐานข้อมูล บางงานวิจัยได้นำภาษาดาต้าล็อกเข้ามาประยุกต์ใช้งาน ซึ่งภาษาโปรล็อกและดาต้าล็อกนั้นมีลักษณะรูปแบบเดียวกัน เพียงแต่ภาษาดาต้าล็อกเป็นสับเซตของในภาษาโปรล็อก และไม่มีความซับซ้อนของการวนซ้ำ (Recursion) เหมาะที่จะนำมาจัดการกับฐานข้อมูล จากที่ได้ทำการศึกษางานวิจัยพบว่ามีหลายงานวิจัยที่ศึกษาเกี่ยวกับการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานข้อมูล XML แต่รูปแบบในการแปลงก็จะมีวิธีการที่ต่างกันออกไป ในปัจจุบันภาษาโปรล็อกเริ่มมีผู้สนใจมากขึ้น และนำมาประยุกต์ใช้งานกับการจัดการกับฐานข้อมูลเพื่อเพิ่มความสามารถให้ระบบจัดการฐานข้อมูล ผู้วิจัยจึงมีความสนใจที่จะทำการแปลงฐานข้อมูลเชิงสัมพันธ์ให้อยู่ในรูปแบบของฐานข้อมูลนิรนัย โดยใช้ภาษาโปรล็อกและภาษาดาต้าล็อก พร้อมทั้งยังนำเทคนิคทำเหมืองข้อมูล (Data Mining) เข้ามาประยุกต์ใช้งานร่วมด้วย เพื่อช่วยสร้างข้อมูลในรูปแบบ IDB ซึ่งจะเพิ่มประสิทธิภาพให้กับการสอบถามหรือค้นหาข้อมูลในฐานข้อมูล โดยในงานวิจัยนี้จะใช้ทั้งภาษาโปรล็อกในการเขียนอัลกอริทึมการทำเหมืองข้อมูลและใช้ภาษาดาต้าล็อกในการแปลงฐานข้อมูลที่จะกล่าวถึงขั้นตอนและวิธีการดำเนินการต่อไปในบทที่ 3

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.1สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

บทความวิจัยที่เกี่ยวข้องประกอบด้วย 1 = Sander Goos (2009), 2 = Kazem Taghvaและ Jayalakshmi Jayaraman (2008), 3 = Ignacis Blanco และคณะ (2009), 4 = Joseph Fong และคณะ (2001), 5 = การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย (งานวิจัยของวิทยานิพนธ์ฉบับนี้)

กระบวนการทำงาน	**งานวิจัยที่เกี่ยวข้อง**				
	1	**2**	**3**	**4**	**5**
ฐานความรู้เริ่มต้นของระบบ					
ประเภทของฐานข้อมูล					
ฐานความรู้ที่ใช้ได้จาก Relational Database	✓			✓	✓
ฐานความรู้ที่ใช้ได้จาก Information System		✓			
ฐานความรู้ที่ใช้ได้จาก GEFRED Database			✓		
การพัฒนา					
พัฒนาโดยใช้รูปแบบของภาษาโปรล็อก	✓		✓		✓
พัฒนาโดยใช้รูปแบบของภาษาดาต้าล็อก		✓			✓
พัฒนาโดยใช้รูปแบบของภาษา XML				✓	
ฐานความรู้ที่ได้					
ประเภทของฐานข้อมูล					
ฐานข้อมูลประเภท Prolog Database	✓		✓		
ฐานข้อมูลประเภท Dedustive Database		✓			✓
ฐานข้อมูลประเภท XML Document				✓	
เทคนิคในการแปลงข้อมูล					
แปลงข้อมูลจากโครงสร้างของข้อมูล				✓	
แปลงข้อมูลจากข้อมูลในตาราง	✓	✓	✓		✓
แปลงข้อมูลจากกฎ (Rule) ที่ได้จากการทำเหมืองข้อมูล					✓

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.1สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย (ต่อ)

กระบวนการทำงาน	**งานวิจัยที่เกี่ยวข้อง**				
	1	**2**	**3**	**4**	**5**
การประยุกต์ใช้ระบบ					
วิจัยเพื่อทดสอบประสิทธิภาพเท่านั้น	✓	✓	✓		✓
วิจัยเพื่อทดสอบความถูกต้อง	✓				✓
วิจัยเพื่อประยุกต์ใช้กับข้อมูลจริง				✓	

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 3
วิธีดำเนินการวิจัย

ในบทนี้จะนำเสนอถึง วิธีการวิจัย เครื่องมือที่ใช้ในการวิจัย และกระบวนการต่าง ๆ ของการวิจัย โดยมีรายละเอียดดังนี้

3.1 วิธีการวิจัย

ในการวิจัยนี้ จะมีวิธีการในการดำเนินการแยกออกเป็นขั้นตอนต่าง ๆ ดังนี้

3.1.1 ศึกษาและวิเคราะห์ถึงปัญหาในการลดเวลาของการค้นหาข้อมูลหรือสอบถามข้อมูล

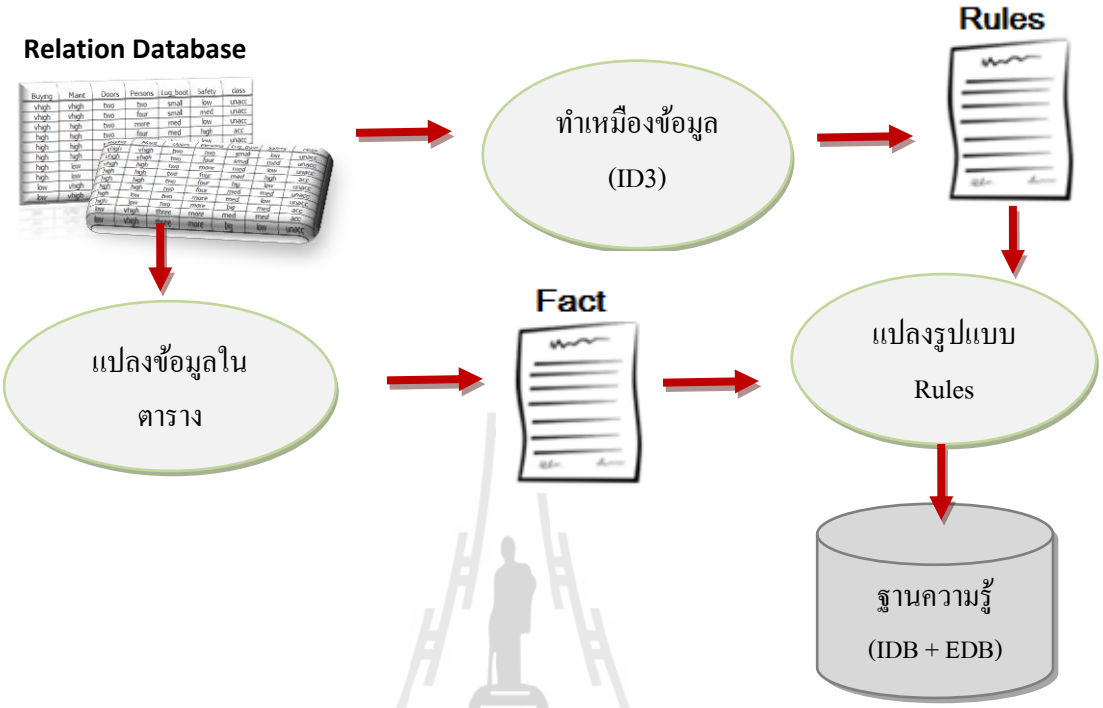
ในส่วนของการประมวลผลในการค้นหาข้อมูลของฐานข้อมูลเชิงสัมพันธ์ ในการค้นหาจะกระทำการค้นหาทีละแถวของข้อมูลที่มีอยู่ ทำให้เวลาในการค้นหานั้นจะขึ้นอยู่กับจำนวนของข้อมูล หากข้อมูลมีจำนวนมากเวลาในการค้นหาก็จะนานขึ้นไปด้วย ในงานวิจัยนี้จึงได้มีแนวคิดในการแปลงฐานข้อมูลเชิงสัมพันธ์ให้เป็นฐานข้อมูลนิรนัย ซึ่งฐานข้อมูลนิรนัยนี้เป็นฐานข้อมูลที่จะประกอบไปด้วยส่วนของข้อมูล และส่วนของรูปแบบของข้อมูล (Pattern) ที่ได้จากการทำเหมืองข้อมูลทำให้ช่วยลดเวลาในการค้นหาข้อมูล เนื่องจากเมื่อเราทำการค้นหาข้อมูล ระบบจะมาทำการตรวจสอบที่รูปแบบของข้อมูลก่อน

3.1.2 ออกแบบและพัฒนากรอบแนวคิดของการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย

แนวคิดการแปลงฐานข้อมูลเชิงสัมพันธ์ในงานวิจัยนี้ จะเน้นไปที่ส่วนของการแปลงข้อมูลในฐานข้อมูลเชิงสัมพันธ์ให้อยู่ในรูปแบบของฐานข้อมูลนิรนัย โดยข้อมูลจากตารางในฐานข้อมูลเชิงสัมพันธ์แต่ละแถวจะถูกแปลงให้เป็น Fact และ วิว จะถูกแปลงให้อยู่ในรูปแบบของ Rule ในงานวิจัยนี้ยังนำเอาการทำเหมืองข้อมูลเข้ามาช่วยในการเพิ่มรูปแบบของข้อมูลในฐานข้อมูลเข้าไปด้วย เพื่อช่วยให้การค้นหาข้อมูลหรือสอบถามข้อมูลเป็นไปได้รวดเร็วขึ้น

แนวคิดในการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย มีการทำงานหลัก 3 ส่วน คือ 1. ส่วนของการจัดเตรียมข้อมูลโดยการ Join ตารางในฐานข้อมูลให้เป็นตารางเดียวกัน 2. ส่วนของการทำเหมืองข้อมูล และ 3. ส่วนของการแปลงข้อมูลให้สามารถนำไปใช้

ในฐานข้อมูลนิรนัยส่วนประกอบทั้งหมดนี้แสดงเป็นแผนภาพได้ดังรูปที่ 3.1



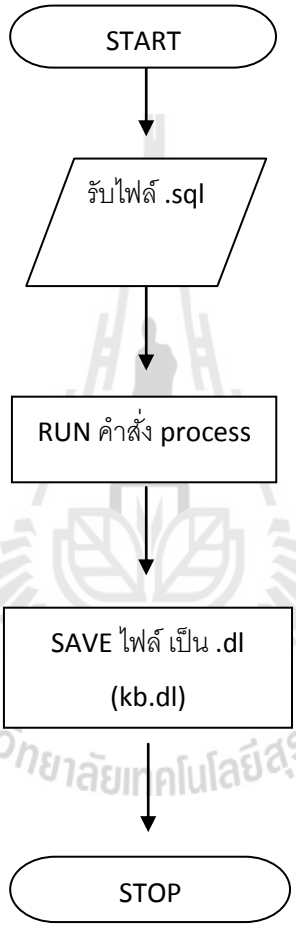
รูปที่ 3.1 กรอบแนวคิดของการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานข้อมูลนิรนัย

1) ส่วนของการจัดเตรียมข้อมูล คือการนำข้อมูลจากฐานข้อมูลเชิงสัมพันธ์ที่ประกอบด้วยหลายตารางมา Join กันและจัดรูปแบบให้ตรงตามรูปแบบที่จะใช้ในการทำเหมืองข้อมูลในขั้นตอนต่อไป ซึ่งส่วนนี้จะประกอบด้วยหลายขั้นตอนย่อย ไม่ว่าจะเป็นการรวบรวมข้อมูล การตัดข้อมูลที่ไม่จำเป็นออก แล้วจัดการแปลงให้อยู่ในรูปแบบที่เหมาะสม

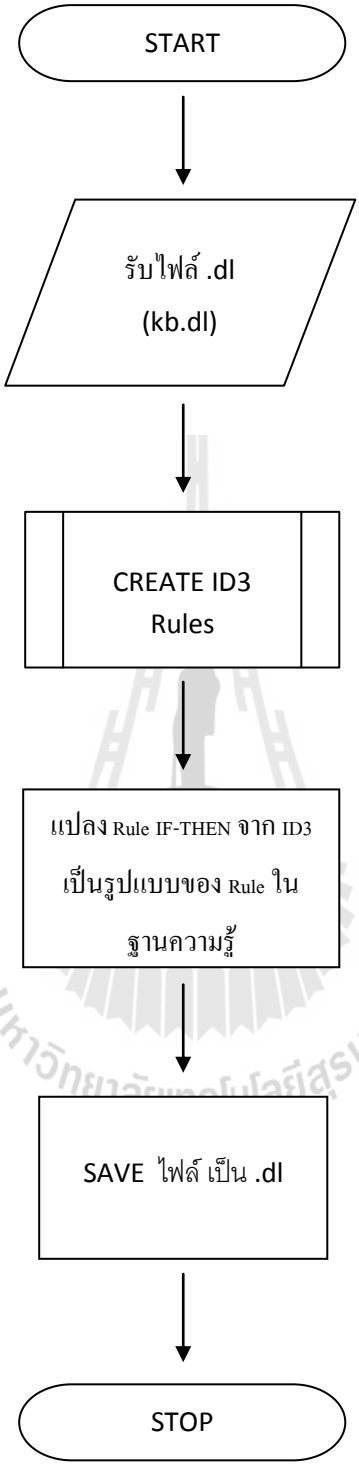
2) ส่วนของการทำเหมืองข้อมูล (Data Mining) ในส่วนนี้ผู้วิจัยได้ใช้อัลกอริทึม ID3 (Quinlan, 1979) และพัฒนาเป็นโปรแกรมโดยใช้ภาษาโปรล็อก อัลกอริทึมนี้จะเป็นการหาความสัมพันธ์ของข้อมูลจากการตัดสินใจแบบต้นไม้ (Decision Tree) และนำเอาความสัมพันธ์ที่ได้ส่งไปยังส่วนของการแปลงข้อมูล

3) ส่วนของการแปลงฐานข้อมูลเป็นฐานข้อมูลนิรนัย ในส่วนนี้จะเป็นส่วนที่จะทำการแปลงข้อมูลในฐานข้อมูลเชิงสัมพันธ์จากทุก ๆ ตารางที่สนใจ วิวของข้อมูล และความสัมพันธ์ที่ได้จากการทำเหมืองข้อมูลให้อยู่ในรูปแบบของภาษาดาต้าล็อก คือจะได้ฐานข้อมูลที่มีแต่ Rule และ Fact เท่านั้น

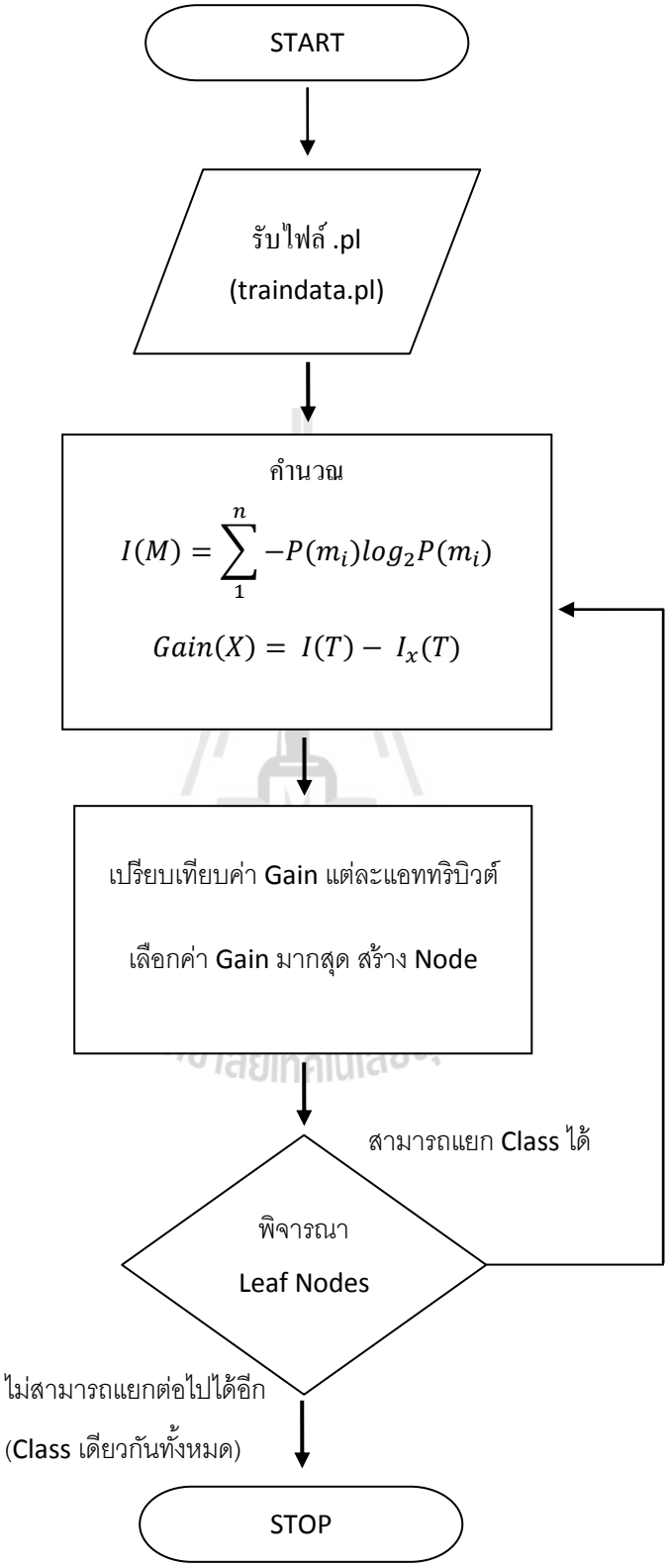
จากแนวคิดในการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานข้อมูลนิรนัย จะเป็นไปตามแผนผังการทำงานดังรูปที่ 3.2 รูปที่ 3.3 จะเป็นการแสดงในส่วนของการทำงานของการเพิ่มกฎที่ได้จากการทำเหมืองข้อมูล และรูปที่ 3.4 จะเป็นการแสดงถึงการหารูปแบบของข้อมูลโดยใช้วิธีการสร้างต้นไม้ตัดสินใจ



รูปที่ 3.2 แสดงแผนผังการทำงานในการแปลงข้อมูลเป็นฐานข้อมูลนิรนัย



รูปที่ 3.3 แสดงแผนผังการทำงานเพิ่มส่วนที่เป็นกฎที่ได้จากการทำเหมืองข้อมูล



รูปที่ 3.4 แสดงแผนผังการทำงานของการทำเหมืองข้อมูล

3.1.3 ทดสอบการทำงาน ประเมินผลและปรับปรุง

ในการประเมินและทดสอบประสิทธิภาพของการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย จะเน้นที่ความถูกต้องและเวลาที่ใช้ในการค้นหาข้อมูลหรือการสอบถามข้อมูลในฐานข้อมูล ดังนั้นในการประเมินประสิทธิภาพจะประเมินโดยการทำการทดสอบการค้นหาข้อมูลหรือการสอบถามข้อมูลจากฐานข้อมูลปกติเทียบกับฐานข้อมูลที่ได้ทำการแปลง แล้วทำการจับเวลาในการตอบคำตอบของทั้งสองฐานข้อมูล

3.2 เครื่องมือที่ใช้ในงานวิจัย

ในส่วนนี้จะกล่าวถึงเครื่องมือที่ใช้ในการพัฒนา โดยมีรายละเอียดดังนี้

3.2.1 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1) เครื่องคอมพิวเตอร์สำหรับการพัฒนา มีรายละเอียดดังนี้
   - หน่วยประมวลผลกลาง : Intel Core 2 Duo
   - หน่วยความจำสำรอง : 160 GB
   - หน่วยความจำหลัก : 4 GB
   - อุปกรณ์เสริมอื่น ๆ เช่น เมาส์ แป้นพิมพ์ เป็นต้น
2) ระบบปฏิบัติการและโปรแกรมประยุกต์สำหรับการพัฒนา ประกอบไปด้วย
   - ระบบปฏิบัติการ : Windows 7 Home Basic 64 bit
   - โปรแกรม : SWI-Prolog, EditPlus
   - เครื่องมือที่ใช้ในการพัฒนา
     - เครื่องมือพัฒนาส่วนของอัลกอริทึม : Prolog และ Datalog
     - โปรแกรมในการทดสอบการทำงาน : DES

3.2.2 เครื่องมือที่ใช้ในการประเมินประสิทธิภาพ

เครื่องมือที่ใช้ในการประเมินประสิทธิภาพของระบบ คือ การบันทึกเวลาส่วนของการทดสอบการสอบถาม (Query) ของฐานข้อมูลเชิงสัมพันธ์ และส่วนของการสอบถามข้อมูล (Query) จากฐานข้อมูลนิรนัยที่ได้จากการแปลงข้อมูลของงานวิจัย แล้วทำการเก็บค่าของเวลาในการประมวลผลของการตอบคำถาม มาทำการเปรียบเทียบกัน

3.3 วิธีการแปลงฐานข้อมูลไปเป็นฐานความรู้ในฐานข้อมูลนิรนัย

ในส่วนนี้จะเป็นการนำเสนอเกี่ยวกับวิธีการที่นำมาใช้ในการแปลงฐานข้อมูลเป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย ในส่วนของการแปลงฐานข้อมูลนั้น จะใช้วิธีการสร้างข้อมูลขึ้นมาในรูปแบบของ Fact ในภาษาดาต้าล็อก และจากนั้นจะนำเอาวิธีการทำเหมืองข้อมูลเข้ามาช่วยในการสร้างรูปแบบของข้อมูลในฐานความรู้ โดยรูปแบบจะอยู่ในลักษณะของ Rule

3.3.1 อัลกอริทึมในการทำเหมืองข้อมูล

ในงานวิจัยนี้ มีอัลกอริทึมในการทำเหมืองข้อมูลใช้ในการค้นหาความสัมพันธ์ของข้อมูล เพื่อที่จะนำความสัมพันธ์ที่ได้มาสร้างเป็นกฎในฐานความรู้เพื่อใช้ในฐานข้อมูลเชิงนิรนัย

อัลกอริทึมที่ใช้นี้จะเป็นอัลกอริทึม ID3 (ยุทธพงษ์ ปลื้มภิรมย์, 2549) เป็นอัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคที่นิยมใช้กันมากที่สุด ซึ่งวิธีการสร้างต้นไม้ตัดสินใจแบบ ID3 จะใช้ค่าเกนของสารสนเทศ (Information Gain) ในการพิจารณาเลือกโหนดที่จะสร้างในต้นไม้วิธีการในการคำนวณหาค่าเกนจะสามารถคำนวณได้จากสมการทั้ง 2 สมการต่อไปนี้ คือ

$$I(M) = \sum_{1}^{n} -P(m_i) log_2 P(m_i) \qquad (1)$$

$$Gain(X) = I(T) - I_x(T) \qquad (2)$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

Buying	Maint	Doors	Persons	Lug_boot	Safety	class
vhigh	vhigh	2	2	small	low	unacc
vhigh	vhigh	2	4	small	med	unacc
vhigh	high	2	more	med	low	unacc
high	high	2	4	med	high	acc
high	high	2	4	big	low	unacc
high	high	2	4	med	med	unacc
high	low	2	more	med	low	unacc
high	low	2	more	big	med	acc
low	vhigh	3	more	med	med	acc
low	vhigh	3	more	big	low	unacc

รูปที่ 3.5 ตัวอย่างตารางข้อมูล

จากรูปที่ 3.5 จะเป็นตัวอย่างของตารางข้อมูลที่นำมาแสดงวิธีการคำนวณการสร้างต้นไม้ตัดสินใจ โดยใช้ค่าเกนในการพิจารณาเลือกว่าแอททริบิวต์ใดจะเป็นโหนดในต้นไม้ตัดสินใจ ข้อมูลจะประกอบไปด้วยข้อมูล 7 แอททริบิวต์ 10 เรคคอร์ด แอททริบิวต์ที่สนใจจากข้อมูลตัวอย่าง คือ แอททริวบิวต์ class ที่ประกอบด้วย acc และ unacc

ในการคำนวณ จะพิจารณาข้อมูลไปทีละแอททริบิวต์ โดยทำการคำนวณหาค่าเกนของแต่ละแอททริบิวต์ จากนั้นทำการเปรียบเทียบเลือกแอททริบิวต์ที่มีค่าเกนมากที่สุด มาเป็นโหนดในต้นไม้ตัดสินใจ

ขั้นตอนแรกในการสร้างต้นไม้ตัดสินใจ คือ หาแอททริบิวต์ที่เป็นโหนดที่อยู่บนสุด (Root Node) เริ่มต้นจะคำนวณหาค่า Info ของแอททริบิวต์ที่สนใจ คือ Class

$$I(class) = \sum_{1}^{n} -P(m_i) log_2 P(m_i)$$

= [— (ความถี่ของข้อมูลคลาส unacc/จำนวนข้อมูลทั้งหมด) x $\log_2$(ความถี่ของข้อมูลคลาส unacc/จำนวนข้อมูลทั้งหมด)]+ [— (ความถี่ของข้อมูลคลาส acc/จำนวนข้อมูลทั้งหมด) x $\log_2$(ความถี่ของข้อมูลคลาส acc/จำนวนข้อมูลทั้งหมด)]

= — (7/10) x $\log_2$ (7/10) – (3/10) x $\log_2$ (3/10)

= 0.881

ต่อมาทำการหาค่า Info ของแต่ละแอททริบิวต์ที่เหลืออยู่อีก 6 แอททริบิวต์ ซึ่งในการคำนวณค่า Info ของแต่ละแอททริบิวต์ จะต้องคำนวณสัดส่วนข้อมูลในทุกค่าที่เป็นไปได้ของแอททริบิวต์นั้น เพื่อประกอบการเลือก ดังตัวอย่างการคำนวณต่อไปนี้

$I(Buying)$ = (3/10) [– (3/3) x $\log_2$ (3/3) – (0/3) x $\log_2$ (0/3)]
+ (5/10) [– (2/5) x $\log_2$ (2/5) – (3/5) x $\log_2$ (3/5)]
+ (2/10) [– (1/2) x $\log_2$ (1/2) – (1/2) x $\log_2$ (1/2)]
= 0.685

$I(Maint)$ = (4/10) [– (3/4) x $\log_2$ (3/4) – (1/4) x $\log_2$ (1/4)]
+ (4/10) [– (3/4) x $\log_2$ (3/4) – (1/4) x $\log_2$ (1/4)]
+ (2/10) [– (1/2) x $\log_2$ (1/2) – (1/2) x $\log_2$ (1/2)]

$= 0.849$

$I(Doors) = (8/10)\ [-(6/8) \times \log_2(6/8) - (2/8) \times \log_2(2/8)]$
$+ (2/10)\ [-(1/2) \times \log_2(1/2) - (1/2) \times \log_2(1/2)]$
$= 0.849$

$I(Persons) = (1/10)\ [-(1/1) \times \log_2(1/1) - (0/1) \times \log_2(0/1)]$
$+ (4/10)\ [-(3/4) \times \log_2(3/4) - (1/4) \times \log_2(1/4)]$
$+ (5/10)\ [-(3/5) \times \log_2(3/5) - (2/5) \times \log_2(2/5)]$
$= 0.809$

$I(Lug_boot) = (2/10)\ [-(2/2) \times \log_2(2/2) - (0/2) \times \log_2(0/2)]$
$+ (5/10)\ [-(3/5) \times \log_2(3/5) - (2/5) \times \log_2(2/5)]$
$+ (3/10)\ [-(2/3) \times \log_2(2/3) - (1/3) \times \log_2(1/3)]$
$= 0.761$

$I(Safety) = (5/10)\ [-(5/5) \times \log_2(5/5) - (0/5) \times \log_2(0/5)]$
$+ (4/10)\ [-(2/4) \times \log_2(2/4) - (2/4) \times \log_2(2/4)]$
$+ (1/10)\ [-(0/1) \times \log_2(0/1) - (1/1) \times \log_2(1/1)]$
$= 0.4$

เมื่อทำการคำนวณหาค่า Info ของแต่ละแอททริบิวต์แล้ว สามารถคำนวณหาค่าเกนจากการเลือกแต่ละแอททริบิวต์ดังนี้

$$Gain(X) = I(T) - I_x(T)$$

$\text{Gain}_{(\text{Buying})} = 0.881 - 0.685$
$= 0.196$

$\text{Gain}_{(\text{Maint})} = 0.881 - 0.849$
$= 0.032$

$\text{Gain}_{(\text{Doors})} = 0.881 - 0.849$
$= 0.032$

$\text{Gain}_{(\text{Persons})} = 0.881 - 0.809$

= 0.072

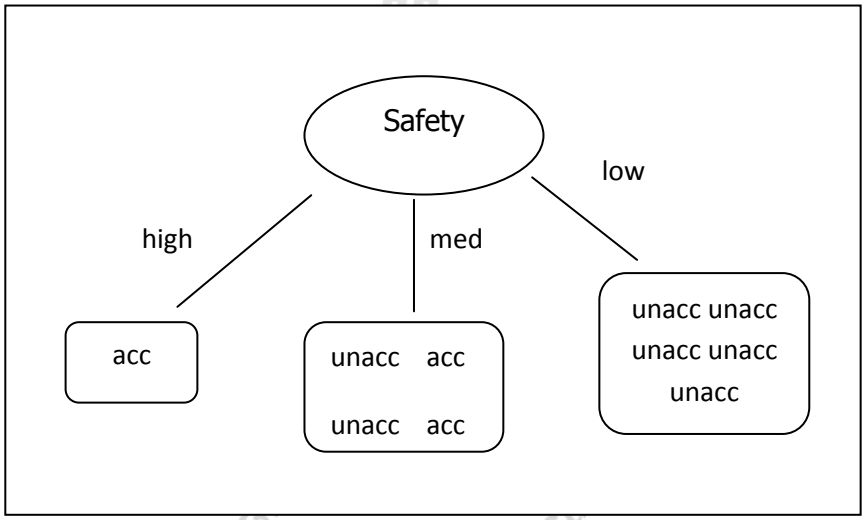
Gain($_{Lug_boot}$) = 0.881 – 0.761

= 0.120

Gain($_{Safety}$) = 0.881 – 0.4

= 0.481

จากค่าเกนของแต่ละแอททริบิวต์ พบว่าแอททริบิวต์ $_{Safety}$ มีค่ามากที่สุด ดังนั้นจะทำการเลือกแอททริบิวต์ $_{Safety}$ เป็นโหนดแรกสุด ดังรูปที่ 3.6



รูปที่ 3.6 แสดงตัวอย่างต้นไม้ตัดสินใจในขั้นแรก

จากรูปที่ 3.6 จะพบว่าต้นไม้ที่ได้เมื่อแอททริบิวต์ $_{Safety}$ เป็นโหลดแรก ต้นไม้ที่ได้จะมีกิ่งแตกออกมาสามกิ่ง เมื่อพิจารณาแต่ละกิ่งพบว่า ส่วนที่ $_{Safety}$=med นั้น ยังสามารถแยกต่อไปได้อีก เพื่อให้ต้นไม้ที่ได้สามารถแยกคลาสได้เป็นคลาสเดียวกันทั้งหมด ดังนั้นจากรูปยังจะต้องทำการคำนวณหาค่าเกนของแต่ละแอททริบิวต์ที่อยู่ในเงื่อนไขของ $_{Safety}$=med ต่ออีก เพื่อสร้างโหนดในต้นไม้ตัดสินใจ เพื่อให้สามารถแยกข้อมูลออกเป็นคลาสเดียวกันทั้งหมด

$I(med)$ = – (2/4) x $\log_2$ (2/4) – (2/4) x $\log_2$ (2/4)

= 1

พิจารณาค่า Info ของแต่ละแอททริบิวต์ที่เหลือและสามารถนำมาสร้างโหนดภายใต้ขอบเขตของ Safety=med

$I(Buying) =$ (1/4) [– (1/1) x $\log_2$ (1/1) – (0/1) x $\log_2$ (0/1)]
+ (2/4) [– (1/2) x $\log_2$ (1/2) – (1/2) x $\log_2$ (1/2)]
+ (1/4) [– (0/1) x $\log_2$ (0/1) – (1/1) x $\log_2$ (1/1)]
= 0.5

$I(Maint) =$ (2/4) [– (1/2) x $\log_2$ (1/2) – (1/2) x $\log_2$ (1/2)]
+ (1/4) [– (1/1) x $\log_2$ (1/1) – (0/1) x $\log_2$ (0/1)]
+ (1/4) [– (0/1) x $\log_2$ (0/1) – (1/1) x $\log_2$ (1/1)]
= 0.5

$I(Doors) =$ (3/4) [– (2/3) x $\log_2$ (2/3) – (1/3) x $\log_2$ (1/3)]
+ (1/4) [– (0/1) x $\log_2$ (0/1) – (1/1) x $\log_2$ (1/1)]
= 0.689

$I(Persons) =$ (0/4) [– (0/0) x $\log_2$ (0/0) – (0/0) x $\log_2$ (0/0)]
+ (2/4) [– (2/2) x $\log_2$ (2/2) – (0/2) x $\log_2$ (0/2)]
+ (2/4) [– (0/2) x $\log_2$ (0/2) – (2/2) x $\log_2$ (2/2)]
= 0

$I(Lug_boot) =$ (1/4) [– (1/1) x $\log_2$ (1/1) – (0/1) x $\log_2$ (0/1)]
+ (2/4) [– (1/2) x $\log_2$ (1/2) – (1/2) x $\log_2$ (1/2)]
+ (1/4) [– (0/1) x $\log_2$ (0/1) – (1/1) x $\log_2$ (1/1)]
= 0.5

จากนั้นเมื่อได้ค่า Info แล้วนำมาหาค่าเกนของแต่ละแอททริบิวต์

$Gain(X) = I(T) - I_x(T)$

Gain$_{(Buying)}$ = 1 – 0.5
= 0.5

$Gain(_{Maint})$ = 1 – 0.5

= 0.5

$Gain(_{Doors})$ = 1 – 0.689

= 0.311

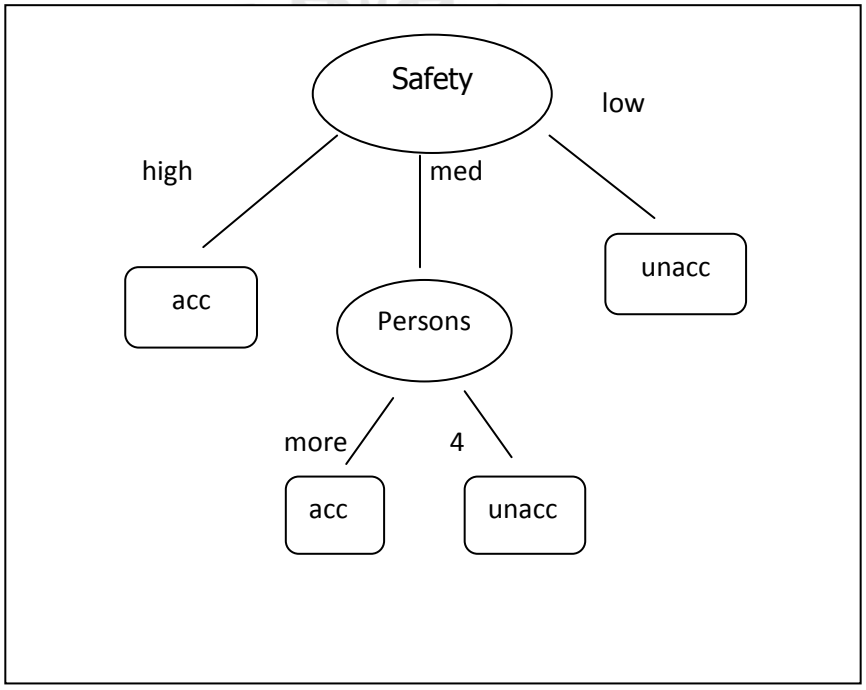
$Gain(_{Persons})$ = 1 – 0

= 1

$Gain(_{Lug_boot})$ = 1 – 0.5

= 0.5

จากค่าเกนที่ได้ พบว่าแอททริบิวต์ที่มีค่าเกนมากที่สุดคือ Persons ดังนั้นจะทำการเลือกแอททริบิวต์นี้มาเป็นโหนดลูกต่อจากโหนด Safety=med ต้นไม้ตัดสินใจจะเป็นไปตามรูปที่ 3.7



รูปที่ 3.7 แสดงตัวอย่างต้นไม้ตัดสินใจ

จากรูปที่ 3.7 จะเห็นว่าในต้นไม้ทุกโหนดสามารถแยกคลาสออกเป็นคลาสเดียวกันได้หมด แสดงว่าเป็นการสิ้นสุดการคำนวณหาต้นไม้สินใจ แล้วเมื่อได้ต้นไม้ตัดสินใจดังรูปที่ 3.7 แล้วสามารถที่จะแปลงต้นไม้ดังกล่าวให้อยู่ในรูปแบบกฎของรูปแบบข้อมูลได้ดังนี้

rule 1 : IF (safety = high) then class = acc

rule 2 : IF (safety = med) AND (persons = more) THEN class = acc

rule 3 : IF (safety = med) AND (persons = 4) THEN class = unacc

rule 4 : IF (safety = low) then class = unacc

เมื่อได้รูปแบบของข้อมูลดังกล่าวในลักษณะของกฎแล้ว จะนำรูปแบบที่ได้เข้าไปเพิ่มในฐานความรู้ที่ได้จากการแปลงฐานข้อมูลนิรนัย โดยจะนำรูปแบบของข้อมูลที่ได้ไปแปลงให้อยู่ในรูปแบบของประโยคในฐานข้อมูลนิรนัย ซึ่งจะกล่าวถึงในหัวข้อถัดไป

3.3.2 อัลกอริทึมในการแปลงข้อมูล

ในงานวิจัยที่ทำการศึกษาเป็นการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ในหัวข้อนี้จึงจะกล่าวถึงอัลกอริทึมที่ทำหน้าที่ในการแปลงจากฐานข้อมูลเชิงสัมพันธ์ให้อยู่ในรูปแบบของแหล่งความรู้ ในส่วนของการแปลงข้อมูลนี้ผู้วิจัยจะใช้ภาษาดาต้าล็อกในการแปลงข้อมูลผ่านโปรแกรม DES

ในการแปลงข้อมูลเป็นแหล่งความรู้ในงานวิจัย จะทำการแปลงข้อมูล ทั้งส่วนที่เป็นข้อมูลในตาราง และส่วนของรูปแบบข้อมูลหรือกฎที่ได้จากการทำเหมืองข้อมูล

(1) ส่วนของการแปลงข้อมูลจากตาราง

ในส่วนของการแปลงข้อมูลจากตารางในฐานข้อมูลเชิงสัมพันธ์ไปเป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ส่วนนี้เป็นการแปลงข้อมูลจากตารางในฐานข้อมูล โดยใช้การ Query ข้อมูลในตารางออกมา แล้วนำมาสร้างเป็นตารางใหม่ในโปรแกรม DES โดยใช้ภาษา Datalog ข้อมูลที่ได้จากข้อมูลในตารางความสัมพันธ์อยู่ในรูปแบบของ Fact

ตัวอย่างในการแปลงข้อมูลจากตารางในฐานข้อมูลเชิงสัมพันธ์ไปเป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัยจากรูปที่ 3.5 จะเป็นข้อมูลตัวอย่างจากตารางในฐานข้อมูลเชิงสัมพันธ์ จากนั้นจะทำการแปลงข้อมูลในตารางให้อยู่ในรูปแบบของ Fact โดยการใช้ภาษาดาต้าล็อกในการแปลงตารางข้อมูล ในการแปลงจะทำการเปลี่ยนข้อมูลในตารางให้อยู่ในรูปแบบของภาษาเอสคิวแอลก่อน ดังรูปที่ 3.8

```
/sql

create table data(buying string,maint string,doors string,persons string,lugboot
string,safety string,careva string);


insert into data values('vhigh','vhigh','two','two','small','low','unacc');

insert into data values('vhigh','vhigh','two','four','small','med','unacc');

insert into data values('vhigh','high','two','more','med','low','unacc');

insert into data values('high','high','two','four','med','high','acc');

insert into data values('high','high','two','four','big','low','unacc');

insert into data values('high','high','two','four','med','med','unacc');

insert into data values('high','low','two','more','med','low','unacc');

insert into data values('high','low','two','more','big','med','acc');

insert into data values('low','vhigh','three','more','med','med','acc');

insert into data values('low','vhigh','three','more','big','low','unacc');


select * from data;
```

รูปที่ 3.8 ตัวอย่างข้อมูลในรูปแบบเอสคิวแอล

จากนั้น เมื่อได้ข้อมูลเป็นเอสคิวแอลแล้ว จะนำไฟล์นี้ไปรันในโปรแกรม DES เพื่อให้ฟังก์ชันในโปรแกรมทำการสร้างข้อมูลออกมาเป็นข้อมูลในรูปแบบ Fact ซึ่งคำสั่งที่ใช้ในโปรแกรม DES คือ คำสั่ง /process จากนั้นโปรแกรมจะสร้างข้อมูลออกมาดังรูปที่ 3.9

```
data(high,high,two,four,big,low,unacc).
data(high,high,two,four,med,med,unacc).
data(high,high,two,four,med,high,acc).
data(high,low,two,more,big,med,acc).
data(high,low,two,more,med,low,unacc).
data(low,vhigh,three,more,big,low,unacc).
data(low,vhigh,three,more,med,med,acc).
data(vhigh,high,two,more,med,low,unacc).
data(vhigh,vhigh,two,four,small,med,unacc).
data(vhigh,vhigh,two,two,small,low,unacc).
:-type(data(buying:string,maint:string,doors:string,persons:string,
lugboot:string,safety:string,careva:string)).
```

รูปที่ 3.9 ข้อมูลที่อยู่ในรูปแบบ Fact

ในส่วนของงานวิจัยที่ทำการศึกษาจะทำการแปลงข้อมูลในตารางของข้อมูลเพื่อเปลี่ยนเป็นแหล่งความรู้ จากนั้นจะทำการเก็บบันทึกข้อมูลใหม่ไว้ที่ไฟล์ .dl เพื่อจัดเก็บเป็นแหล่งความรู้ในฐานข้อมูลนิรนัย

(2) ส่วนของการแปลงรูปแบบข้อมูลที่ได้จากอัลกอริทึมการทำเหมืองข้อมูล

ในส่วนนี้จะเป็นการรับรูปแบบของข้อมูลหรือกฎของข้อมูลที่ได้จากอัลกอริทึมการทำเหมืองข้อมูล ผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลจะได้รูปแบบของข้อมูลในลักษณะของ IF….THEN…. แต่การแปลงรูปแบบข้อมูลที่ได้เป็น Rule ในภาษา Datalog นั้น ลักษณะของ Rule ในภาษา Datalog จะเขียนอยู่ในรูปแบบตรงข้ามกับรูปแบบข้อมูลที่ได้จากการทำเหมืองข้อมูล คือ THEN….IF….

จากตัวอย่างข้อมูลในตาราง (รูปที่ 3.5) เมื่อนำข้อมูลมาทำเหมืองข้อมูลจะได้รูปแบบของข้อมูล ดังรูปที่ 3.10 ข้อมูลที่ได้นั้น จะนำมาทำการแปลงให้อยู่ในรูปแบบของภาษา

Datalog ซึ่งรูปแบบของการแปลงจะอยู่ในรูปแบบดังรูปที่ 3.11 จากรูปจะเห็นได้ว่าลักษณะของ Rule ที่จะมีลักษณะคล้ายกับรูปแบบข้อมูล เพียงแต่ส่วนของIF และ THEN จะสลับตำแหน่งกัน

```
if[safety=high]then[acc/1].
if[safety=med, persons=more]then[acc/2].
if[safety=med, persons=4]then[unacc/2].
if[safety=low]then[unacc/5].
```

รูปที่ 3.10 รูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล

```
rule(C1,C2,C3,C4,C5,C6,acc):-data(C1,C2,C3,C4,C5,C6),C6=high.
rule(C1,C2,C3,C4,C5,C6,acc):-data(C1,C2,C3,C4,C5,C6),C6=med,C4=more.
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=med,C4=4.
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=low.
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

รูปที่ 3.11 รูปแบบของข้อมูลในภาษาดาต้าล็อก

3.3.3 ลักษณะที่ใช้ในการทดสอบแหล่งความรู้

ลักษณะที่ใช้ในการทดสอบหาประสิทธิภาพของแหล่งความรู้ที่ได้จากการแปลงฐานข้อมูลเชิงสัมพันธ์ ได้ทำการทดสอบระหว่างแหล่งความรู้ 2 แบบว่าแบบใดที่จะมีประสิทธิภาพมากกว่า ซึ่งรูปแบบของแหล่งความรู้ที่ใช้ทั้ง 2 แบบ จะมีลักษณะดังนี้

1) แหล่งความรู้แบบที่ 1 ประกอบด้วยข้อมูล (Fact) เพียงอย่างเดียว ซึ่งจะมีลักษณะข้อมูลคล้ายกับฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

2) แหล่งความรู้แบบที่ 2 ประกอบด้วยข้อมูล (Fact) และกฎ (Rule) ที่ได้จากการทำเหมืองข้อมูล

โดยการออกแบบคำถามที่ใช้ในการสอบถามข้อมูลในฐานความรู้ให้ครอบคลุมกับข้อมูลที่มีทั้งหมด ไม่ว่าจะเป็นการสอบถามข้อมูลที่ตรงตามเงื่อนไข และสอบถามข้อมูลที่ไม่ตรงตามเงื่อนไขของรูปแบบข้อมูลที่ได้มาจากการทำเหมืองข้อมูล

3.3.4 ขั้นตอนการทดสอบ

ขั้นตอนในการแปลงฐานข้อมูลเป็นแหล่งความรู้ และทำการทดสอบแหล่งความรู้ ขั้นตอนแรกนั้น จะต้องทำการเตรียมข้อมูลที่จะใช้ในส่วนนี้ดังนี้

1) ส่วนของการทำเหมืองข้อมูล ต้องทำการจัดรูปแบบของข้อมูลให้อยู่ในรูปแบบที่เหมาะสม เพื่อที่จะนำข้อมูลที่ได้ไปใช้ในการหารูปแบบข้อมูล แล้วทำการบันทึกไว้ในไฟล์ .pl ลักษณะของรูปแบบข้อมูลที่จะใช้จะเป็นไปดังรูปที่ 3.12

2) ส่วนของข้อมูลที่จะทำการแปลงเป็นแหล่งความรู้ ทำการเตรียมข้อมูลให้อยู่ในรูปของภาษา SQL แล้วทำการบันทึกไว้ในไฟล์ .sql ดังตัวอย่างรูปที่ 3.13

```
example(1,unacc,[buying=vhigh,maint=vhigh,doors=2,persons=2,lugboot=small,safety=low]).
example(2,unacc,[buying=vhigh,maint=vhigh,doors=2,persons=4,lugboot=small,safety=med]).
example(3,unacc,[buying=vhigh,maint=high,doors=2,persons=more,lugboot=med,safety=low]).
example(4,acc,[buying=high,maint=high,doors=2,persons=4,lugboot=med,safety=high]).
example(5,unacc,[buying=high,maint=high,doors=2,persons=4,lugboot=big,safety=low]).
example(6,unacc,[buying=high,maint=high,doors=2,persons=4,lugboot=med,safety=med]).
example(7,unacc,[buying=high,maint=low,doors=2,persons=more,lugboot=med,safety=low]).
example(8,acc,[buying=high,maint=low,doors=2,persons=more,lugboot=big,safety=med]).
example(9,acc,[buying=low,maint=vhigh,doors=3,persons=more,lugboot=med,safety=med]).
example(10,unacc,[buying=low,maint=vhigh,doors=3,persons=more,lugboot=big,safety=low]).
```

รูปที่ 3.12 ตัวอย่างรูปแบบของข้อมูลที่ใช้ในการทำเหมืองข้อมูล

```
/sql

create table data(buying string,maint string,doors string,persons string,lugboot
string,safety string,careva string);


insert into data values('vhigh','vhigh','two','two','small','low','unacc');

insert into data values('vhigh','vhigh','two','four','small','med','unacc');

insert into data values('vhigh','high','two','more','med','low','unacc');

insert into data values('high','high','two','four','med','high','acc');

insert into data values('high','high','two','four','big','low','unacc');

insert into data values('high','high','two','four','med','med','unacc');

insert into data values('high','low','two','more','med','low','unacc');

insert into data values('high','low','two','more','big','med','acc');

insert into data values('low','vhigh','three','more','med','med','acc');

insert into data values('low','vhigh','three','more','big','low','unacc');


select * from data;
```

รูปที่ 3.13 ตัวอย่างรูปแบบของข้อมูลเอสคิวแอล

เมื่อทำการเตรียมข้อมูลทั้งสองเสร็จเรียบร้อยแล้ว ในงานวิจัยนี้จะแบ่งแหล่งความรู้ที่ได้เป็น 2 แบบ แหล่งความรู้แบบแรกนั้น จะเป็นเพียงข้อมูลที่อยู่ในรูปแบบของ Fact ซึ่งขั้นตอนนี้เริ่มจากการแปลงข้อมูลตัวอย่างที่ได้เตรียมไว้แล้วในไฟล์ .sql มาใช้งานในโปรแกรม DES โดยทำการรันไฟล์ .sql แล้วทำการบันทึกข้อมูลที่ได้จากการรันข้อมูลที่อยู่ในลักษณะของ Fact ไว้ในไฟล์ .pl สำหรับแหล่งความรู้แบบที่ 2 จะเป็นฐานความรู้ที่ประกอบไปด้วยข้อมูล (Fact) กับกฎ (Rule) ที่ได้จากการทำเหมืองข้อมูล ในส่วนของแหล่งความรู้แบบที่ 2 นี้ จะทำการรันไฟล์ ID3.pl

เพื่อให้ได้รูปแบบที่ได้จากการทำเหมืองและเพิ่มรูปแบบนี้เข้าไปในแหล่งความรู้ที่ได้จากการรันไฟล์ .sql

ในส่วนของไฟล์ ID3.pl นั้นเมื่อทำการรันขึ้นมา ตัวโปรแกรมจะทำการถามผู้ใช้งาน เพื่อใส่ข้อมูล คำถามแรก จะเป็นการถามถึงชื่อไฟล์ของข้อมูลที่ได้ทำการเตรียมไว้ในตอนต้น ต่อจากนั้นจะถามชื่อไฟล์ที่ได้จากการรันไฟล์ .sql และสุดท้ายจะถามชื่อไฟล์ที่ต้องการจะให้ไฟล์เป็นแหล่งความรู้

3.4 การเปรียบเทียบประสิทธิภาพของแหล่งความรู้

ในหัวข้อนี้จะกล่าวถึงการเปรียบเทียบประสิทธิภาพของแหล่งความรู้ที่ได้ทั้ง 2 แบบ ว่าแหล่งความรู้แบบใดจะมีประสิทธิภาพมากกว่ากัน ซึ่งในการเปรียบเทียบประสิทธิภาพนี้ จะนำเอาแหล่งความรู้ที่ได้สองแบบไปทำการทดสอบในโปรแกรม DES โดยการสอบถามข้อมูลจากแหล่งข้อมูลทั้งสองในลักษณะของคำถามที่เหมือนกัน ซึ่งลักษณะของคำถามที่ใช้ในการสอบถามข้อมูลจะมีลักษณะ

1) ตรงกับเงื่อนไขของรูปแบบข้อมูล
2) ไม่ตรงกับเงื่อนไขของรูปแบบข้อมูล

จากนั้นเมื่อได้รูปแบบของการคำถามที่จะใช้ในการสอบถาม จะทำการสอบถามไปยังแหล่งความรู้ทั้งสองแบบ แล้วพิจารณาเวลาที่ใช้ในการตอบคำถามของแหล่งความรู้ทั้งสองแบบ แล้วนำเวลาที่ได้มาทำการเปรียบเทียบวิเคราะห์และอภิปรายผล

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4
ผลการวิเคราะห์ข้อมูลและอภิปรายผล

การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ซึ่งในการแปลงฐานข้อมูลของงานวิจัยนี้ ทำการแปลงฐานข้อมูลออกมาเป็นแหล่งความรู้ 2 แบบ คือ แบบที่ 1 แหล่งความรู้ที่ได้จะประกอบไปด้วยข้อมูลเพียงอย่างเดียว และแบบที่ 2 แหล่งความรู้จะประกอบไปด้วยข้อมูลรวมกับรูปแบบของข้อมูลที่ได้มาจากการทำเหมืองข้อมูลโดยใช้เทคนิค ID3 ในการหารูปแบบหรือโมเดลของข้อมูล จากนั้นทำการทดสอบประสิทธิภาพของแหล่งความรู้ที่ได้จากการแปลงทั้งสองแบบว่า แหล่งความรู้แบบใดมีประสิทธิภาพมากกว่า การทดสอบประสิทธิภาพของแหล่งความรู้จะพิจารณาจากการใช้ระยะเวลาในการประมวลผลเพื่อค้นหาคำตอบของการป้อนแบบสอบถาม (Query) เข้าไป การสอบถามที่ผู้วิจัยได้ทำการป้อนจะแบ่งออกเป็นการสอบถามที่ตรงตามเงื่อนไขของรูปแบบข้อมูล และการสอบถามที่ไม่ตรงกับเงื่อนไขของรูปแบบข้อมูล การทดสอบจะใช้โปรแกรม DES เข้ามาช่วยในการอ่านไฟล์แหล่งความรู้ที่อยู่ในรูปแบบของภาษาดาต้าล็อก

สำหรับเนื้อหาที่ได้ทำการนำเสนอในบทนี้ จะเป็นการนำเสนอตั้งแต่ขั้นตอนการแปลงข้อมูลออกมาเป็นแหล่งความรู้ไปจนถึงขั้นตอนการทดสอบประสิทธิภาพของแหล่งความรู้ที่ได้ โดยมีรายละเอียดดังนี้ หัวข้อ 4.1 อธิบายถึงฐานข้อมูลที่ใช้ในการแปลงเป็นแหล่งความรู้ หัวข้อ 4.2 อธิบายถึงแหล่งความรู้ที่ได้จากการแปลงฐานข้อมูล หัวข้อ 4.3 อธิบายถึงผลการทดสอบประสิทธิภาพของแหล่งความรู้ และหัวข้อ 4.4 เป็นการอภิปรายสรุปผลการเปรียบเทียบการทดสอบแหล่งความรู้ทั้งสอง

4.1 ฐานข้อมูลที่ใช้ในการแปลงเป็นแหล่งความรู้

ในงานวิจัยฐานข้อมูลที่ใช้ในการแปลงเป็นแหล่งความรู้ จะนำข้อมูลตัวอย่างมาทำการแปลงไปเป็นแหล่งความรู้ ซึ่งข้อมูลที่ใช้ในงานวิจัยนี้จะใช้ข้อมูล 2 ชุดด้วยกันคือ ข้อมูลชุดแรกจะเป็นข้อมูลเกี่ยวกับการวิเคราะห์เกี่ยวกับรถ และข้อมูลชุดที่ 2 จะเป็นข้อมูลเกี่ยวการวิเคราะห์การจัดอันดับใบสมัครเพื่อคัดเลือกผู้สมัครเข้าเรียนในโรงเรียนอนุบาลของประเทศสโลวีเนีย (Nursery) ซึ่งข้อมูลที่ได้นำมาใช้ทั้ง 2 ชุดนี้ได้นำมาจากเว็บไซต์ http://archive.ics.uci.edu/ml/

ข้อมูลชุดแรกชื่อข้อมูล Car เป็นข้อมูลเกี่ยวกับการวิเคราะห์รถ ตัวอย่างข้อมูลดังรูปที่ 4.1 ซึ่งมีรายละเอียดดังนี้

- จำนวนข้อมูล 1728 เรคคอร์ด
- จำนวนแอททริบิวต์ 7 แอททริบิวต์ ได้แก่ buying, maint, doors, persons, lug_boot, safety และ class

Buying	Maint	Doors	Persons	Lug_boot	Safety	class
vhigh	vhigh	two	two	small	low	unacc
vhigh	vhigh	two	four	small	med	unacc
vhigh	high	two	more	med	low	unacc
high	high	two	four	med	high	acc
high	high	two	four	big	low	unacc
high	high	two	four	med	med	unacc
high	low	two	more	med	low	unacc
high	low	two	more	big	med	acc
low	vhigh	three	more	med	med	acc
low	vhigh	three	more	big	low	unacc
vhigh	vhigh	two	two	small	high	unacc
vhigh	vhigh	two	two	med	low	unacc
vhigh	vhigh	two	two	med	med	unacc
vhigh	vhigh	two	two	med	high	unacc
vhigh	vhigh	two	two	big	low	unacc
vhigh	vhigh	two	two	big	med	unacc
vhigh	vhigh	two	two	big	high	unacc
vhigh	vhigh	two	four	small	low	unacc
vhigh	vhigh	two	four	small	med	unacc
vhigh	vhigh	two	four	small	high	unacc
vhigh	vhigh	two	four	med	low	unacc
vhigh	vhigh	two	four	med	med	unacc
vhigh	vhigh	two	four	med	high	unacc
vhigh	vhigh	two	four	big	low	unacc
vhigh	vhigh	two	four	big	med	unacc

รูปที่ 4.1 ตัวอย่างตารางข้อมูลของชุดข้อมูลเกี่ยวกับรถ (Car)

ข้อมูลชุดที่ 2 ชื่อข้อมูล Nursery เป็นข้อมูลเกี่ยวการวิเคราะห์การจัดอันดับใบสมัครเพื่อคัดเลือกผู้สมัครเข้าเรียนในโรงเรียนอนุบาลของประเทศสโลวีเนีย ซึ่งตัวอย่างข้อมูลดังรูปที่ 4.2 มีรายละเอียดดังนี้

– จำนวนข้อมูล 12960 เรคคอร์ด

– จำนวนแอททริบิวต์ 9 แอททริบิวต์ ได้แก่ parents, has_nurs, form, children, housing, finance, social, health และ class

parents	has_nurs	form	children	housing	finance	social	health	class
usual	proper	complete	1	convenient	convenient	nonprob	recommended	recommend
usual	proper	complete	1	convenient	convenient	nonprob	priority	priority
usual	proper	complete	1	convenient	convenient	nonprob	not_recom	not_recom
usual	proper	complete	1	convenient	convenient	slightly_prob	recommended	recommend
usual	proper	complete	1	convenient	convenient	slightly_prob	priority	priority
usual	proper	complete	1	convenient	convenient	slightly_prob	not_recom	not_recom
usual	proper	complete	1	convenient	convenient	problematic	recommended	priority
usual	proper	complete	1	convenient	convenient	problematic	priority	priority
usual	proper	complete	1	convenient	convenient	problematic	not_recom	not_recom
usual	proper	complete	1	convenient	inconv	nonprob	recommended	very_recom
usual	proper	complete	1	convenient	inconv	nonprob	priority	priority
usual	proper	complete	1	convenient	inconv	nonprob	not_recom	not_recom
usual	proper	complete	1	convenient	inconv	slightly_prob	recommended	very_recom
usual	proper	complete	1	convenient	inconv	slightly_prob	priority	priority
usual	proper	complete	1	convenient	inconv	slightly_prob	not_recom	not_recom
usual	proper	complete	1	convenient	inconv	problematic	recommended	priority
usual	proper	complete	1	convenient	inconv	problematic	priority	priority
usual	proper	complete	1	convenient	inconv	problematic	not_recom	not_recom
usual	proper	complete	1	less_conv	convenient	nonprob	recommended	very_recom
usual	proper	complete	1	less_conv	convenient	nonprob	priority	priority
usual	proper	complete	1	less_conv	convenient	nonprob	not_recom	not_recom
usual	proper	complete	1	less_conv	convenient	slightly_prob	recommended	very_recom
usual	proper	complete	1	less_conv	convenient	slightly_prob	priority	priority
usual	proper	complete	1	less_conv	convenient	slightly_prob	not_recom	not_recom
usual	proper	complete	1	less_conv	convenient	problematic	recommended	priority
usual	proper	complete	1	less_conv	convenient	problematic	priority	priority
usual	proper	complete	1	less_conv	convenient	problematic	not_recom	not_recom
usual	proper	complete	1	less_conv	inconv	nonprob	recommended	very_recom
usual	proper	complete	1	less_conv	inconv	nonprob	priority	priority
usual	proper	complete	1	less_conv	inconv	nonprob	not_recom	not_recom
usual	proper	complete	1	less_conv	inconv	slightly_prob	recommended	very_recom
usual	proper	complete	1	less_conv	inconv	slightly_prob	priority	priority
usual	proper	complete	1	less_conv	inconv	slightly_prob	not_recom	not_recom
usual	proper	complete	1	less_conv	inconv	problematic	recommended	priority
usual	proper	complete	1	less_conv	inconv	problematic	priority	priority
usual	proper	complete	1	less_conv	inconv	problematic	not_recom	not_recom
usual	proper	complete	1	critical	convenient	nonprob	recommended	very_recom
usual	proper	complete	1	critical	convenient	nonprob	priority	priority
usual	proper	complete	1	critical	convenient	nonprob	not_recom	not_recom
usual	proper	complete	1	critical	convenient	slightly_prob	recommended	very_recom
usual	proper	complete	1	critical	convenient	slightly_prob	priority	priority
usual	proper	complete	1	critical	convenient	slightly_prob	not_recom	not_recom
usual	proper	complete	1	critical	convenient	problematic	recommended	priority

รูปที่ 4.2 ตัวอย่างตารางข้อมูลของชุดข้อมูลเกี่ยวกับการส่งเด็กไปยังสถานที่รับเลี้ยงเด็ก (Nursery)

4.2 แหล่งความรู้ที่ได้จากการแปลงฐานข้อมูล

แหล่งความรู้ที่ได้จากการแปลงฐานข้อมูลตัวอย่างในงานวิจัยนี้ จะทำการแปลงข้อมูลโดยใช้โปรแกรม DES และใช้ภาษาดาต้าล็อกในการช่วยแปลงข้อมูลให้อยู่ในรูปแบบของแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย แหล่งความรู้ที่ได้จากการแปลงฐานข้อมูล ผู้วิจัยได้จัดทำแหล่งความรู้ขึ้นมาสองรูปแบบ คือ แหล่งความรู้ที่เประกอบด้วยข้อมูล (Fact) เพียงอย่างเดียว และแหล่งความรู้ที่ประกอบด้วยข้อมูล (Fact) และกฎ (Rule) ซึ่งเป็นรูปแบบหรือโมเดลของข้อมูลที่ได้จากการทำเหมืองข้อมูล

แหล่งความรู้ สำหรับชุดข้อมูลที่ 1 ข้อมูลเกี่ยวกับรถ เป็นไปดังรูปที่ 4.3 แสดงแหล่งความรู้ที่ประกอบด้วยข้อมูลเพียงอย่างเดียว และรูปที่ 4.4 แหล่งความรู้ที่ประกอบไปด้วยข้อมูลและรูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล

```
data(high,high,four,more,big,high,acc).
data(high,high,more,more,med,high,acc).
data(high,high,more,more,small,high,acc).
data(high,high,more,two,small,high,unacc).
data(high,low,more,more,small,low,unacc).
data(high,med,more,four,big,med,acc).
        …
        …
data(high,vhigh,four,four,big,low,unacc).
data(low,vhigh,four,four,big,med,acc).
data(vhigh,med,two,two,big,med,unacc).
data(vhigh,vhigh,two,four,big,high,unacc).
data(vhigh,vhigh,two,two,small,med,unacc).
:-type(data(buying:string,maint:string,doors:string,persons:string,lugboot:string,
safety:string,careva:string)).
```

รูปที่ 4.3 แหล่งความรู้แบบที่ 1 ของข้อมูล Car

```
data(high,high,four,more,big,high).
data(high,high,more,more,med,high).
data(high,high,more,more,med,low).
data(high,high,more,more,small,high).
data(high,high,more,more,small,low).
data(high,high,more,more,small,med).
        …
        ...
data(vhigh,vhigh,two,two,big,low).
data(vhigh,vhigh,two,two,big,med).
data(vhigh,vhigh,two,two,med,low).
data(vhigh,vhigh,two,two,small,high).
data(vhigh,vhigh,two,two,small,low).
data(vhigh,vhigh,two,two,small,med).
:-type(data(buying:string,maint:string,doors:string,persons:string,lugboot:string,safety:string)).
rule(C1,C2,C3,C4,C5,C6,unacc):-
data(C1,C2,C3,C4,C5,C6),C6=med,C4=more,C1=high,C5=med,C3=two.
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=med,C4=more,C1=high,C5=small
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=high,C4=two.
rule(C1,C2,C3,C4,C5,C6,acc):-data(C1,C2,C3,C4,C5,C6),C6=high,C4=four,C1=low,C2=vhigh.
        …
        …
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=med,C4=four,C1=vhigh,C2=vhigh.
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=med,C4=two.
rule(C1,C2,C3,C4,C5,C6,unacc):-data(C1,C2,C3,C4,C5,C6),C6=low.
```

รูปที่ 4.4 แหล่งความรู้แบบที่ 2 ของข้อมูล Car

แหล่งความรู้ สำหรับชุดข้อมูลที่ 2 ข้อมูลเกี่ยวกับการวิเคราะห์ถึงการส่งเด็กไปยังสถานที่รับเลี้ยงเด็ก (Nursery) เป็นไปดังรูปที่ 4.5 ซึ่งแสดงแหล่งความรู้ที่มีเพียงแต่ข้อมูล และรูปที่ 4.6 แสดงแหล่งความรู้ที่ประกอบไปด้วยข้อมูลและรูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล

```
data(greatpret,critical,complete,more,convenient,convenient,nonprob,notrecom,notrecom).
data(greatpret,critical,complete,more,convenient,convenient,nonprob,priority,specprior).
data(greatpret,critical,complete,more,lessconv,convenient,nonprob,notrecom,notrecom).
data(greatpret,critical,complete,more,lessconv,convenient,nonprob,priority,specprior).
data(greatpret,proper,foster,one,lessconv,inconv,nonprob,notrecom,notrecom).
data(greatpret,proper,foster,one,lessconv,inconv,nonprob,priority,specprior).
data(pretentious,critical,completed,two,convenient,inconv,slightlyprob,notrecom,notrecom).
        ...
data(usual,critical,complete,two,critical,inconv,nonprob,priority,specprior).
data(usual,critical,complete,two,critical,inconv,nonprob,recommended,priority).
data(usual,improper,incomplete,one,critical,convenient,problematic,recommended,priority).
data(usual,improper,incomplete,one,critical,convenient,slightlyprob,notrecom,notrecom).
data(usual,verycrit,foster,one,critical,inconv,problematic,recommended,specprior).
data(usual,verycrit,foster,one,critical,inconv,slightlyprob,notrecom,notrecom).
data(usual,verycrit,foster,one,critical,inconv,slightlyprob,priority,specprior).
data(usual,verycrit,foster,one,critical,inconv,slightlyprob,recommended,specprior).
data(usual,verycrit,foster,one,lessconv,convenient,nonprob,notrecom,notrecom).
data(usual,verycrit,foster,one,lessconv,convenient,nonprob,priority,specprior).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,recommended,specprior).
:- :-type(data(parents:string,hasnurs:string,form:string,children:string,housing:string,
finance:string,social:string,health:string,nursary:string)).
```

รูปที่ 4.5 แหล่งความรู้แบบที่ 1 ของข้อมูล Nursery

```
data(greatpret,critical,complete,two,critical,inconv,problematic,priority).

data(greatpret,critical,complete,two,critical,inconv,problematic,recommended).

data(greatpret,critical,complete,two,critical,inconv,slightlyprob,notrecom).

        ...

data(pretentious,proper,completed,two,convenient,convenient,nonprob,notrecom).

data(pretentious,proper,completed,two,convenient,inconv,slightlyprob,priority).

data(pretentious,proper,completed,two,convenient,inconv,slightlyprob,recommended).

data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,notrecom).

data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,priority).

data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,recommended).

:-type(data(parents:string,hasnurs:string,form:string,children:string,housing:string,finance:string
,social:string,health:string)).

rule(C1,C2,C3,C4,C5,C6,C7,C8,priority):-
data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,C2=improper,C1=usual,C7=nonprob,C5=lessc
onv,C4=three.

rule(C1,C2,C3,C4,C5,C6,C7,C8,veryrecom):-
data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,C2=improper,C1=usual,C7=nonprob,C5=lessc
onv,C4=two,C3=complete.

        ..

        ..

rule(C1,C2,C3,C4,C5,C6,C7,C8,veryrecom):-
data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,C2=improper,C1=usual,C7=nonprob,C5=conve
nient,C6=convenient.

rule(C1,C2,C3,C4,C5,C6,C7,C8,specprior):-
data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,C2=lessproper,C7=problematic,C1=greatpret,C
5=critical,C4=three.

rule(C1,C2,C3,C4,C5,C6,C7,C8,specprior):-
data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,C2=lessproper,C7=problematic,C1=greatpret,C
5=critical,C4=two.
```

รูปที่ 4.6 แหล่งความรู้แบบที่ 2 ของข้อมูล Nursery

4.3 ผลการทดสอบประสิทธิภาพของแหล่งความรู้

การทดสอบประสิทธิภาพของแหล่งความรู้ที่ได้จากการแปลงฐานข้อมูล จะทำการทดสอบประสิทธิภาพของแหล่งความรู้โดยการสอบถามข้อมูลหรือการค้นหาข้อมูล ซึ่งคำถามที่ใช้ในการสอบถามข้อมูลจะเป็นคำถามที่ครอบคลุมกับข้อมูล คือการสอบถามข้อมูลที่ตรงตามเงื่อนไขและไม่ตรงตามเงื่อนไขของรูปแบบข้อมูล

4.3.1 การทดสอบกับข้อมูลในฐานข้อมูลเชิงสัมพันธ์

ตารางที่ 4.1 และตารางที่ 4.2 เป็นตารางแสดงรูปแบบคำถามที่ใช้ในการสอบถามข้อมูลของข้อมูลทั้งชุดที่ 1 และชุดที่ 2 ในรูปแบบของภาษาเอสคิวแอล ซึ่งทำการสอบถามข้อมูลในรูปแบบของข้อมูลที่เป็นฐานข้อมูลเชิงสัมพันธ์ เพื่อใช้ในการเปรียบเทียบกับแหล่งความรู้ทั้ง 2 แบบ ของงานวิจัยนี้ และแสดงตารางที่ 4.3 และ 4.4 เป็นการแสดงผลการทดสอบในการสอบถามข้อมูลทั้งที่เป็นการสืบค้นข้อมูลที่ตรงตามเงื่อนไขและไม่ตรงตามเงื่อนไขของรูปแบบข้อมูล

ตารางที่ 4.1 แสดงรูปแบบการสอบถามข้อมูล Car

รูปแบบ	Query ตรงตามเงื่อนไข
1	select * from data where careva = 'unacc'
2	select * from data where buying='low' and maint='low' and safety='high' and careva = 'vgood'
3	select * from data where buying='low' and maint='high' and persons='four' and lugboot = 'med' and careva = 'unacc'
4	select * from data where buying='low' and maint='high' and doors='two' and persons = 'four' and lugboot = 'med' and safety='low'and careva = 'unacc'
รูปแบบ	Query ไม่ตรงกับเงื่อนไข
1	select * from data where buying='low' and doors='five'
2	select * from data where buying='low' and maint='high' and doors='four' and persons = 'two' and lugboot = 'big' and safety='low' and careva = 'good'
3	select * from data where buying='med' and maint='med' and safety='high' and careva = 'good'
4	select * from data where careva = 'un'

ตารางที่ 4.2 แสดงรูปแบบการสอบถามข้อมูล Nursery

รูปแบบ	Query ตรงตามเงื่อนไข
1	select * from data where nursary= 'priority'
2	select * from data where parents = 'greatpret' and children = 'more' and finance ='inconv' and health ='priority'
3	select * from data where hasnurs = 'improper' and children = 'two' and housing ='critical' and finance ='inconv' and health ='priority' and nursary= 'specprior'
4	select * from data where parents = 'pretentious' and hasnurs = 'improper' and form = 'foster' and housing ='convenient' and social ='problematic' and health ='notrecom' and nursary = 'notrecom'
รูปแบบ	Query ไม่ตรงกับเงื่อนไข
1	select * from data where children = 'mo'
2	select * from data where finance = 'test' and health = 'notrecom'
3	select * from data where parents = 'usual' and hasnurs = 'improper' and form = 'completed' and children ='one' and finance= 'incomplete' and health ='notrecom' and nursary= 'specprior '
4	select * from data where parents = 'pretentious' and hasnurs = 'improper' and form = 'completed' and children ='three' and housing ='convenient' and finance= 'inconv' and social ='nonprob' and health ='notrecom' and nursery= ' veryrecom '

ตารางที่ 4.3 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Car

รูปแบบ	เวลา (ms)
Query แบบตรงกับ Rule	
แบบที่ 1	190
แบบที่ 2	27
แบบที่ 3	20
แบบที่ 4	25
Query แบบไม่ตรงกับ Rule	
แบบที่ 1	15
แบบที่ 2	7
แบบที่ 3	14
แบบที่ 4	16

ตารางที่ 4.4 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Nursery

รูปแบบ	เวลา (ms)
Query แบบตรงกับ Rule	
แบบที่ 1	60386
แบบที่ 2	4097
แบบที่ 3	1177
แบบที่ 4	430
Query แบบไม่ตรงกับ Rule	
แบบที่ 1	70
แบบที่ 2	67
แบบที่ 3	62
แบบที่ 4	63

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.3.2 การทดสอบประสิทธิภาพของข้อมูลชุดที่ 1

ตารางที่ 4.5 และตารางที่ 4.6 เป็นตารางแสดงรูปแบบคำถามที่ใช้ในการสอบถามข้อมูลของแหล่งความรู้ทั้ง 2 แบบ ซึ่งรูปแบบการสอบถามข้อมูลจะแบ่งออกเป็นการสอบถามที่ตรงตามเงื่อนไขของข้อมูลและไม่ตรงตามเงื่อนไขของข้อมูล และแสดงตารางที่ 4.7 เป็นการแสดงผลการทดสอบในการสอบถามข้อมูลทั้งที่เป็นการสืบค้นข้อมูลที่ตรงตามเงื่อนไขและไม่ตรงตามเงื่อนไขของรูปแบบข้อมูลในแต่ละแบบ

ตารางที่ 4.5 แสดงรูปแบบการสอบถามข้อมูล Car ที่ตรงตามเงื่อนไข

รูปแบบ	Query ของแหล่งความรู้แบบที่ 1	Query ของแหล่งความรู้แบบที่ 2
1	data(A,B,C,D,E,F,unacc).	rule(A,B,C,D,E,F,unacc).
2	data (low,low,C,D,E,high,vgood).	rule(low,low,C,D,E,high,vgood).
3	data (low,high,C,four,med,F,unacc).	rule(low,high,C,four,med,F,unacc).
4	data(low,high,two,four,med,low,unacc).	rule(low,high,two,four,med,low,unacc).

ตารางที่ 4.6 แสดงรูปแบบการสอบถามข้อมูล Car ที่ไม่ตรงตามเงื่อนไข

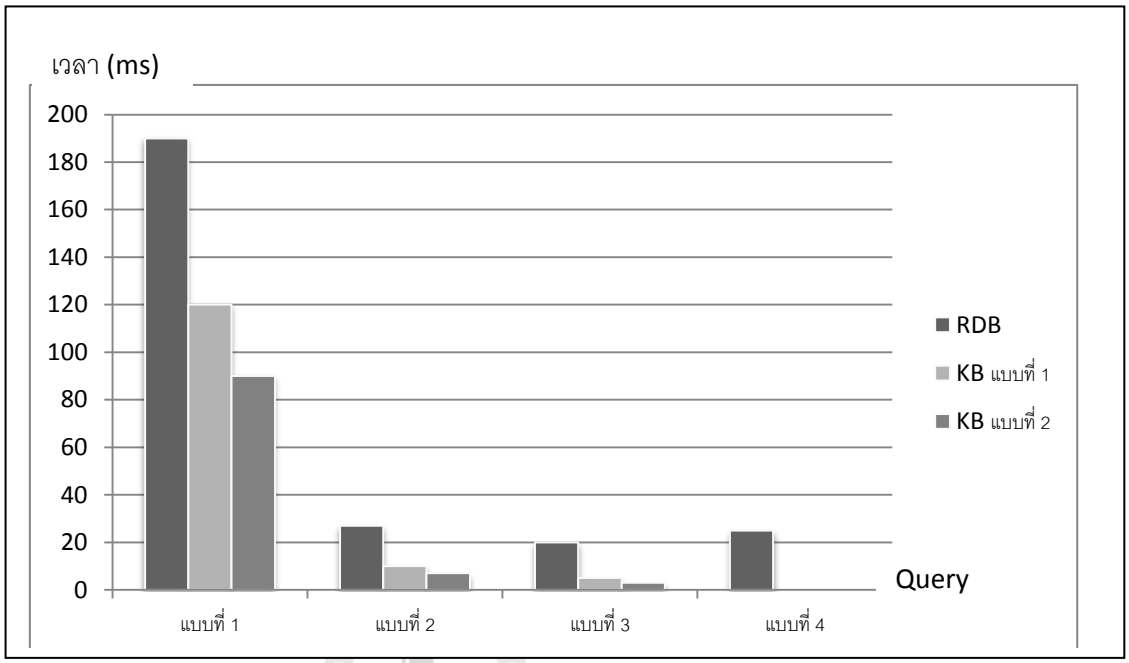
รูปแบบ	Query ของแหล่งความรู้แบบที่ 1	Query ของแหล่งความรู้แบบที่ 2
1	data (low,B,five,D,E,F,Re).	rule(low,B,five,D,E,F,Re).
2	data (low,high,four,two,big,low,good).	rule(low,high,four,two,big,low,good).
3	data (med,med,C,D,E,high,good).	rule(med,med,C,D,E,high,good).
4	data (A,B,C,D,E,F,un).	rule(A,B,C,D,E,F,un).

ตารางที่ 4.7 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Car

รูปแบบ/เวลา (ms)	KB แบบที่ 1	KB แบบที่ 2
Query แบบตรงกับ Rule		
แบบที่ 1	120	90
แบบที่ 2	10	7
แบบที่ 3	5	3
แบบที่ 4	0	0
Query แบบไม่ตรงกับ Rule		
แบบที่ 1	0	4
แบบที่ 2	0	3
แบบที่ 3	0	5
แบบที่ 4	0	4

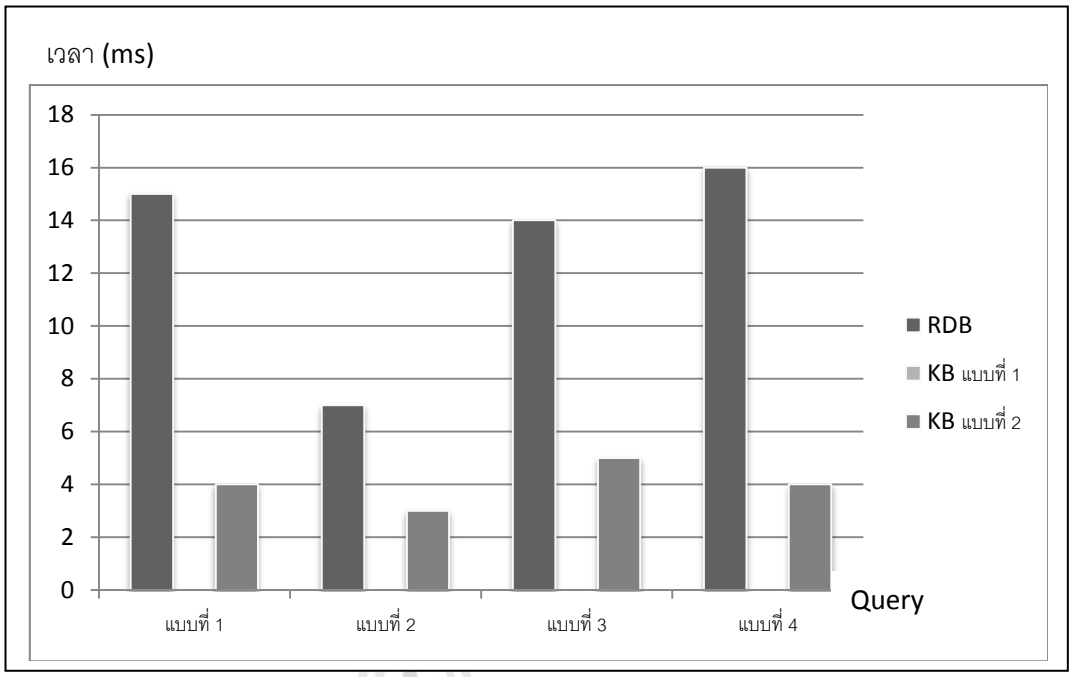
มหาวิทยาลัยเทคโนโลยีสุรนารี

เมื่อนำค่าที่ได้จากตารางมาวิเคราะห์ สามารถแสดงผลได้เป็นกราฟดังรูปที่ 4.7 และรูปที่ 4.8



รูปที่ 4.7 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Car ที่ตรงกับเงื่อนไข

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 4.8 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Car ที่ไม่ตรงกับเงื่อนไข

4.3.3 การทดสอบประสิทธิภาพของข้อมูลชุดที่ 2

ตารางที่ 4.8 และตารางที่ 4.9 เป็นตารางแสดงรูปแบบคำถามที่ใช้ในการสอบถามข้อมูลของแหล่งความรู้ทั้ง 2 แบบ ซึ่งรูปแบบการสอบถามข้อมูลจะแบ่งออกเป็นการสอบถามที่ตรงตามเงื่อนไขของข้อมูลและไม่ตรงตามเงื่อนไขของข้อมูล และแสดงตารางที่ 4.10 เป็นการแสดงผลการทดสอบในการสอบค้นข้อมูล

ตารางที่ 4.8 แสดงรูปแบบการสอบถามข้อมูล Nursery ที่ตรงตามเงื่อนไข

รูปแบบ	Query ของแหล่งความรู้แบบที่ 1	Query ของแหล่งความรู้แบบที่ 2
1	data(C1,C2,C3,C4,C5,C6,C7,C8, priority).	rule(C1,C2,C3,C4,C5,C6,C7,C8, priority).
2	data(greatpret,C2,C3,more,C5,inconv, C7,priority,R).	rule(greatpret,C2,C3,more,C5,inconv ,C7,priority,R).
3	data(C1,improper,C3,two,critical, inconv,C7,priority,specprior).	rule(C1,improper,C3,two,critical, inconv,C7,priority,specprior).
4	data(pretentious,improper,foster,C4 ,convenient,C6,problematic,notrecom, notrecom).	rule(pretentious,improper,foster,C4 ,convenient,C6,problematic,notrecom, notrecom).

ตารางที่ 4.9 แสดงรูปแบบการสอบถามข้อมูล Nursery ที่ไม่ตรงตามเงื่อนไข

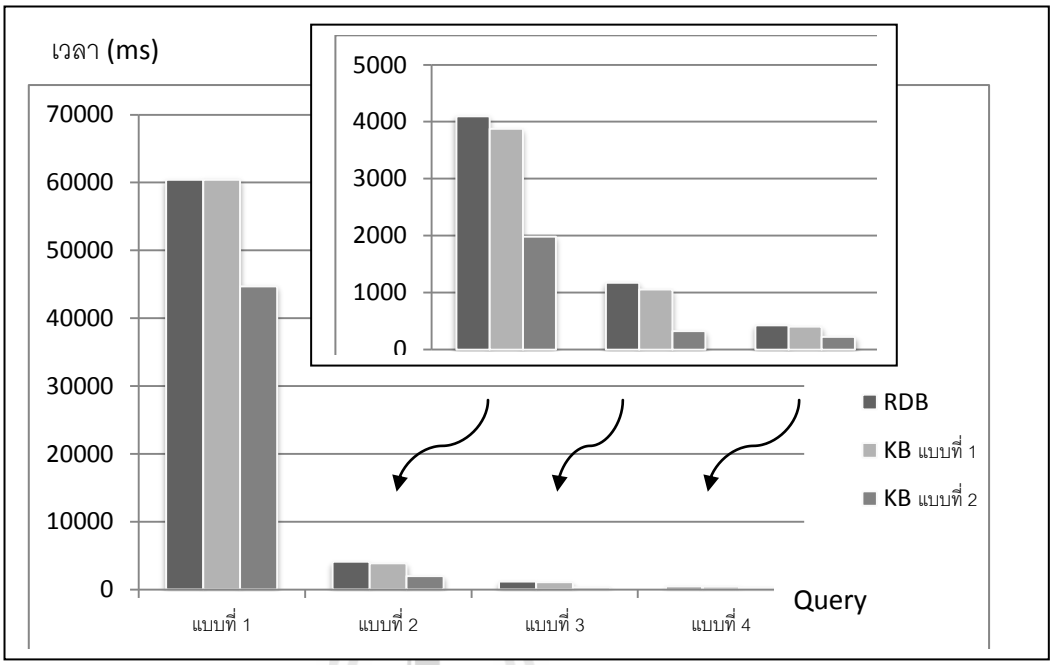
รูปแบบ	Query ของแหล่งความรู้แบบที่ 1	Query ของแหล่งความรู้แบบที่ 2
1	data(C1,C2,C3,mo,C5,C6,C7,C8,Re).	rule(C1,C2,C3,mo,C5,C6,C7,C8,Re).
2	data(C1,C2,C3,C4,C5,test,C7,notrecom,R).	rule(C1,C2,C3,C4,C5,test,C7,notrecom,R).
3	data(usual,improper,complete,one,C5, incomplete,C7,notrecom,specprior).	rule(usual,improper,complete,one,C5, incomplete,C7,notrecom,specprior).
4	data(pretentious,improper,completed,three, convenient,inconv,nonprob, notrecom,veryrecom).	rule(pretentious,improper,completed,three, convenient,inconv,nonprob, notrecom,veryrecom).

ตารางที่ 4.10 แสดงเวลาที่ใช้ในการสอบถามข้อมูล Nursery

รูปแบบ/เวลา (ms)	KB แบบที่ 1	KB แบบที่ 2
Query แบบตรงกับ Rule		
แบบที่ 1	60418	44685
แบบที่ 2	3877	1987
แบบที่ 3	1057	327
แบบที่ 4	403	225
Query แบบไม่ตรงกับ Rule		
แบบที่ 1	16	73
แบบที่ 2	16	81
แบบที่ 3	16	81
แบบที่ 4	15	75

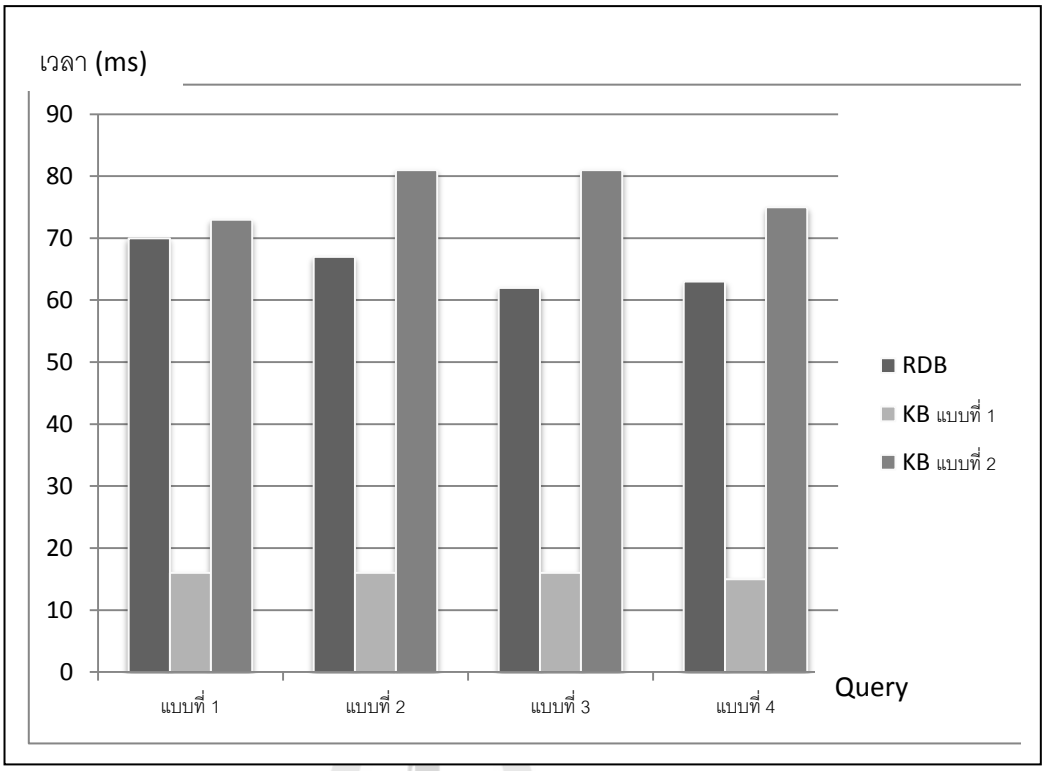
มหาวิทยาลัยเทคโนโลยีสุรนารี

เมื่อนำค่าที่ได้จากตารางมาวิเคราะห์ สามารถแสดงผลได้เป็นกราฟดังรูปที่ 4.9 และรูปที่ 4.10



รูปที่ 4.9 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Nursery ที่ตรงกับเงื่อนไข

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 4.10 กราฟแสดงเวลาที่ใช้ในการค้นหาคำตอบของการสอบถามข้อมูล Nursery ที่ไม่ตรงกับเงื่อนไข

4.3.4 การตรวจสอบความถูกต้องของผลลัพธ์ที่ได้

จากการทดสอบหาประสิทธิภาพของแหล่งความรู้ที่ได้ทั้ง 2 แบบ แล้ว จำเป็นต้องทำการทดสอบความถูกต้องของผลลัพธ์ที่ได้มาจากแหล่งความรู้ทั้ง 2 แบบ ด้วยว่าผลลัพธ์ที่ได้นั้น มีความถูกต้องมากเพียงใด โดยทำการทดสอบความถูกต้องจากการสอบถามกับฐานข้อมูลเชิงสัมพันธ์เทียบกับแหล่งความรู้แบบที่ 1 จากนั้นทำการตรวจสอบผลลัพธ์ที่ได้ ดังรูปที่ 4.11 ผลลัพธ์ที่ได้จากฐานความรู้เชิงสัมพันธ์ และรูปที่ 4.12 แสดงผลลัพธ์ที่ได้จากแหล่งความรู้แบบที่ 1

จากการทดสอบตัวอย่าง เมื่อทำการสอบถามด้วยคำถาม select * from data where buying='low' and maint='low' and safety='high' and careva = 'vgood' กับฐานข้อมูลนิรนัย และสอบถามด้วยคำถาม data (low,low,C,D,E,high,vgood). กับแหล่งความรู้แบบที่ 1 ผลลัพธ์ที่ได้พบว่าผลลัพธ์ที่ได้มีถูกต้อง สังเกตได้จากผลลัพธ์ที่ได้นั้น ได้คำตอบเท่ากัน คือ 13 ทูเพิล

```
DES-SQL> select * from data where buying='low' and maint='low' and safety='high' and careva = 'vgood'

answer(data.buying:string(varchar),data.maint:string(varchar),data.doors:string(varchar),data.persons:string
{
  answer(low,low,four,four,big,high,vgood),
  answer(low,low,four,four,med,high,vgood),
  answer(low,low,four,more,big,high,vgood),
  answer(low,low,four,more,med,high,vgood),
  answer(low,low,more,four,big,high,vgood),
  answer(low,low,more,four,med,high,vgood),
  answer(low,low,more,more,big,high,vgood),
  answer(low,low,more,more,med,high,vgood),
  answer(low,low,three,four,big,high,vgood),
  answer(low,low,three,more,big,high,vgood),
  answer(low,low,three,more,med,high,vgood),
  answer(low,low,two,four,big,high,vgood),
  answer(low,low,two,more,big,high,vgood)
}
Info: 13 tuples computed.
```

รูปที่ 4.11 แสดงผลลัพธ์ที่ได้จากการสอบถามของฐานข้อมูลเชิงสัมพันธ์

```
DES> data (low,low,C,D,E,high,vgood).

{
  data(low,low,four,four,big,high,vgood),
  data(low,low,four,four,med,high,vgood),
  data(low,low,four,more,big,high,vgood),
  data(low,low,four,more,med,high,vgood),
  data(low,low,more,four,big,high,vgood),
  data(low,low,more,four,med,high,vgood),
  data(low,low,more,more,big,high,vgood),
  data(low,low,more,more,med,high,vgood),
  data(low,low,three,four,big,high,vgood),
  data(low,low,three,more,big,high,vgood),
  data(low,low,three,more,med,high,vgood),
  data(low,low,two,four,big,high,vgood),
  data(low,low,two,more,big,high,vgood)
}
Info: 13 tuples computed.
```

รูปที่ 4.12 แสดงผลลัพธ์ที่ได้จากการสอบถามของแหล่งความรู้แบบที่ 1

4.4 อภิปรายสรุปผลการเปรียบเทียบการทดสอบแหล่งความรู้

ในการทดสอบประสิทธิภาพของแหล่งความรู้ทั้ง 2 แบบที่ได้ทำการแปลงมาจากฐานข้อมูลเชิงสัมพันธ์พบว่า แหล่งความรู้ในแบบที่ 2 จะมีประสิทธิภาพมากกว่าในการค้นหาข้อมูล สังเกตได้จากเวลาที่ใช้ในการค้นหาคำตอบนั้นจะใช้เวลาที่เร็วกว่าแบบที่ 1 และจากการทดสอบพบว่าแหล่งความรู้ทั้ง 2 แบบนี้จะแสดงถึงความแตกต่างในการค้นหาคำตอบเมื่อข้อมูลที่ใช้นั้นมีขนาดใหญ่มาก ๆ แหล่งความรู้ทั้ง 2 นี้ยังมีทั้งข้อดีและข้อเสียที่แตกต่างกันคือ แหล่งความรู้แบบที่ 1 นั้นจะมีขนาดของข้อมูลที่น้อยกว่า ทำให้เวลาที่ใช้ในการค้นหาคำตอบของคำถามที่ไม่ตรงกับเงื่อนไขของรูปแบบข้อมูลนั้นใช้เวลาน้อยกว่าแบบที่ 2 เนื่องจากว่าในภาษาดาต้าล็อกนั้นจะทำการค้นหาคำตอบไปทีละบรรทัดของแหล่งความรู้ ดังนั้นจึงสรุปได้ว่าการเพิ่มส่วนของเงื่อนไข (Rule) เข้าไปในแหล่งความรู้จะช่วยเพิ่มประสิทธิภาพให้กับแหล่งความรู้ได้ดียิ่งขึ้นในกรณีที่การสอบถามข้อมูลมีเงื่อนไขที่ตรงกับ Rule

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 5
บทสรุป

5.1 สรุปผลการวิจัย

จากการศึกษาวิจัยและออกแบบการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัยนั้น สามารถสรุปได้ว่าในการแปลงฐานข้อมูลเป็นแหล่งความรู้ในฐานข้อมูลนิรนัยนั้น เพื่อให้แหล่งความรู้ที่ได้มีประสิทธิภาพมากขึ้น แหล่งความรู้จะต้องเพิ่มในส่วนของรูปแบบข้อมูลหรือโมเดลเข้าไป ซึ่งรูปแบบข้อมูลนี้ได้มาจากการทำเหมืองข้อมูล ในงานวิจัยนี้เลือกใช้วิธีการ ID3 ซึ่งเป็นการแบ่งคลาสของข้อมูลมาหารูปแบบของข้อมูลทั้งหมด เนื่องจากวิธีการนี้เป็นวิธีการหารูปแบบของข้อมูลที่เข้าใจง่าย และผลลัพธ์ที่ได้จะได้รูปแบบข้อมูลที่อยู่ในลักษณะของ IF-THEN Rule

ในการศึกษางานวิจัยนี้ จึงมุ่งเน้นไปยังแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัยที่ได้จากการแปลงจากฐานข้อมูลเชิงสัมพันธ์ แหล่งความรู้ที่ได้จากการแปลงนี้จะแบ่งออกเป็น 2 แบบ คือ แหล่งความรู้แบบที่ 1 ที่ประกอบด้วยข้อมูลอยู่ในรูปของ Fact เพียงอย่างเดียวและแหล่งความรู้แบบที่ 2 ที่ประกอบด้วยข้อมูล (Fact) และรูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูลด้วยวิธีการ ID3 อยู่ในรูปแบบของ Rule จากนั้นจะทำการเปรียบเทียบประสิทธิภาพของแหล่งความรู้ที่ได้ โดยทำการทดสอบประสิทธิภาพจากการสอบถามข้อมูลหรือค้นหาข้อมูลจากเงื่อนไขที่ครอบคลุมกับข้อมูล คือ สอบถามข้อมูลที่ตรงตามเงื่อนไขและไม่ตรงตามเงื่อนไขของรูปแบบข้อมูล จากนั้นทำการจับเวลาที่ใช้ในการประมวลผลหาคำตอบของทั้งสองแหล่งความรู้

ผลลัพธ์ที่ได้จากการทดสอบประสิทธิภาพของแหล่งความรู้สามารถสรุปได้ว่า แหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัยที่ประกอบไปด้วยข้อมูลและรูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูลจะมีประสิทธิภาพในการสอบถามข้อมูลหรือค้นหาข้อมูลที่มีเงื่อนไขตรงกับรูปแบบของข้อมูลได้ดีขึ้นกว่าแหล่งความรู้ที่มีเพียงแต่ข้อมูลเพียงอย่างเดียว แต่พบข้อเสียคือขนาดของแหล่งความรู้จะมีขนาดใหญ่เพิ่มขึ้นจากแหล่งความรู้ที่ประกอบด้วยข้อมูล และข้อจำกัดอีกอย่าง คือ การค้นหาข้อมูลที่ไม่ตรงตามเงื่อนไขนั้นเวลาที่ใช้ในการค้นหา จะใช้เวลามากกว่าแหล่งความรู้ที่ประกอบด้วยข้อมูลเพียงอย่างเดียว

5.2 ประโยชน์ของการแปลงฐานข้อมูลเป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย

1) ช่วยลดเวลาในการสอบถามข้อมูลหรือค้นหาข้อมูลในฐานข้อมูลที่มีขนาดใหญ่มาก ๆ

2) เพิ่มประสิทธิภาพให้กับแหล่งความรู้ที่จะใช้ในฐานข้อมูลนิรนัย

5.3 ข้อจำกัดของการแปลงข้อมูลเป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย

1) ในส่วนของการทำเหมืองข้อมูลในงานวิจัยนี้ อัลกอริทึมที่ใช้ในการหารูปแบบของข้อมูลนั้น ยังใช้ได้กับข้อมูลที่เป็นแบบ Categorical ยังไม่สามารถใช้ได้กับข้อมูลที่เป็นข้อมูลจริงๆ เนื่องจากว่าอัลกอริทึมที่ใช้นั้นยังไม่สามารถค้นหาข้อมูลที่เป็นตัวเลขได้

2) ในการแปลงฐานข้อมูลเป็นแหล่งความรู้นี้ ต้องใช้ผู้ที่มีความสามารถพื้นฐานในการเขียนภาษาโปรล็อกและภาษาดาต้าล็อกในการแปลงข้อมูลและในการสอบถามข้อมูลหรือค้นหาข้อมูล อีกทั้งผู้ใช้ยังต้องมีความรู้ในการใช้งานโปรแกรม DES

5.4 แนวทางในการพัฒนาต่อ

ในการแปลงฐานข้อมูลเป็นแหล่งความรู้ในงานวิจัยนี้ แหล่งความรู้ที่ได้นั้นยังมีข้อด้อยในกรณีทั้งแหล่งความรู้แบบที่ 1 ที่มีเพียงข้อมูลเพียงอย่างเดียวและแหล่งความรู้แบบที่ 2 ที่มีทั้งข้อมูลและรูปแบบข้อมูลที่ได้จากการทำเหมืองข้อมูล ดังนั้นแนวทางพัฒนาต่อในอนาคต เพื่อปรับปรุงให้แหล่งความรู้ที่ได้นั้นมีประสิทธิภาพมากขึ้น มีข้อเสนอแนะดังนี้

1) การประยุกต์แหล่งความรู้ทั้งสองแบบให้รวมเป็นแบบเดียวกัน คือ เมื่อทำการค้นหาข้อมูลให้ทำการตรวจสอบก่อนว่ามีข้อมูลที่ค้นหาอยู่ในแหล่งความรู้แบบที่มีเฉพาะข้อมูล ถ้าพบว่ามีก็จะสามารถตอบคำถามด้วยแหล่งรู้นี้เพียงอย่างเดียว แต่ถ้าไม่มีข้อมูลที่ตรงกับคำถามจึงจะทำการค้นหาจากแหล่งความรู้ประเภทที่ 2 เพื่อค้นหา Rule

2) การลด Rule ของแหล่งความรู้ให้เหลือเพียง Rule ที่มีความสำคัญมาก ๆ เท่านั้น เพื่อลดเวลาที่ใช้ในการค้นหาและลดขนาดของแหล่งความรู้

3) หาเทคนิคอื่น ๆ ของการทำเหมืองข้อมูลที่ให้ผลลัพธ์ออกมาแล้วสามารถแปลงเป็น Rule เพิ่มเข้าไปในแหล่งความรู้เพื่อช่วยให้แหล่งความรู้มี Rule ที่หลากหลายประเภทมากขึ้น

รายการอ้างอิง

จารุณี ซามาตย์. (2551). **ข้อดีของการใช้ Relational Database**. [ออนไลน์]. ได้จากhttp://www.gotoknow.org/blog/dbdesign/209806.

ชญามน บุญประสิทธิ์. (2550). **รูปแบบฐานข้อมูล**.[ออนไลน์]. ได้จากhttp://www.thaigoodview.com/library/teachershow/lopburi/chayamon_b/com/sec01p03.html.

นิตยา เกิดประสพ. (2547). **เอกสารประกอบการสอนวิชาการค้นหาความรู้และการขุดค้นข้อมูล.** มหาวิทยาลัยเทคโนโลยีสุรนารี.

พูนศักดิ์ ประคำทอง. (2551). **การศึกษาเทคนิคของ Web Mining**. [ออนไลน์]. ได้จากhttp://wiki.nectec.or.th/ru/IT630_1_2008Students/MuekPoonsakWebminingReport.

แพรตะวัน จารุตัน.(2554). **ฐานข้อมูลเชิงสัมพันธ์ Relational Database**. [ออนไลน์]. http://teacher.snru.ac.th/peartawan/admin/news/files/Chapter%203_29-06-54.pdf

ยุทธพงษ์ ปลิ้มภิรมย์. (2549). **การทำเหมืองข้อมูล : Data Mining**. [ออนไลน์]. ได้จาก http://www.dae.mi.th/articles/DATA%20MINING.pdf.

วิทยา พรพัชรพงศ์. (2549). **Data Mining คืออะไร.** [ออนไลน์]. ได้จากhttp://www.gotoknow.org/blog/business-intelligence/43216.

ศศลักษณ์ ทองขาว. (2547). **เอกสารประกอบการเรียนการสอนวิชา Artificial Intelligence**. มหาวิทยาลัยราชภัฏสงขลา.

ศุภชัย จิวะรังสินี และ ขจรศักดิ์ สังข์เจริญ. (2549). **ระบบฐานข้อมูล Oracle Database 10g Express Edition**. กรุงเทพมหานคร : TrainLyst. 490 หน้า.

สุนิสา ริมเจริญ. (2554). **เอกสารประกอบการเรียนการสอนวิชา Artificial Intelligence**. มหาวิทยาลัยบูรพา.

อดุลย์ ยิ้มงาน. (2552). **การทำเหมืองข้อมูล (Data Mining)**. [ออนไลน์]. ได้จาก http://compcenter.bu.ac.th/index.php?option=com_content&task=view&id=75&Itemid=172.

โอฬาริก สุรินต๊ะ. (2552). **ฐานข้อมูลเชิงสัมพันธ์ Relational Database.** [ออนไลน์]. ได้จากhttp://www.wbi.msu.ac.th/file/744/doc_33.pdf.

I. J. Blanco et al. (2009). Deduction in a GEFRED database using Datalog. **Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology.** pp550-553.

L. Breiman et al. (1984). **Classification ang Regression Tree**. Belmont, CA: Wadsworth

Datalog. (2011). **Datalog Educational System (DES).** [On-line]. Available: http://www.fdi.ucm.es/profesor/fernan/des/index.html.

Datalog Tutorials. (2011). **Datalog Educational System V2.5 User's Manual**. [On-line]. Available: http://cdnetworks-kr-1.dl.sourceforge.net/project/des/des/des2.5/manualDES2.5.pdf.

J. Fong, F. Pang and C. Bloor (2001). Converting relational database into XML document. **Procudings of the 12th International Workshop on Database and Expert System Applications,** pp.61-65.

S. Goos. (2009). On the correctness of a Prolog-based federated database. **Procuding of the 10th Twente Student Conference on IT**.

J. McCarthy. (2011). **Datalog: Deductive database programming.** [On-line]. Available: http://docs.racket-lang.org/datalog/

J. M. Medina, O. Pons, J. C. Cubero and M. A. Vila. (1994). **A fuzzy relational deductive database interface**. International Journal of Intelligent Systems, pp. 597-613.

J. R. Quinlan. (1979). **Discovering rules by induction from large collections of examples**. In Expert Systems in the Micro-electronic Age, D. Michie, Ed. Edinburgh University Press, pp. 168-201.

J. R. Quinlan. (1986). **Induction of decision trees**. Machine Learning, vol. 1, no. 1. pp. 81-106.

J. R. Quinlan. (1992). **C4.5: Programs for Machine Learning**. Morgan Kaufmann.

K. Taghva and J. Jayaraman. (2008). Implementation of recursive queries for information systems. **Proceedings of the 19th International Conference On Systems Engineering,** pp. 273-275.

A. Tomasic. (1993). **Correct View Update Translations via Containment**. Standford University. Computer Science Technical Note STAN-CS-TN_93-9.

O. Zaiane. (1999). **Principles of Knowledge Discovery in Databases.** University of Alberta. [On-line]. Available: http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/

ภาคผนวก ก

รหัสต้นฉบับโปรแกรม

```
id3 :-
        write('Enter name of file to Data: '), read(F),
        write('Enter name of file Knowleage: '), read(K),
        retractall(node(_,_,_)),
        findall(N,example(N,_,_),E),
         example(_,_,L), !,
        get_attributes(L,A),
        idt(E,root,A,3),
        tell(K),
        write_fact(F),
        write_rule(A),
        told,
        write('You have one Knowleage Base for DDB'),!.


idt(E,Parent,_,Tr) :-   length(E,Len),
                        Len=<Tr,
                        distr(E, Distr),
                        assertz(node(leaf,Distr,Parent)), !.
idt(E,Parent,_,_) :-    distr(E, [C]),
                         assertz(node(leaf,[C],Parent)).
idt(Es,Parent,As,Tr) :-
                        choose_attribute(Es,As,A,Values,Rest), !,
                        partition(Values,A,Es,Parent,Rest,Tr).
idt(E,Parent,_,_) :- !,
                         node(Parent,Test,_).


get_attributes([],[]) :- !.
get_attributes([A=_|T],[A|W]) :-    get_attributes(T,W).


partition([],_,_,_,_,_) :- !.
```

```
partition([V|Vs],A,Es,Parent,Rest,Tr) :-
                        get_subset(Es,A=V,Ei), !,
                        gen_name(Node),
                        assertz(node(Node,A=V,Parent)),
                         idt(Ei,Node,Rest,Tr), !,
                        partition(Vs,A,Es,Parent,Rest,Tr).


choose_attribute(Es,As,A,Values,Rest) :-
                        length(Es,LenEs),
                        information_content(Es,LenEs,I), !,
                        findall((A-Values)/Gain, (member(A,As),
                        get_values(Es,A,[],Values),
                        split_into_subsets(Values,Es,A,Ess),
                         residual_information(Ess,LenEs,R),Gain is I - R),All),
                        maximum(All,(A-Values)/_),
                        efface(A,As,Rest), !.


split_into_subsets([],_,_,[]) :- !.
split_into_subsets([V|Vs],Es,A,[Ei|Rest]) :-
                        get_subset(Es,A=V,Ei), !,
                        split_into_subsets(Vs,Es,A,Rest).


residual_information([],_,0) :- !.
residual_information([Ei|Es],Len,Res) :-
                        length(Ei,LenEi),
                        information_content(Ei,LenEi,I), !,
                        residual_information(Es,Len,R),
                         Res is R + I*LenEi/Len.


information_content(Es,Len,I) :-
```

```
                    setof(C,E^L^(member(E,Es),example(E,C,L)),Classes), !,
                    sum_terms(Classes,Es,Len,I).


sum_terms([],_,_,0) :- !.
sum_terms([C|Cs],Es,Len,Info) :-
                    findall(E,(member(E,Es),example(E,C,_)),InC),
                    length(InC,N),
                    sum_terms(Cs,Es,Len,I),
                    Info is I - (N/Len)*(log(N/Len)/log(2)).


get_values([],_,Values,Values) :- !.
get_values([E|Es],A,Vs,Values) :-
                    example(E,_,L),
                    member(A=V,L), !,
                    (member(V,Vs), !, get_values(Es,A,Vs,Values);
                    get_values(Es,A,[V|Vs],Values)  ).


get_subset([],_,[]) :- !.
get_subset([E|Es],A,[E|W]) :-
                    example(E,_,L),
                    member(A,L), !,
                    get_subset(Es,A,W).


get_subset([_|Es],A,W) :- get_subset(Es,A,W).


path(Parent,[],Class) :-  node(leaf,Class,Parent), !.
path(Parent,[A|Path],Leaf) :-
                     node(Son,A,Parent),
                    path(Son,Path,Leaf).
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
distr(S,Dist) :-
                        setof(C,X^L^(member(X,S),example(X,C,L)),Cs),
                         countc(Cs,S,Dist).


countc([],_,[]) :- !.
countc([C|L],E,[C/N|T]) :-
                        findall(X,(member(X,E),example(X,C,_)),W),
                        length(W,N), !,
                        countc(L,E,T).


gen_name(M) :-    retract(nam(N)),
                        M is N+1,
                        assert(nam(M)), !.
gen_name(1) :-    assert(nam(1)).


efface(X,[X|T],T) :- !.
efface(X,[Y|T],[Y|Z]) :-  efface(X,T,Z).


subset([],_) :- !.
subset([X|T],L) :-  member(X,L), !,
                        subset(T,L).


maximum([X],X) :- !.
maximum([X/M|T],Y/N) :-
                        maximum(T,Z/K),
                        (M>K,Y/N=X/M ; Y/N=Z/K), !.


/*------------------- write_rules -------------------------*/
write_rule(A) :-
                        path(root,Path,Conclusion),
```

```
                        cutt(Conclusion,B),
                        length(A,D),
                        check_attibute(A,'C',1,D,H),
                        cut(H,At),
                        write('rule('),attibute(At),write(','),write(B),write(')'),write(':-'),
                        write('data('),attibute(At),write('),'),
                        rule(Path,H),writeln('.'),
                        fail.
write_rule(_).
/*-------------------End  write_rules -------------------------*/


check_attibute([A|B],S,E,D,[A=W|T]):- atomic_list_concat([S,E],W),
                        T1 is E+1,check_attibute(B,S,T1,D,T).
check_attibute([A],S,D,D,[A=W]):-  atomic_list_concat([S,D],W),!.

rule([A=B|T],At) :-         check(A,At,R),
                        write(R),write('='),write(B),rule1(T,At).
rule1([A=B|T],At) :-        check(A,At,R),
                        write(','),write(R),write('='),write(B),rule1(T,At).
rule1([],_) :-       !.


check(L,[_=_|T],R):- check(L,T,R).
check(L,[L=B|_],B):-!.


cut([],[]):-!.
cut([_=A|T],[A|W]):- cut(T,W).
cutt([A/_],A):- !.


attibute([H|T]) :- T\=[],write(H),write(','),attibute(T).
attibute([H]) :- write(H),!.
```

```
/*-------------------write_fact --------------------------*/
write_fact(A) :-  see(A),
                        repeat,
                        read(Data),
                        checkd(Data),
                        seen,!.
write_fact(_).
checkd(A) :- A\=end_of_file,write(A),writeln('.'),fail.
checkd(end_of_file) :- !.



/*-------------------End  write_fact ---------------------*/
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ข

บทความทางวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

รายชื่อบทความที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

ประภัสสร สิเนหะวัฒนะ (2012). **การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย**. การประชุมวิชาการ “ศรีนครินทรวิโรฒวิชาการ” ครั้งที่ 6. 29-30 พฤษภาคม, 2012.,หน้า 86.

SWU6-1026: การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย
TRANSFORMATION OF RELATIONAL DATABASE TO A KNOWLEDGE REPOSITORY IN DEDUCTIVE DATABASE

ประภัสสร สิเนหะวัฒนะ[*], นิตยา เกิดประสพ, กิตติศักดิ์ เกิดประสพ
Prapatsorn Sinahawattana[*], Nittaya Kerdprasop, Kittisak Kerdprasop

สาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี
School of Computer Engineering, Faculty of Engineering, Suranaree University of Technology, Thailand.
[*]**Corresponding author, E-mail:** prapatsorn.sin@gmail.com

บทคัดย่อ

งานวิจัยนี้เป็นการนำเสนอเกี่ยวกับการแปลงฐานข้อมูลเชิงสัมพันธ์เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ฐานความรู้ที่ได้จะประกอบด้วย 2 ส่วนคือ ส่วนที่เป็น Intentional Database (IDB) และส่วน Extensional Database (EDB) ในฐานข้อมูลนิรนัย ผู้วิจัยจะแสดงการแปลงข้อมูลโดยใช้ภาษาดาต้าล็อก และมีการนำกฎหรือรูปแบบข้อมูลที่ได้จากการทำเหมืองข้อมูลชนิด Decision-tree induction เข้ามาเพิ่มเข้าไปในส่วนของ IDB จากนั้นจะทำการเปรียบเทียบประสิทธิภาพของฐานความรู้ที่ได้พบว่า ฐานความรู้มีประสิทธิภาพมากขึ้นในการค้นหาข้อมูลหรือค้นหาคำตอบจากการสอบถามข้อมูล

คำสำคัญ: ฐานข้อมูลนิรนัย โปรล็อก ดาต้าล็อก ต้นไม้ตัดสินใจ (ID3)

Abstract

This paper aims to transform Relational Database to Knowledge repository in Deductive Database. Knowledge Base consists of an intentional (IDB) and extensional databases (EDB). The proposed procedure consists of 2 components: the transformed database to be used an EDB, and the transformed rules from ID3 result to be used an IDB. Then, we compare the performance of the knowledge Base alone to the Knowledge Base with induced rules stored as IDB, the experiment results show that Knowledge Base with induced rules are more effective to search information or answer to queries.

Keywords: Deductive Database, Prolog, Datalog, DecisionTree (ID3)

บทนำ

ปัจจุบันในองค์กรต่างๆ ไม่ว่าจะเป็นบริษัทขนาดเล็กหรือขนาดใหญ่ต่างก็มีการเก็บรวบรวมข้อมูลไว้จำนวนมาก ในอดีตนั้นการจัดเก็บข้อมูลจะทำการจัดเก็บข้อมูลต่างๆ ในรูปแบบของแฟ้มเอกสาร ซึ่งการจัดเก็บแบบแฟ้มนั้นมีข้อเสียมาก ไม่ว่าจะเป็นความซับซ้อนของข้อมูล จำนวนข้อมูลที่มีขนาดมาก ยากต่อการแก้ไขและค้นหา ต่อมาเมื่อมีการนำคอมพิวเตอร์เข้ามาใช้งานมากขึ้น การจัดเก็บข้อมูลจึงถูกนำมาเก็บไว้ในคอมพิวเตอร์ที่เรียกว่า ระบบฐานข้อมูล ซึ่งทำให้การจัดเก็บข้อมูลมีระบบมากขึ้น ฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

เป็นฐานข้อมูลที่ได้รับความนิยมใช้กันอย่างมาก [1] เป็นการจัดเก็บข้อมูลในรูปแบบของตาราง 2 มิติ ประกอบไปด้วย แถวและคอลัมน์

การทำเหมืองข้อมูล [2-3] เป็นวิธีการหารูปแบบของข้อมูลที่ซ่อนอยู่ในข้อมูลที่มีจำนวนมาก โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่องและหลักทางคณิตศาสตร์ ประโยชน์ของการทำเหมืองข้อมูลนั้นจะได้รูปแบบที่ช่วยในการตัดสินใจ การวางแผนและการดำเนินงานบางอย่างได้ การทำเหมืองข้อมูลจำเป็นต้องอาศัยบุคคลและความรู้จำนวนมากถึงจะได้รับประโยชน์สูงสุด

จากที่กล่าวมาจะเห็นได้ว่าปัจจุบันการจัดเก็บข้อมูลมีมากขึ้น ส่งผลให้ข้อมูลที่ถูกเก็บไว้เพิ่มขึ้นเป็นจำนวนมาก ซึ่งข้อมูลที่มีจำนวนมากๆ นี้ บางข้อมูลก็ไม่ได้ถูกนำกลับมาใช้ประโยชน์อีก และข้อมูลที่มีจำนวนมากๆ นี้ก็ส่งผลให้ใช้ระยะเวลาในการค้นหาข้อมูล หรือการนำข้อมูลมาวิเคราะห์ เกิดความซับซ้อน ยุ่งยาก และใช้เวลานานมากขึ้น ผู้วิจัยจึงได้เสนอแนวคิดในการที่จะแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานความรู้ที่จะนำไปใช้ในฐานข้อมูลนิรนัย [4] ฐานข้อมูลนิรนัยจะเป็นออกเป็นส่วน Extensional Database (EDB) จะประกอบไปด้วย fact ของข้อมูลและ Intentional Database (IDB) จะประกอบด้วย rule ที่ได้จากการนำ fact มาสร้างเงื่อนไข

ฐานข้อมูลนิรนัยในงานวิจัยนี้เป็นฐานข้อมูลที่เพิ่มในส่วนของการทำเหมืองข้อมูลเข้ามาช่วย ทำให้ฐานข้อมูลนิรนัยที่ได้ต่างจากฐานข้อมูลทั่วไป การทำเหมืองข้อมูลจะทำให้เราได้รูปแบบของข้อมูลขึ้นมา รูปแบบของข้อมูลนี้จะช่วยทำให้การค้นหาข้อมูลและการวิเคราะห์ข้อมูลเป็นไปได้ง่ายขึ้น พร้อมทั้งยังช่วยลดเวลาในการค้นหาอีกด้วย

วัตถุประสงค์ของการวิจัย

งานวิจัยนี้จัดทำขึ้น โดยมีวัตถุประสงค์เพื่อที่จะทำการแปลงข้อมูลจากฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานความรู้ในฐานข้อมูลนิรนัยที่จะประกอบไปด้วยส่วนที่เป็น EDB และ IDB ในส่วนของ IDB จะได้มาจากการทำเหมืองข้อมูล โดยใช้อัลกอริทึม ID3 ฐานความรู้ที่ได้จากงานวิจัยนี้จะเพิ่มประสิทธิภาพในการค้นหาข้อมูลหรือการสอบถามข้อมูลที่มีขนาดใหญ่ๆ ได้ดีขึ้น ใช้เวลาในการค้นหาคำตอบได้เร็วขึ้นกว่าการค้นหาข้อมูลจากฐานข้อมูลทั่วไป

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการแปลงฐานข้อมูลไปเป็นฐานความรู้ที่ใช้ในฐานข้อมูลนิรนัย ฐานความรู้ที่ได้จะอยู่ในรูปแบบของ fact และ rule แนวคิดในการแปลงข้อมูลเป็นฐานความรู้ภาษาที่ใช้ในการเขียนอัลกอริทึมคือ ภาษาดาต้าล็อกและภาษาโปรล็อก งานวิจัยจะแบ่งการทำงานออกเป็น 2 ส่วน คือ ส่วนของการแปลงข้อมูลเป็นฐานความรู้ และส่วนของการทำเหมืองข้อมูล

1. ส่วนของการแปลงข้อมูลเป็นฐานความรู้ ส่วนนี้เป็นการจัดการกับข้อมูลที่มีอยู่ โดยใช้โปรแกรม DES [5] เข้ามาช่วยในการแปลงข้อมูลให้อยู่ในรูปแบบของ fact

ตัวอย่างในการแปลงข้อมูลจากตารางในฐานข้อมูลเชิงสัมพันธ์ไปเป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย จากรูปที่ 1 จะเป็นข้อมูลตัวอย่างจากตารางในฐานข้อมูลเชิงสัมพันธ์ จากนั้นเราจะทำการแปลงข้อมูลในตารางให้อยู่ในรูปแบบของฐานความรู้ โดยการใช้ภาษาดาต้าล็อกในการสร้างตารางข้อมูลขึ้นมาจากตารางข้อมูลที่มีอยู่ดังรูปที่ 2 และจากรูปที่ 3 จะเป็นรูปแบบของข้อมูลที่ได้ทำการแปลงแล้วจะอยู่ในรูปแบบของ fact

งานวิจัยที่ทำการศึกษานี้จะทำการแปลงข้อมูลในตารางทุกตารางของข้อมูลที่นำมาแปลงเป็นฐานความรู้ จากนั้นจะทำการเก็บบันทึกข้อมูลใหม่ไว้ที่ไฟล์ .dl เพื่อจัดเก็บเป็นฐานความรู้ในฐานข้อมูลนิรนัย

No	Parents	Hasnurs	Form	Children	Housing	Finance	Social	Health	Class
1	usual	proper	complete	one	convenient	convenient	nonprob	recommended	recommend
2	usual	proper	complete	one	convenient	convenient	nonprob	priority	priority
3	usual	proper	complete	one	convenient	convenient	nonprob	not_recom	not_recom
4	usual	proper	complete	one	convenient	convenient	slightly_prob	recommended	recommend
5	usual	proper	complete	one	convenient	convenient	slightly_prob	priority	priority
6	usual	proper	complete	one	convenient	convenient	slightly_prob	not_recom	not_recom
7	usual	proper	complete	one	convenient	convenient	problematic	recommended	priority
8	usual	proper	complete	one	convenient	convenient	problematic	priority	priority
9	usual	proper	complete	one	convenient	convenient	problematic	not_recom	not_recom
10	usual	proper	complete	one	convenient	inconv	nonprob	recommended	very_recom
11	usual	proper	complete	one	convenient	inconv	nonprob	priority	priority
12	usual	proper	complete	one	convenient	inconv	nonprob	not_recom	not_recom
13	usual	proper	complete	one	convenient	inconv	slightly_prob	recommended	very_recom

ภาพที่ 1 ตัวอย่างตารางข้อมูล

```
/sql
create table data(parents string,hasnurs string,form string,children string,housing
                  string,finance string,social string,health string);

insert into data values('usual','proper','complete','one','convenient','convenient',
                        'nonprob','recommended');
insert into data values('usual','proper','complete','one','convenient','convenient',
                        'nonprob','priority');
insert into data values('usual','proper','complete','one','convenient','convenient',
                        'nonprob','notrecom');
                        ..
                        ..
                        ..
                        ..
                        ..
insert into data values('greatpret','verycrit','foster','more','critical','inconv',
                        'problematic','priority');
insert into data values('greatpret','verycrit','foster','more','critical','inconv',
                        'problematic','notrecom');
select * from data;
```

ภาพที่ 2 ตัวอย่างของภาษาดาต้าล็อกที่ใช้ในการสร้างฐานความรู้

```
data(greatpret,critical,complete,more,convenient,convenient,nonprob,
        notrecom,notrecom).
data(greatpret,critical,complete,more,convenient,convenient,nonprob,
        priority,specprior).
data(greatpret,critical,complete,more,convenient,convenient,nonprob,
        recommended,priority).
                ..
                ..
                ..
                ..
                ..
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,
        notrecom,notrecom).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,
        priority,specprior).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,
        recommended,specprior).
:-type(data(parents:string,hasnurs:string,form:string,children:string,
housing:string,finance:string,social:string,health:string,nursary:string)).
```

ภาพที่ 3 ตัวอย่างของข้อมูลในฐานความรู้ที่ได้

2. ส่วนของการทำเหมืองข้อมูล ส่วนนี้จะเป็นส่วนการค้นหาความสัมพันธ์ของข้อมูล เพื่อที่จะนำความสัมพันธ์ที่ได้มาสร้างเป็นกฎในฐานความรู้เพื่อใช้ในฐานข้อมูลเชิงอุปนัย อัลกอริทึมที่ใช้คือ อัลกอริทึม ID3 เป็นอัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ (Decision Tree) จะใช้ค่าเกนของสารสนเทศในการตัดสินใจที่จะเลือกคุณสมบัติแอททริบิวต์ที่จะใช้เป็นรากหรือโหนดในต้นไม้ โดยดูจากค่าเกนสูงที่สุด

ในงานวิจัยนี้จะใช้ภาษาโปรล็อกในการเขียนอัลกอริทึม ID3 ขึ้นมาเพื่อใช้กับข้อมูลในฐานข้อมูลที่จะทำการแปลงเป็นฐานความรู้ ผลลัพธ์ที่ได้จากอัลกอริทึมนี้จะออกมาได้รูปแบบของกฎ (IF-THEN) แล้วจากนั้นจะทำการแปลงกฎที่ได้ให้อยู่ในรูปแบบของ rule ในฐานความรู้ แล้วทำการเพิ่มเข้ารวมเข้าไปกับข้อมูลที่ได้จากส่วนของการแปลงข้อมูลเป็นฐานความรู้

จากตัวอย่างข้อมูลในตาราง (ภาพที่ 1) เมื่อนำข้อมูลมาทำเหมืองข้อมูลจะได้รูปแบบของข้อมูล ดังภาพที่ 4 ข้อมูลที่ได้นั้น เราจะนำมาทำการแปลงให้อยู่ในรูปแบบของภาษาดาต้าล็อก ซึ่งรูปแบบของการแปลงจะอยู่ในรูปแบบดังภาพที่ 5 จากรูปจะเห็นได้ว่าลักษณะของ Rule ที่จะมีลักษณะคล้ายกับรูปแบบข้อมูล เพียงแต่ส่วนของIF และ THEN จะสลับตำแหน่งกัน

```
if[health=notrecom]then[notrecom].
if[health=priority, hasnurs=verycrit, children=more]then[specprior].
if[health=priority, hasnurs=verycrit, children=two]then[specprior].
                        ..
                        ..
                        ..
                        ..
                        ..

if[health=recommended, hasnurs=proper, parents=usual, social=nonprob,
       housing=convenient, finance=convenient, children=more]then[veryrecom].
if[health=recommended, hasnurs=proper, parents=usual, social=nonprob,
       housing=convenient, finance=convenient, children=one, form=completed]
       then[veryrecom].
if[health=recommended, hasnurs=proper, parents=usual, social=nonprob,
       housing=convenient, finance=convenient, children=one, form=complete]
       then[recommend].
.
```

ภาพที่ 4 รูปแบบของข้อมูลที่ได้จากการทำเหมืองข้อมูล

```
rule(C1,C2,C3,C4,C5,C6,C7,C8,notrecom):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=notrecom.
rule(C1,C2,C3,C4,C5,C6,C7,C8,specprior):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=priority,
                                          C2=critical,C1=usual,C5=critical,C4=two.
rule(C1,C2,C3,C4,C5,C6,C7,C8,specprior):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=priority,
                                          C2=critical,C1=usual,C5=critical,C4=one,C3=foster.
                        ..
                        ..
                        ..
                        ..
                        ..

rule(C1,C2,C3,C4,C5,C6,C7,C8,priority):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=priority,
                                          C2=improper,C1=usual.
rule(C1,C2,C3,C4,C5,C6,C7,C8,veryrecom):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=proper,C1=usual,C7=nonprob,C5=convenient,
                                          C6=convenient,C4=three.
rule(C1,C2,C3,C4,C5,C6,C7,C8,priority):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=critical,C1=usual,C7=nonprob.
rule(C1,C2,C3,C4,C5,C6,C7,C8,recommend):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=proper,C1=usual,C7=nonprob,C5=convenient,
                                          C6=convenient,C4=one,C3=complete.
```

ภาพที่ 5 รูปแบบของข้อมูลในภาษาดาต้าล็อก

การทดสอบและประเมินผลของงานวิจัยนี้ จะมุ่งเน้นไปที่การเปรียบเทียบความถูกต้องและเวลาที่ใช้ในการสอบถามข้อมูลหรือค้นหาข้อมูลในฐานข้อมูลที่เป็นฐานความรู้เพื่อใช้ในฐานข้อมูลนิรนัย ซึ่งฐานความรู้ที่ทำการเปรียบเทียบจะมีสองรูปแบบ คือ ฐานความรู้ที่เป็น Fact ข้อมูลเพียงอย่างเดียว กับฐานความรู้ที่มี Rule จากการทำเหมืองข้อมูลมาเพิ่มเข้าไป จากนั้นจะทำการเปรียบเทียบประสิทธิภาพฐานความรู้ทั้งสองแบบ

ผลการวิจัย

งานวิจัยนี้จะทำการทดลองเปรียบเทียบประสิทธิภาพของการค้นหาข้อมูลหรือสอบถามข้อมูลจากฐานความรู้ที่ใช้ในฐานข้อมูลนิรนัยแบบธรรมดา และแบบที่ทำการเพิ่ม rule จากที่ได้จากอัลกอริทึม ID3 เพิ่มเข้ามา

ข้อมูลที่ใช้ทดลองในงานวิจัยนี้ใช้ข้อมูล 9 แอททริบิวต์ 12690 เรคคอร์ด มาทำการแปลงข้อมูลให้อยู่ในรูปแบบของฐานความรู้ ฐานความรู้ที่ได้จะประกอบไปด้วยข้อมูลและรูปแบบของข้อมูลที่ได้จากอัลกอริทึม ID3 จากนั้นจะทำการเปรียบเทียบหาประสิทธิภาพของฐานความรู้ที่ได้กับฐานความรู้ธรรมดาที่ไม่มีรูปแบบของข้อมูล

จากภาพที่ 6 จะเป็นฐานความรู้ที่ได้จากงานวิจัยนี้ จะประกอบไปด้วยข้อมูล (fact) และรูปแบบของข้อมูลที่ได้มาจาก ID3 และทำการแปลงให้อยู่ในรูปแบบของ rule

จากการทดลองของงานวิจัยจะทำการเปรียบเทียบการสอบถามข้อมูลกับฐานความรู้ธรรมดาที่มีเพียงแต่ข้อมูลเพียงอย่างเดียว ดังภาพที่ 7 ในส่วนของการสอบถามข้อมูลจะทำการทดสอบโดยการสอบถามข้อมูลออกเป็น 2 แบบ คือ สอบถามข้อมูลที่ตรงกับรูปแบบของข้อมูลและสอบถามข้อมูลที่ไม่ตรงกับรูปแบบของข้อมูล จากนั้นจะทำการจับเวลาในการค้นหาข้อมูลแต่ละคำถาม ผลที่ได้จากการสอบถามข้อมูลจะเป็นไปตามตารางที่ 1 และภาพที่ 8-9 จะเป็นการแสดงกราฟเปรียบเทียบเวลาของการสอบถามข้อมูลทั้งสองแบบพบว่า เวลาในการสอบถามข้อมูลของฐานความรู้ที่มี rule จากอัลกอริทึม ID3 เพิ่มเข้าไปจะช่วยให้เวลาที่ใช้ในการหาคำตอบนั้นน้อยกว่ากว่าฐานความรู้ธรรมดา

```
data(greatpret,critical,complete,more,convenient,convenient,nonprob,notrecom).
data(greatpret,critical,complete,more,convenient,convenient,nonprob,priority).
                          ..
                          ..
                          ..
                          ..

data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,priority).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,recommended).
:-type(data(parents:string,hasnurs:string,form:string,children:string,housing:string,
            finance:string,social:string,health:string)).
rule(C1,C2,C3,C4,C5,C6,C7,C8,notrecom):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=notrecom.
rule(C1,C2,C3,C4,C5,C6,C7,C8,specprior):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=priority,
                                          C2=verycrit,C4=more.
rule(C1,C2,C3,C4,C5,C6,C7,C8,priority):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=proper,C1=usual,C7=nonprob,C5=convenient,
                                          C6=inconv,C4=more.
rule(C1,C2,C3,C4,C5,C6,C7,C8,veryrecom):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=proper,C1=usual,C7=nonprob,C5=convenient,
                                          C6=convenient,C4=three.
rule(C1,C2,C3,C4,C5,C6,C7,C8,recommend):-data(C1,C2,C3,C4,C5,C6,C7,C8),C8=recommended,
                                          C2=proper,C1=usual,C7=nonprob,C5=convenient,
                                          C6=convenient,C4=one,C3=complete.
```

ภาพที่ 6 ตัวอย่างฐานความรู้ของงานวิจัย

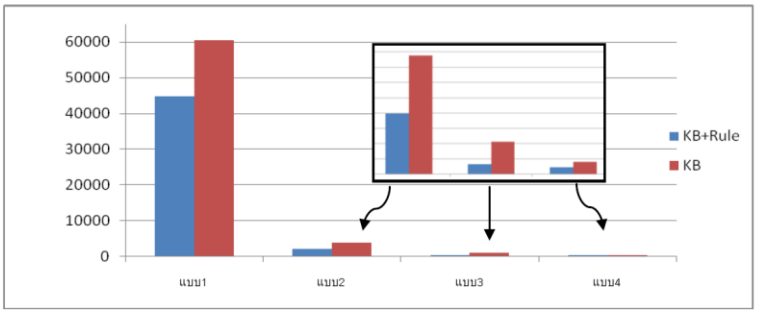
```
data(greatpret,critical,complete,more,convenient,convenient,nonprob,notrecom,notrecom).
data(greatpret,critical,complete,more,convenient,convenient,nonprob,priority,specprior).
                          ..
                          ..
                          ..
                          ..

data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,notrecom,notrecom).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,priority,specprior).
data(usual,verycrit,incomplete,two,lessconv,inconv,slightlyprob,recommended,specprior).
:-type(data(parents:string,hasnurs:string,form:string,children:string,housing:string,
finance:string,social:string,health:string,nursary:string)).
```

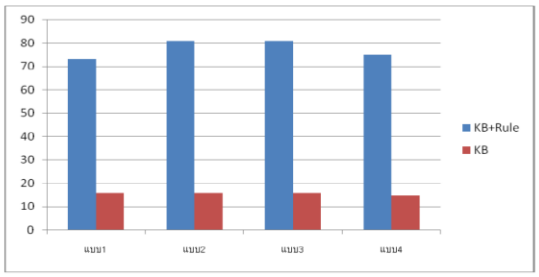
ภาพที่ 7 ฐานความรู้ที่นำมาเปรียบเทียบ

ตารางที่ 1 ตารางแสดงเวลาที่ใช้ในการสอบถามข้อมูล

Query	KB+rule	KB
	เวลา (ms)	
ตรงกับรูปแบบข้อมูล (rule)		
(C1,C2,C3,C4,C5,C6,C7,C8,priority)	44685	60418
(greatpret,C2,C3,more,C5,inconv,C7,priority,R)	1987	3877
(C1,improper,C3,two,critical,inconv,C7,priority,specprior)	327	1057
(pretentious,improper,foster,C4,convenient,C6,problematic,notrecom,notrecom)	225	403
ไม่ตรงกับรูปแบบข้อมูล(rule)		
(C1,C2,C3,mo,C5,C6,C7,C8,Re)	73	16
(C1,C2,C3,C4,C5,test,C7,notrecom,R)	81	16
(usual,improper,complete,one,C5,incomplete,C7,notrecom,specprior)	81	16
(pretentious,improper,completed,three,convenient,inconv,nonprob, notrecom,veryrecom)	75	15



ภาพที่ 8 การเปรียบเทียบเวลาที่ใช้ในการสอบถามข้อมูลแบบตรงกับรูปแบบข้อมูล (Rule)



ภาพที่ 9 การเปรียบเทียบเวลาที่ใช้ในการสอบถามข้อมูลแบบไม่ตรงกับรูปแบบข้อมูล (Rule)

สรุปและอภิปรายผล

จากผลทดลองดังกล่าว จะเห็นได้ว่าเมื่อมีการนำ rule ที่ได้จากอัลกอริทึม ID3 เข้ามาในฐานความรู้แล้วจะช่วยให้เวลาที่ใช้ในการตรวจสอบหาคำตอบให้กับการสอบถามข้อมูลลดลงไปจากเวลาที่ใช้ตรวจสอบหาคำตอบจากฐานความรู้ธรรมดาที่มีเพียงข้อมูลอย่างเดียว เนื่องจากว่าฐานความรู้ที่มี rule อยู่ด้วยนั้น เมื่อทำการป้อนการสอบถามเข้าไปแล้ว จะเข้าไปตรวจเช็คที่ rule ที่มีก่อนจึงทำให้เวลาที่ได้เร็วกว่าแบบธรรมดาที่ต้องเข้าไปตรวจสอบทุกข้อมูลที่มีอยู่จนกว่าจะพบคำตอบที่ต้องการ แต่เมื่อทำการป้อนการสอบถามข้อมูลที่ไม่ตรงกับ rule แล้วจะพบว่าเวลาที่ได้จะมากกว่าแหล่งข้อมูลธรรมดาที่ไม่มี rule ดังนั้นแหล่งข้อมูลที่ได้จากการเพิ่ม rule ของอัลกอริทึม ID3 เข้าไปจะช่วยเพิ่มประสิทธิภาพให้กับการค้นหาคำตอบของการสอบถามข้อมูลที่มีอยู่ในแหล่งข้อมูลนั้น และยังพบข้อด้อยเมื่อทำการสอบถามข้อมูลที่ไม่ตรงกับรูปแบบที่มีอยู่จะใช้เวลาได้นานกว่าฐานความรู้ที่ไม่มี rule

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนจากมหาวิทยาลัยเทคโนโลยีสุรนารี

เอกสารอ้างอิง

[1] รวิวรรณ เทนอิสสระ. (2543). *ฐานข้อมูลและการออกแบบ*. กรุงเทพฯ: เอิร์ดเวฟ เอ็ดดูเคชั่น.

[2] นิตยา เกิดประสพ. (2547). *เอกสารประกอบการสอนวิชาการค้นหาความรู้และการขุดค้นข้อมูล*. นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุรนารี.

[3] Tom M, Mitchell. (1997). *Machine Learning*. Singapore: McGraw-Hill.

[4] Miguel Nussbaum. (1991). *Building a deductive database*. New Jersey: Ablex.

[5] Fernando Sáenz-Pérez. (2011). *Datalog Educational System V2.5User's Manual*. Retrieved September 22, 2011, from http://cdnetworks-kr-1.dl.sourceforge.net/project/des/des/des2.5/manualDES2.5.pdf

มหาวิทยาลัยเทคโนโลยีสุรนารี

ประวัติผู้เขียน

นางสาวประภัสสร สิเนหะวัฒนะ เกิดเมื่อวันที่ 21 พฤศจิกายน พ.ศ. 2530 ที่ อำเภอเมือง จังหวัดมุกดาหาร เริ่มเข้าศึกษาในระดับประถมศึกษาที่โรงเรียนเซนต์ยอแซฟ มุกดาหาร ชั้นมัธยมศึกษาที่โรงเรียนมุกดาหาร จากนั้นเข้าศึกษาต่อระดับปริญญาตรีในปี 2549 ที่สาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี และสำเร็จการศึกษาเมื่อปี 2552 ภายหลังสำเร็จการศึกษาในระดับปริญญาตรี ได้เข้าศึกษาต่อในระดับปริญญาโท สาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี ในปีการศึกษา 2553

ในระหว่างการศึกษาได้รับความอนุเคราะห์อย่างดียิ่งจากคณาจารย์ในสาขาวิชา โดยได้รับความไว้วางใจให้เป็นผู้ช่วยสอนปฏิบัติการวิชา Computer Programming

ผลงานวิจัย : ได้เสนอบทความเข้าร่วมในงานประชุมวิชาการ “ศรีนครินทรวิโรฒวิชาการ” ครั้งที่ 6 ประจำปี 2555 เรื่อง การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นแหล่งความรู้ที่ใช้ในฐานข้อมูลนิรนัย

มหาวิทยาลัยเทคโนโลยีสุรนารี