สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลในระบบสารสนเทศ

AN EXPERT SYSTEM FOR ASSESSING DATA RELIABILITY IN AN INFORMATION SYSTEM

นายพีระศักดิ์ เสรีกุล

MR. PEERASAK SERIKUL

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2539

ISBN 974-621-513-2

เลขหมู่.....
เลขทะเบียน 27131
วัน, เดือน, ปี 11 ก.พ. 2540

ลิขสิทธิ์ของบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

AN EXPERT SYSTEM FOR ASSESSING DATA RELIABILITY IN AN INFORMATION SYSTEM

PEERASAK SERIKUL

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
MASTER OF SCIENCE PROGRAM IN COMPUTER SCIENCE
AND INFORMATION TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
1996

ISBN 974-621-513-2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลในระบบสารสนเทศ
นักศึกษา	นายพีระศักดิ์ เสรีกุล
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ดร.เอื้อน ปิ่นเงิน
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.	2539

บทคัดย่อ

สารสนเทศที่มีคุณค่าจะต้องได้มาจากข้อมูลที่เชื่อถือได้ อย่างไรก็ตามวิธีการประเมินค่าความน่าเชื่อถือของข้อมูลก็ยังมิได้รับการยอมรับกันอย่างกว้างขวางในระบบสารสนเทศ วิทยานิพนธ์ฉบับนี้จะกล่าวถึง การนำเอาการประเมินความน่าเชื่อถือที่ใช้อยู่ในระบบการควบคุมคุณภาพ(Quality Control) มาประยุกต์ใช้กับระบบสารสนเทศ โดยจะกล่าวถึงวิธีการในการประเมินความน่าเชื่อถือของข้อมูล 3 ลักษณะ คือ ความน่าเชื่อถือของข้อมูลที่เป็นไปตามลักษณะพื้นฐานที่ยอมรับกันโดยทั่วไปของรายการข้อมูลแต่ละรายการ(Internal reliability), ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความสัมพันธ์และข้อกำหนดของระบบฐานข้อมูล(Relative reliability), ความน่าเชื่อถือของข้อมูลเมื่อเปรียบเทียบกับข้อมูลตามความเป็นจริง(Absolute reliability)

วิทยานิพนธ์ฉบับนี้ยังได้พัฒนาระบบผู้เชี่ยวชาญ(Expert system) เพื่อประยุกต์ใช้ในการประเมินความน่าเชื่อถือ โดยจะอ้างอิงตามลักษณะการประเมินความน่าเชื่อถือ ทั้ง 3 ลักษณะที่กล่าวมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

Thesis Title	An Expert System for Assessing Data Reliability in an Information System
Student	Mr.Peerasak Serikul
Thesis Adivisor	Dr.Ouen Pin-ngern
Level of Study	Master of Science Program in Computer Science and Information Technology
Department	Mathematics and Computer Science , Faculty of Science King Mongkut's Institute of Technology Ladkrabang
Year	1996

ABSTRACT

Information is valuable if it derives from reliable data. However, measurements for data reliability have not been widely established in the area of information systems. This thesis attempts to draw some concepts of reliability from the field of quality control and to apply them to information systems. This thesis presents three measurements for data reliability. *Internal reliability* : Reliability whose assessment is based on commonly accepted criterias about the characteristics of the data items. *Relative reliability* : Reliability of the data in view of the user requirements. *Absolute reliability* : Reliability which is measured by performing direct comparisons between the content of the database and reality.

Also implementing in this thesis is applying an expert system in evaluating the reliability referring to the three measurements.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบคุณ ดร.เอื้อน ปิ่นเงิน ที่ให้ความกรุณารับเป็นอาจารย์ที่ปรึกษา และให้คำแนะนำที่เป็นประโยชน์ ตลอดจนช่วยเหลือในตรวจสอบและแก้ไขเนื้อหาเป็นอย่างดี

ขอขอบคุณท่านอาจารย์ทุกท่านที่ประสิทธิ์ประสาทความรู้ ตลอดระยะเวลาที่ผ่านมา รวมถึงเพื่อนร่วมรุ่นของข้าพเจ้าทุกท่านที่คอยให้กำลังใจด้วยดีตลอดมา

ขอบคุณพี่ และญาติทุกท่านรวมถึงคุณตา คุณยายที่ค่อยเป็นห่วง และให้กำลังใจอยู่เสมอ

ท้ายที่สุดนี้ขอกราบขอบพระคุณ บิดา มารดา ที่มอบความอบอุ่น และกำลังใจ ตลอดจนให้การสนับสนุนทางด้านทุนทรัพย์ด้วยดีตลอดมา

พีระศักดิ์ เสรีกุล

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญภาพ VIII

บทที่

1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 1

1.3 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย 2

1.4 ขอบเขตของการวิจัย 3

1.5 วิธีที่ใช้ในการดำเนินการวิจัย 3

2 ระบบผู้เชี่ยวชาญ(Expert System) 4

2.1 กล่าวนำ 4

2.2 โครงสร้างพื้นฐานของระบบผู้เชี่ยวชาญ 5

2.3 การแสดงความรู้ในรูปของกฎ 6

2.4 ขั้นตอนการพัฒนาระบบผู้เชี่ยวชาญ 9

3 การประเมินความน่าเชื่อถือของข้อมูล(Data Reliability) 14

3.1 ค่าความน่าเชื่อถือ(Reliability) 14

3.2 ค่าความน่าเชื่อถือของข้อมูลในระบบฐานข้อมูล 16

3.3 วิธีการประเมินความน่าเชื่อถือของข้อมูล 18

4 ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูล
ในแต่ละรายการข้อมูล(Expert System for Internal Reliability) 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

บทที่		หน้า
	4.1 ส่วนประกอบของระบบผู้เชี่ยวชาญลักษณะ R_i	20
	4.2 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญ	24
	4.3 ผลการทดลอง	31
5	ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูล ที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Expert System for Relative Reliability)	36
	5.1 ส่วนประกอบของระบบผู้เชี่ยวชาญลักษณะ R_r	36
	5.2 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญลักษณะ R_r	42
	5.3 ผลการทดลอง	46
6	ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลตามความเป็นจริง (Expert System for Absolute Reliability)	54
	6.1 เทคนิคการสุ่มตัวอย่าง	55
	6.2 การประเมินความน่าเชื่อถือของข้อมูลด้วยการสุ่มตัวอย่าง	60
	6.3 การสร้างระบบผู้เชี่ยวชาญ	60
	6.4 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญ	61
	6.5 ผลการทดลอง	62
7	สรุปผลการวิจัยและข้อเสนอแนะ	67
	7.1 สรุปผลการวิจัย	67
	7.2 ประเมินผลการวิจัย	68
	7.3 ข้อเสนอแนะ	69
	รายการอ้างอิง	72
	ภาคผนวก ก โครงสร้างของระบบฐานข้อมูลนักศึกษา	73
	ภาคผนวก ข บทความที่ส่งตีพิมพ์ในวารสาร	82
	ประวัติผู้เขียน	89

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
1	แสดงสัญญลักษณ์ ชื่อเรียก และความหมายของเครื่องหมายเงื่อนไข	22
2	แสดง inverse condition operater	25
3	แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 1	27
4	แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 2	28
5	แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 3	28
6	แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 4	29
7	แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 5	29
8	แสดงตารางกฎเกณฑ์ที่ใช้ในการทดลองการประเมินความน่าเชื่อถือของข้อมูลชนิด R_i	32
9	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 1	33
10	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 2	34
11	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 3	33
12	แสดงตัวอย่างตารางความสัมพันธ์	38
13	แสดงตัวอย่างตารางเงื่อนไขที่จัดเก็บส่วนของเงื่อนไขในขั้นตอนที่ 3	44
14	แสดงตัวอย่างตารางความสัมพันธ์ที่เก็บส่วนของความสัมพันธ์ของเงื่อนไขในขั้นตอนที่ 3	44
15	แสดงตารางความสัมพันธ์ของเงื่อนไขที่ใช้ในการทดลอง	48
16	แสดงตารางเงื่อนไขที่จัดเก็บส่วนของเงื่อนไขในการทดลอง	49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง(ต่อ)

ตารางที่		หน้า
17	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_r ด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 1	51
18	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_r ด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 2	51
19	แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_r ด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญจากการทดลองครั้งที่ 3	52
20	แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ก. จากจำนวนรายการข้อมูลทั้งหมด 1000 รายการ ในช่วงความคลาดเคลื่อน 0.05	63
21	แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ข. จากจำนวนรายการข้อมูลทั้งหมด 1000 รายการ ในช่วงความคลาดเคลื่อน 0.1	64
22	แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ค. จากจำนวนรายการข้อมูลทั้งหมด 2000 รายการ ในช่วงความคลาดเคลื่อน 0.05	64
23	แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ง. จากจำนวนรายการข้อมูลทั้งหมด 2000 รายการ ในช่วงความคลาดเคลื่อน 0.1	65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

ภาพที่		หน้า
1	แสดงส่วนประกอบและขั้นตอนของระบบประเมินความน่าเชื่อถือของข้อมูล	2
2	แสดงส่วนประกอบของระบบผู้เชี่ยวชาญ	5
3	แสดงโครงสร้างและส่วนประกอบของการแสดงความรู้ในรูปของกฎ	7
4	แสดงขั้นตอนการทำงานของการแสดงความรู้ในรูปของกฎ	8
5	แสดงขั้นตอนการสร้างระบบผู้เชี่ยวชาญ	9
6	แสดงขั้นตอนการทำงานของระบบการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i	30
7	แสดงขั้นตอนการทำงานของระบบการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_t	45
8	แสดงรูปแบบของการประเมินค่าความน่าเชื่อถือของข้อมูลลักษณะ R_i, R_t และ R_a	70

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้ระบบสารสนเทศหรือระบบฐานข้อมูลได้เข้ามามีส่วนสนับสนุนให้การปฏิบัติงานในระบบต่าง ๆ ให้มีประสิทธิภาพมากยิ่งขึ้น ไม่ว่าจะเป็นการปฏิบัติงานในภาครัฐ ภาคเอกชน แม้กระทั่งในภาคการศึกษา โดยเฉพาะอย่างยิ่งในองค์กรที่ใช้ระบบฐานข้อมูลที่ได้มาจากการประมวลผลโดยอาศัยคอมพิวเตอร์ด้วยแล้วจะเห็นถึงความสำคัญของระบบฐานข้อมูลเป็นอย่างสูง หรือจะกล่าวได้ว่าองค์กรที่จะประสบความสำเร็จจะต้องประกอบไปด้วยระบบสารสนเทศหรือระบบฐานข้อมูลที่ถูกต้องสมบูรณ์และมีความน่าเชื่อถือสูง เพราะฉะนั้นการที่สามารถประเมินความน่าเชื่อถือของข้อมูลในระบบฐานข้อมูลในขณะเวลาใดขณะเวลาหนึ่งได้นั้น ก็จะทำให้ทราบว่าสมควรที่จะทำการแก้ไขและปรับปรุงระบบสารสนเทศที่มีอยู่หรือไม่ เพื่อทำให้ได้มาซึ่งระบบสารสนเทศที่ถูกต้องและสมบูรณ์ที่สุด

และเป็นที่ทราบกันดีแล้วว่าทุกวันนี้บุคลากรทางด้านคอมพิวเตอร์ได้ขาดแคลนเป็นจำนวนมาก รวมทั้งผู้เชี่ยวชาญทางด้านคอมพิวเตอร์ในงานแขนงต่าง ๆ โดยรวมไปถึงบุคลากรหรือผู้เชี่ยวชาญที่คอยทำหน้าที่การประเมินความน่าเชื่อถือของข้อมูลด้วย เพราะฉะนั้นในการวิจัยจึงพยายามที่จะสร้างระบบผู้เชี่ยวชาญ(Expert System)เพื่อการประเมินความน่าเชื่อถือของข้อมูลขึ้น เพื่อชดเชยบุคลากรหรือผู้เชี่ยวชาญที่มีอยู่อย่างจำกัด

1.2 วัตถุประสงค์ของการวิจัย

1) เพื่อสร้างระบบการประเมินความน่าเชื่อถือของระบบสารสนเทศหรือระบบฐานข้อมูลโดยใช้ระบบผู้เชี่ยวชาญ

2) เพื่อลดภาระและหน้าที่ของบุคลากรหรือผู้เชี่ยวชาญที่มีหน้าที่ดูแลระบบฐานข้อมูล

3) เพื่อให้ได้มาซึ่งระบบสารสนเทศหรือระบบฐานข้อมูลที่มีความถูกต้องและสมบูรณ์ของข้อมูลมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) เพื่อเป็นแนวทางในการพัฒนาระบบการประเมินความน่าเชื่อถือของระบบสารสนเทศหรือระบบฐานข้อมูลให้กับผู้ที่สนใจต่อไป

1.3 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย

ในการวิจัยพยายามที่จะแบ่งลักษณะการประเมินความน่าเชื่อถือของข้อมูลออกเป็น 3 ลักษณะดังนี้ คือ

1) ความน่าเชื่อถือของข้อมูลในแต่ละรายการข้อมูล (Internal reliability : R_1) เป็นค่าความน่าเชื่อถือของข้อมูลที่เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการ โดยที่ลักษณะพื้นฐานของรายการข้อมูลนี้จะเป็นมาตรฐานที่ยอมรับกันโดยทั่ว ๆ ไป

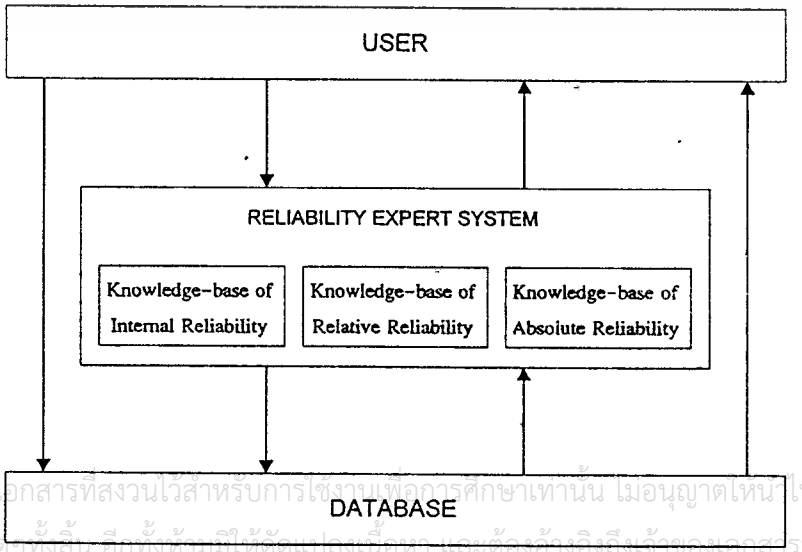
2) ความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Relative reliability : R_2) เป็นค่าความน่าเชื่อถือที่ประเมินได้จากการเปรียบเทียบข้อมูลว่าเป็นจริงตามความสัมพันธ์และเงื่อนไขของระบบฐานข้อมูลที่ได้กำหนดไว้หรือไม่

3) ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความเป็นจริง (Absolute reliability : R_3) เป็นค่าความน่าเชื่อถือของข้อมูลที่ได้จากการเปรียบเทียบข้อมูลที่มีอยู่ในฐานข้อมูลกับข้อมูลที่มีอยู่จริง

เมื่อทราบถึงลักษณะของความน่าเชื่อทั้ง 3 ลักษณะแล้ว ขั้นตอนต่อไปคือพยายามที่จะทำการประเมินหาค่าความน่าเชื่อของข้อมูลทั้ง 3 ลักษณะโดยสร้างเป็นระบบผู้เชี่ยวชาญขึ้น ซึ่งจะมีส่วนประกอบและขั้นตอนการทำงานดังภาพที่ 1

ภาพที่ 1

แสดงส่วนประกอบและขั้นตอนของระบบประเมินความน่าเชื่อถือของข้อมูล



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ขอบเขตของการวิจัย

1) การวิจัยนี้มุ่งพัฒนากับระบบฐานข้อมูลที่มีการทำงานบนเครื่องไมโครคอมพิวเตอร์ หรือกับระบบฐานข้อมูลที่สามารถเชื่อมต่อกับโปรแกรม Microsoft Access เวอร์ชั่น 2.0 ได้

2) การวิจัยนี้จะมุ่งพัฒนากับระบบฐานข้อมูลแบบสัมพันธ์ (Relational Database) และเป็นระบบฐานข้อมูลที่มีความซ้ำซ้อนของข้อมูลน้อยที่สุดโดยผ่านการ Normalization มาแล้วจนอยู่รูปของบรรทัดฐานขั้นที่ 5 (Fifth Normal Form)

3) ทำการสร้างระบบผู้เชี่ยวชาญเพื่อประเมินหาค่าความน่าเชื่อถือข้อมูลทั้ง 3 ลักษณะ โดยการสร้าง rule base แยกกันในแต่ละลักษณะของความน่าเชื่อถือ เพื่อให้ได้มาซึ่งค่าเปอร์เซ็นต์ความน่าเชื่อถือในแต่ละลักษณะ

1.5 วิธีที่ใช้ในการดำเนินการวิจัย

1) ศึกษาถึงทฤษฎีการประเมินความน่าเชื่อถือของข้อมูล

2) ศึกษาถึงทฤษฎีของระบบผู้เชี่ยวชาญที่จะนำมาใช้ช่วยในการประเมินความน่าเชื่อของข้อมูล

3) ทำการสร้างระบบผู้เชี่ยวชาญเพื่อประเมินความน่าเชื่อถือของข้อมูล

4) ศึกษาถึงโครงสร้างและรายละเอียดของระบบฐานข้อมูลที่จะทำการประเมินความน่าเชื่อถือ

5) ทำการทดลองใช้ระบบผู้เชี่ยวชาญที่สร้างขึ้นกับระบบฐานข้อมูลจริง (ซึ่งเป็นระบบฐานข้อมูลนักศึกษา) และเก็บรวบรวมข้อมูล และสรุปผลที่ได้รับ

6) สรุปผลการวิจัย ประเมินผล และเสนอแนะแนวทางในการวิจัยต่อไป

7) เสนอรายงานการวิจัยในรูปวิทยานิพนธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ระบบผู้เชี่ยวชาญ
(Expert System)

2.1 กล่าวนำ

ศาสตราจารย์ Edward Feigenbaum แห่งมหาวิทยาลัยสแตนฟอร์ด ซึ่งเป็นนักค้นคว้าชั้นแนวหน้าในสาขาปัญญาประดิษฐ์ (Artificial Intelligence) ได้ให้คำจำกัดความของระบบผู้เชี่ยวชาญ (Expert System) ไว้ว่า ระบบผู้เชี่ยวชาญคือโปรแกรมคอมพิวเตอร์ที่มีความฉลาดด้วยการใช้ความรู้และขบวนการอนุมาน (inference procedure) ในการแก้ปัญหาที่ยุ่งยากขนาดที่ต้องใช้ประสบการณ์ความชำนาญของมนุษย์จึงจะแก้ได้

กล่าวคือระบบผู้เชี่ยวชาญ คือโปรแกรมคอมพิวเตอร์ที่เก็บทั้งความรู้เกี่ยวกับปัญหาที่จะแก้และขบวนการอนุมานเพื่อนำไปสู่ผลสรุปหรือคำตอบของปัญหานั้น ความรู้ที่เก็บมีทั้งความรู้ที่เป็นความจริงที่อาจจะถูกบันทึกไว้ในรูปของตำราหรือเอกสารทางวิชาการ หรือความรู้ที่ต้องดึงออกมาจากผู้เชี่ยวชาญหรือผู้ชำนาญที่มีประสบการณ์นั้น

ปัญหาที่จะใช้กับระบบผู้เชี่ยวชาญ ส่วนใหญ่จะเป็นปัญหาที่ยุ่งยากและไม่ค่อยมีโครงสร้าง (semi-structured หรือ ill-structured problem) ในปัญหาประเภทนี้คำตอบจะมีโอกาสเป็นได้หลายอย่างทั้งนี้ขึ้นอยู่กับสภาพขณะนั้นของปัญหาและข้อมูลที่เข้ามา ปัญหาประเภทนี้อาจจะอุปมาได้เหมือนกับการเล่นหมากรุก การเดินหมากครั้งต่อไปนั้นจะเดินได้หลายวิธีด้วยกันแต่ตัวหมากที่จะเดินดีที่สุดตัดสินจากสภาพของกระดานหมากในขณะนั้นและหมากที่คิดว่าคู่ต่อสู้จะเดินในครั้งต่อไป ในการแก้ปัญหาประเภทนี้เรามักไม่สามารถจะกำหนดขั้นตอนในการแก้ไขอย่างชัดเจนไว้ล่วงหน้าได้ แต่จะต้องอาศัยความรู้ , ประสบการณ์และสภาพของปัญหาในขณะนั้นรวมกันจึงจะแก้ได้ ดังนั้นวิธีการแก้ปัญหาแบบที่มีมาซึ่งเป็นแบบเขียนโปรแกรมเป็นขั้นตอน การแก้ปัญหาหรืออัลกอริธึม (algorithm) จึงไม่สามารถจะนำมาประยุกต์ใช้ในปัญหาประเภทนี้ได้ ระบบผู้เชี่ยวชาญถึงแม้จะเป็นโปรแกรมคอมพิวเตอร์ชนิดหนึ่ง แต่โครงสร้างและเทคนิคที่ใช้ในการสร้างหรือพัฒนาต่างจากของโปรแกรมที่มีมาและเป้าหมายในการประยุกต์ใช้ก็แตกต่างกัน การประยุกต์ใช้ระบบผู้เชี่ยวชาญที่ประสบความสำเร็จเท่าที่มีมาได้แก่การวินิจฉัยโรค, การสำรวจทรัพยากรธรณี, การวิเคราะห์โครงสร้างสารอินทรีย์เคมี เป็นต้น

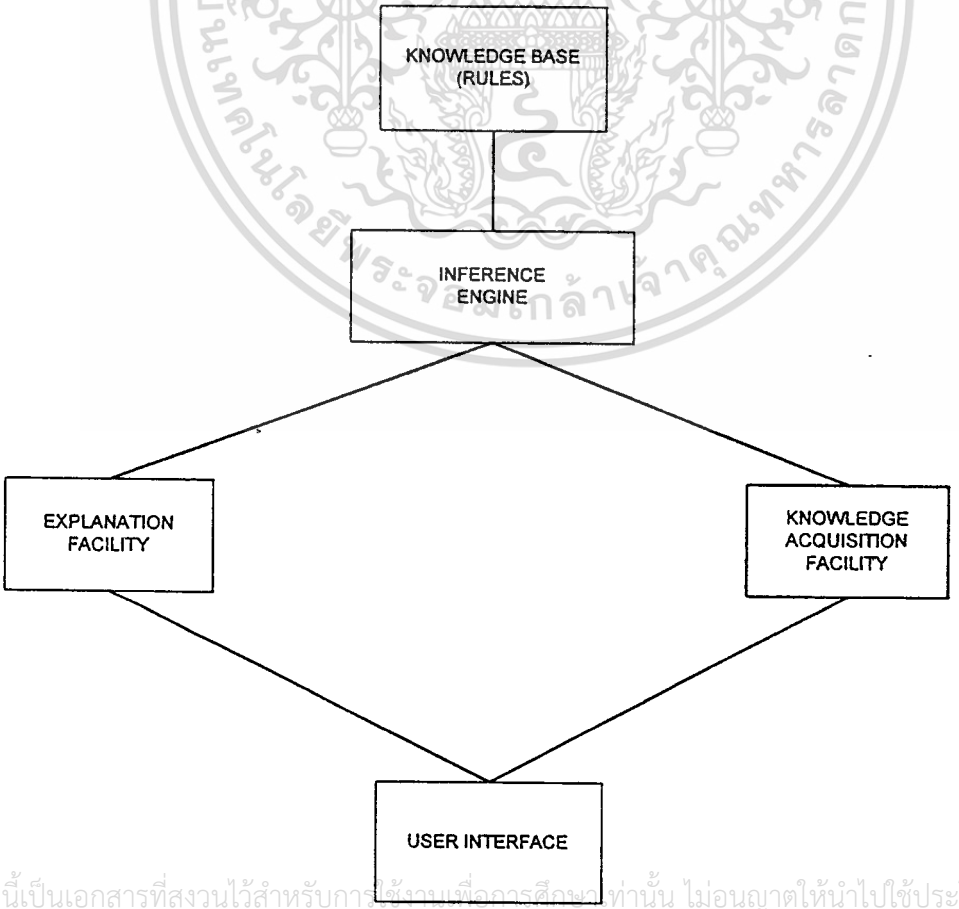
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 โครงสร้างพื้นฐานของระบบผู้เชี่ยวชาญ

ระบบผู้เชี่ยวชาญโดยทั่วไปจะประกอบด้วยส่วนประกอบพื้นฐาน 5 ส่วน ดังแสดงในภาพที่ 2 ส่วนที่เป็นหัวใจที่จะขาดเสียมิได้ คือ ฐานความรู้และเครื่องอนุมาน (inference engine) รายละเอียดโดยย่อของแต่ละส่วนสามารถอธิบายได้ดังนี้

1. **ฐานความรู้** (knowledge base) ส่วนนี้เปรียบเสมือนกับข้อมูลในซอฟต์แวร์ธรรมดาหรือฐานข้อมูล (database) ในระบบสารสนเทศ (information system) เป็นส่วนที่ใช้เก็บความรู้ทุกประเภทไม่ว่าจะเป็นความรู้ที่ได้จากตำราหรือความรู้ที่ได้จากประสบการณ์ ปัญหาหลักของฐานความรู้คือ การเลือกวิธีการแสดงความรู้หรือโครงสร้างสำหรับเก็บความรู้ที่เหมาะสม ปัญหานี้เปรียบได้กับการเลือกโครงสร้างข้อมูลหรือโครงสร้างฐานข้อมูลที่เหมาะสมในระบบซอฟต์แวร์ธรรมดา

ภาพที่ 2
แสดงส่วนประกอบของระบบผู้เชี่ยวชาญ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เครื่องอนุมาน (inference engine) ส่วนนี้เปรียบได้กับอัลกอริธึม เป็นส่วนที่ควบคุมการใช้ความรู้ในฐานความรู้เพื่อแก้ไขปัญหาอย่างมีประสิทธิภาพ วิธีการอนุมานมีหลายแบบแต่แยกเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท คือ อนุมานแบบเดินหน้า (forward chaining inference) และอนุมานแบบย้อนหลัง (backward chaining inference) ทั้งสองวิธีนี้ต่างก็มีจุดดีและจุดเสีย ทั้งนี้ขึ้นอยู่กับลักษณะของปัญหา ในระบบผู้เชี่ยวชาญบางระบบจะใช้วิธีอนุมานทั้งสองวิธีรวมกัน

3. ส่วนดึงความรู้ (knowledge acquisition facility) เป็นส่วนของระบบผู้เชี่ยวชาญที่ใช้ช่วยในการดึงเอาความรู้จากตำราหรือฐานข้อมูลและจากผู้เชี่ยวชาญ การดึงเอาความรู้จากตำราหรือฐานข้อมูลนั้นทำได้ไม่ยาก ถ้าหากเราสามารถจัดความรู้จากแหล่งดังกล่าวให้เป็นระบบ และเข้ากันได้กับโครงสร้างของฐานความรู้ เราก็จะสามารถบรรจุความรู้เหล่านั้นเข้าไปในฐานข้อมูลได้ แต่ทว่าการดึงเอาความรู้จากผู้เชี่ยวชาญนั้นทำได้ยาก จำเป็นต้องใช้เทคนิคต่าง ๆ เข้าช่วยหรือไม่ก็ทำให้ระบบผู้เชี่ยวชาญสามารถเรียนรู้ด้วยตนเองในบางส่วนได้ ปัจจุบันการเรียนรู้ (learning) เป็นหัวข้อค้นคว้าที่นักค้นคว้าในสาขาปัญญาประดิษฐ์ให้ความสนใจมากที่สุดหัวข้อหนึ่ง

4. ส่วนอธิบาย (explanation facility) ส่วนนี้ทำหน้าที่อธิบายรายละเอียดของขั้นตอนการวินิจฉัยต่อผู้ใช้ว่าข้อสรุปหรือคำตอบนั้นได้มาอย่างไรและทำไม

5. ส่วนติดต่อกับผู้ใช้ (user interface) เป็นส่วนที่เป็นตัวกลางระหว่างผู้ใช้กับระบบเพื่อทำให้การสื่อสารระหว่างผู้ใช้กับระบบเป็นไปได้อย่างราบรื่น และช่วยทำให้ผู้ใช้ยอมรับระบบมากขึ้น

ในระบบผู้เชี่ยวชาญบางระบบจะไม่มีส่วนประกอบครบทั้งห้าส่วนดังกล่าวข้างต้น แต่ที่ขาดไม่ได้ คือ ฐานความรู้และเครื่องอนุมาน

2.3 การแสดงความรู้ในรูปของกฎ

การแสดงความรู้ในรูปของกฎมีชื่อเรียกอีกอย่างว่า Production system (PS) PS ถือได้ว่าเป็นโมเดลคอมพิวเตอร์ชนิดหนึ่งที่มีรากฐานทางทฤษฎีอยู่บน post machine ที่เสนอโดย E.L. Post ในปี ค.ศ. 1943 อย่างที่ทราบกันอยู่ว่าเครื่องคอมพิวเตอร์ปัจจุบันมีพื้นฐานทางทฤษฎีอยู่บน Turing machine โดยใน Turing machine เราจะบันทึกขั้นตอนการควบคุมไว้ทั้งหมด การปฏิบัติหรือการประมวลผลของเครื่องจะเป็นไปตามขั้นตอนของการควบคุม แต่ใน post machine ขั้นตอนการปฏิบัติการหรือการประมวลผลจะถูกบันทึกในรูปของเซ็ตของกฎ กฎอันไหนจะถูกใช้ก่อนหลังนั้นไม่ได้ขึ้นอยู่กับการบันทึกกฎ แต่ขึ้นอยู่กับว่าเงื่อนไขของกฎนั้นครบสมบูรณ์หรือไม่ ถ้าครบก็จะมีการปฏิบัติตามกฎนั้น และได้มีคนพิสูจน์แล้วว่าทั้ง Turing และ Post machine มีความสามารถในการประมวลผลเท่าเทียมกัน ดังนั้นความสามารถในการประมวลผลของ PS จึงเท่ากับภาษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมมิงอื่น ๆ เช่น ฟอร์แทรน หรือปาสคาล กล่าวคือ โปรแกรมทุกโปรแกรมที่เขียนด้วยภาษาฟอร์แทรน ปาสคาล หรือภาษาโปรแกรมมิงอื่น ๆ จะแทนได้ด้วยการใช้ PS

การเขียนกฎใน PS จะอยู่ในรูป

IF ..... THEN .....

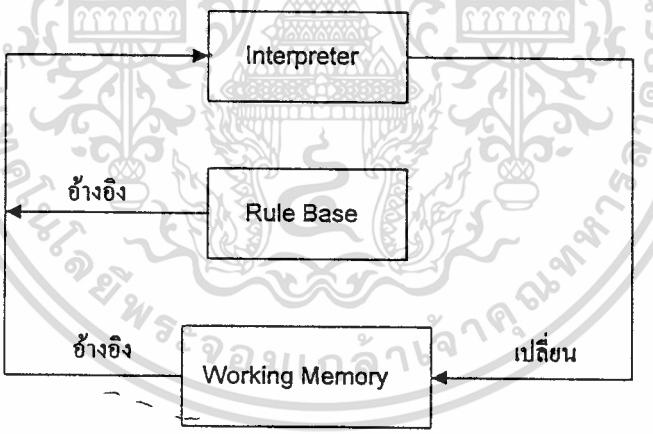
ส่วนของ IF เรียกว่า ส่วนเงื่อนไข และส่วนของ THEN เรียกว่า ส่วนข้อสรุปหรือส่วนการปฏิบัติ

โครงสร้างของ PS จะประกอบด้วยส่วนย่อยหลักสามส่วนด้วยกัน ตามภาพที่ 3 คือ

1. ฐานของกฎ (rule base)
2. ส่วนตีความ (interpreter) หรือ ส่วนอนุมาน
3. ส่วนความจำสำหรับการใช้งาน (working memory)

ภาพที่ 3

แสดงโครงสร้างและส่วนประกอบของการแสดงความรู้ในรูปของกฎ



Rule base เป็นฐานความรู้ที่เก็บอยู่ในรูปของกฎ ส่วน working memory นั้นจะใช้เก็บข้อมูลและสถานะของระบบ โดยที่ข้อมูลและสถานะใน working memory จะเป็นอินพุทของส่วน IF ของกฎ และจะถูกอ้างอิงและเปลี่ยนแปลงโดยกฎใน rule base ซึ่งในส่วนอนุมานนั้นจะใช้ตรวจดูเนื้อหาใน rule base และworking memory แล้วก็จะเลือกกฎใดกฎหนึ่งจากเซตของกฎที่มีเงื่อนไขครบขึ้นมาปฏิบัติการ โดยในการปฏิบัติการแต่ละครั้งจะประกอบด้วยขั้นตอนการทำงานดังต่อไปนี้

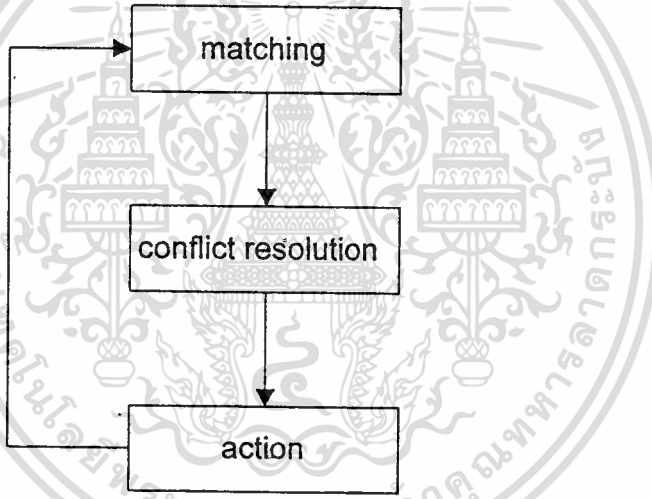
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. matching เป็นการทำการตรวจดูเนื้อหาของ rule base และ working memory เพื่อหากฎทั้งหมดที่มีเงื่อนไขพร้อม

2. conflict resolution เป็นการคัดเลือกกฎจากที่หาได้จากการ matching โดยจะมีการเลือกกฎที่เหมาะสมขึ้นมาหนึ่งกฎ

3. action เป็นการปฏิบัติการตามส่วน THEN ของกฎที่ได้จากการคัดเลือกในข้อ 2 ซึ่งบางการปฏิบัติการอาจเป็นการเปลี่ยนแปลงเนื้อหาของ working memory

ภาพที่ 4
แสดงขั้นตอนการทำงานของการแสดงความรู้ในรูปของกฎ



2.3.1 ทิศทางการอนุมาน

ใน PS มีวิธีการอนุมานสองแบบใหญ่ ๆ ด้วยกัน คือ

1. การอนุมานแบบเดินหน้า (forward chaining) การอนุมานแบบนี้เป็นแบบดั้งเดิมของ PS โดยความเป็นจริงแล้วภาพที่ 4 แสดงวงจรการปฏิบัติการของ PS ภายใต้การอนุมานแบบเดินหน้ากล่าวคือ PS จะเริ่มต้นจาก working memory ไปสู่ขั้นตอนการเลือกกฎ และ ปฏิบัติการตามกฎ โดยจะปฏิบัติการซ้ำ ๆ กันเช่นนี้จนกว่าจะได้คำตอบหรือบรรลุเป้าหมาย การอนุมานแบบนี้มีชื่ออื่นเรียกอีกว่า data-driven inference หรือ bottom-up inference

2. การอนุมานแบบย้อนหลัง (backward chaining) การอนุมานแบบนี้เริ่มต้นจากเป้าหมายไปสู่กฎที่ทำให้เป้าหมายบรรลุผล จนกระทั่งไปสู่เป้าหมายย่อยและกฎที่ทำให้เป้าหมายย่อยทั้ง

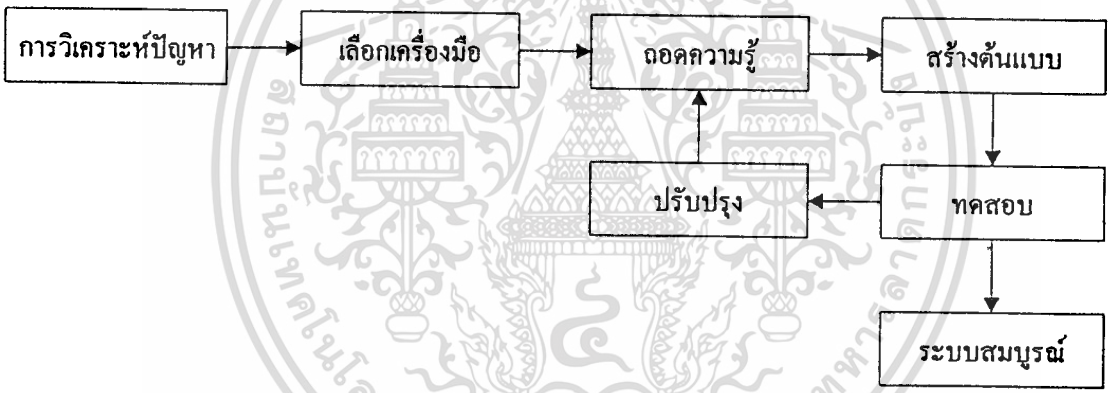
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมดบรรลุผล ในกรณีที่ค้นไม่พบข้อมูลหรือไม่บรรลุผล อาจจะต้องทำการย้อนรอย (back-goal-driven inference) หรือ top-down inference

2.4 ขั้นตอนการพัฒนาระบบผู้เชี่ยวชาญ

การกล่าวถึงการพัฒนาระบบผู้เชี่ยวชาญในที่นี้จะเป็นการกล่าวถึงเฉพาะในส่วนของการพัฒนาระบบที่ต้องอาศัยวิศวกรความรู้ ซึ่งมีขบวนการต่าง ๆ ของการพัฒนาระบบดังแสดงไว้ในภาพที่ 5

ภาพที่ 5
แสดงขั้นตอนการสร้างระบบผู้เชี่ยวชาญ



รูปที่ 5 เป็นการแสดงขั้นตอนการพัฒนาระบบผู้เชี่ยวชาญ เริ่มต้นด้วยการวิเคราะห์ปัญหาเลือกเครื่องมือ จากนั้นก็จะเป็น ขบวนการถอดความรู้ การสร้างระบบต้นแบบ แล้วก็จะทำการทดสอบ ถ้าหากว่าระบบต้นแบบที่ได้ยังไม่ถูกต้อง ต้นแบบนี้จะต้องมีการปรับปรุงแก้ไขระบบใหม่โดยเริ่มต้นจากการถอดความรู้ ปรับปรุงระบบต้นแบบ และทดสอบใหม่จนกระทั่งได้ต้นแบบที่ถูกต้อง

เมื่อได้ต้นแบบที่ถูกต้องแล้ว จึงจะทำการขยายระบบให้เป็นระบบที่สมบูรณ์แล้วก็ประเมินผล ถ้าระบบที่ขยายขึ้นมามีอะไรต้องแก้ไขปรับปรุง แล้วทำการประเมินผลใหม่จนกระทั่งได้ผลเป็นที่พอใจ เมื่อได้ระบบที่พอใจแล้ว ก็ติดตั้งระบบ และวางแผนการบำรุงรักษา เหตุผลของการบำรุงรักษาระบบก็เพื่อทำให้ระบบผู้เชี่ยวชาญนี้มีความรู้ที่ทันสมัยเสมอ ขั้นตอนของการพัฒนาระบบดังที่ได้แสดงในรูปที่ 5 มีรายละเอียดตามหัวข้อต่าง ๆ ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การวิเคราะห์ปัญหา

การวิเคราะห์ปัญหาของระบบผู้เชี่ยวชาญนั้น ผู้พัฒนาระบบจะต้องทำความเข้าใจเบื้องต้นกับปัญหาเหล่านั้น และจะต้องมีการเตรียมการสำหรับการแก้ปัญหานั้นด้วย เรื่องที่สำคัญสำหรับการวิเคราะห์ปัญหาที่จะต้องพิจารณาก็คือ ความจำเป็นหรือความเหมาะสมของการใช้ระบบผู้เชี่ยวชาญการมีความรู้ความเข้าใจปัญหาอย่างถูกต้อง การจัดขั้นตอนสำหรับการแก้ปัญหาและการกำหนดรูปแบบของการให้คำปรึกษา

สำหรับการแก้ปัญหาโดยทั่วไปแล้ว การใช้ระบบผู้เชี่ยวชาญหรือปัญญาประดิษฐ์ อาจจะไม่ใช่ทางออกที่ถูกต้องเสมอไป ปัญหาบางชนิดเหมาะกับการใช้คณิตศาสตร์ ปัญหาบางชนิดเหมาะกับการใช้สูตรสำเร็จ แต่ปัญหาที่ใช้ได้กับระบบผู้เชี่ยวชาญและปัญญาประดิษฐ์คือปัญหาที่ต้องอาศัยฮิวริสติค โดยปกติการที่จะกำหนดว่าปัญหาชนิดใดที่เหมาะกับระบบผู้เชี่ยวชาญนั้น ผู้พัฒนาระบบอาจจะอาศัยหลักการข้อใดข้อหนึ่งดังต่อไปนี้ มาช่วยในการพิจารณาได้ คือ

1. ปัญหาเหล่านี้จะต้องเป็นปัญหาที่ไร้โครงสร้าง คือ ไม่อาจแก้ได้ด้วยสูตรสำเร็จหรือคณิตศาสตร์แต่จะต้องอาศัยฮิวริสติค

2. ปัญหาเกี่ยวข้องกับการวินิจฉัย (inference) หรือการวิเคราะห์ข้อมูล (data analysis)

3. ปัญหาที่เกี่ยวพันธ์กับความรู้ที่ไม่แน่นอน (uncertainty knowledge)

4. ผู้เชี่ยวชาญสามารถแก้ปัญหานั้นได้ในช่วงเวลาที่แน่นอน

5. ความรู้ในการแก้ปัญหาสามารถจัดให้อยู่ในโครงสร้างของ if-then ได้

นอกจากความเข้าใจในเรื่องของการแก้ปัญหาแล้ว การมีความรู้และความเข้าใจในเนื้อหาที่นำมาพัฒนาเป็นระบบผู้เชี่ยวชาญ ก็จะเป็นเรื่องที่ต้องพิจารณาตามมา ผู้พัฒนาจะต้องมีความรู้และความเข้าใจในเรื่องที่กำลังจะทำการพัฒนาเป็นอย่างดี ระบบผู้เชี่ยวชาญจะเกิดขึ้นไม่ได้ถ้าไม่มีผู้เชี่ยวชาญเป็นผู้ช่วย สิ่งที่ผู้เชี่ยวชาญต้องทราบเกี่ยวกับความรู้นั้นคือ ต้องรู้ถึงลำดับของเนื้อหา และขบวนการของการแก้ปัญหาในเรื่องนั้น ๆ

2. การเลือกเครื่องมือ

การพัฒนาระบบผู้เชี่ยวชาญในปัจจุบันได้มีการเปลี่ยนแปลงอย่างรวดเร็วทั้งในแง่ความง่ายและความเร็วที่ใช้ในการพัฒนาระบบ ทั้งนี้เนื่องจากการพัฒนาอย่างรวดเร็วของเครื่องมือที่ใช้ในการพัฒนาระบบผู้เชี่ยวชาญ (Expert System Building Tools; ESBTS) นั้นเอง เครื่องมือ (Tools) เหล่านี้โดยส่วนใหญ่แล้วมักจะถูกสร้างขึ้นมาเพื่อการค้า และมีลักษณะพิเศษในการใช้งานที่ต่างกัน ทั้งนี้ขึ้นอยู่กับวิธีการที่ใช้ในการสร้างเครื่องมือเหล่านี้ โดยในการเลือกเครื่องมือนั้นจะต้องพิจารณาถึงสิ่งต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1 ส่วนแสดงความรู้

เป็นเงื่อนไขประการแรกที่ใช้ในการพิจารณาเลือกเครื่องมือที่ใช้ในการสร้างระบบผู้เชี่ยวชาญ วิธีการของการแสดงความรู้นั้นจะต้องง่ายและสามารถใส่เรื่องที่ต้องการได้ครบถ้วน การแสดงความรู้อย่างน้อยจะต้องสามารถจัดการกับเรื่องเหล่านี้ได้คือ การบรรยายประสงค์ (Object Descriptions) การกำหนดค่าความมั่นใจ (Certainties) และการแสดงการกระทำ (Actions) เรื่องทั้ง 3 นี้เป็นหัวข้อพื้นฐานของการแสดงความรู้ที่จะต้องมี

2.2 เครื่องอนุมาน

เครื่องอนุมานเป็นอีกส่วนที่สามารถกำหนดความสามารถของเครื่องมือได้ โดยปกติแล้วเครื่องอนุมานมีส่วนสำคัญที่ผู้พัฒนาระบบจะต้องคำนึงถึงคือ วิธีการอนุมาน การค้นหากฎข้อที่เหมาะสม การกำหนดทางคณิตศาสตร์ และการเชื่อมโยงกับโปรแกรมอื่น

2.3 การติดต่อกับผู้ใช้

การติดต่อกับผู้ใช้ของระบบผู้เชี่ยวชาญ เป็นส่วนที่ระบบจะติดต่อกับผู้ที่มาขอคำปรึกษา ระบบที่ดีจะต้องมีส่วนที่ทำให้ผู้ใช้สะดวกต่อการขอคำปรึกษา มีการแสดงการตอบโต้ที่ง่ายและชัดเจนในเรื่องต่าง ๆ เหล่านี้ ของเครื่องมือที่ได้มีการพัฒนามาแล้วนั้น การติดต่อกับผู้ใช้สามารถจำแนกได้ออกเป็นชนิดคือ วิธีการโต้ตอบระหว่างผู้ใช้และระบบ การแสดงภาพในระหว่างการขอคำปรึกษา และการเก็บความรู้จากคำตอบของผู้ใช้

2.4 ลักษณะทางซอฟท์แวร์

ในการเลือกเครื่องมือช่วยในการพัฒนาระบบผู้เชี่ยวชาญ ลักษณะทางซอฟท์แวร์เป็นอีกส่วนหนึ่งที่มีความสำคัญ เพราะเป็นลักษณะทางซอฟท์แวร์นี้จะบอกถึงธรรมชาติของระบบนั้น ๆ ว่ามีข้อดีและข้อจำกัดอย่างไร ลักษณะทางซอฟท์แวร์นี้มีเรื่องที่น่าสนใจที่จะต้องพิจารณาคือ ภาษาคอมพิวเตอร์ที่ใช้ในการสร้างเครื่องมือ ความสามารถในการคอมไพล์ ความสามารถในการขยายระบบการใช้ร่วมกับภาษาคอมพิวเตอร์อื่น และระบบจัดการที่ใช้งานร่วมกับเครื่องมือดังกล่าว

2.5 ความสามารถในการติดต่อกับผู้พัฒนาระบบ

ในส่วนนี้จะเป็นส่วนที่เกี่ยวข้องกับผู้พัฒนาระบบโดยตรง สิ่งที่จะต้องพิจารณาในส่วนนี้คือ วิธีการสร้างฐานความรู้ เครื่องมือที่ใช้ในการแก้ไขฐานความรู้ ความสามารถในการสร้างส่วนที่ใช้ติดต่อกับผู้ใช้ และส่วนควบคุมการอนุมาน

3. ขบวนการถอดความรู้

ขบวนการถอดความรู้หมายถึง ขบวนการที่ผู้พัฒนาระบบเรียนรู้และทำความเข้าใจกับความรู้ที่จะนำมาพัฒนาเป็นระบบผู้เชี่ยวชาญในการเรียนรู้และทำความเข้าใจกับความรู้นั้น ผู้พัฒนาสามารถหาได้จากแหล่งความรู้ต่าง ๆ เช่น ผู้เชี่ยวชาญ หนังสือ เป็นต้น นอกจากนั้นแล้ว ผู้พัฒนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบยังจะต้องกำหนดขอบเขตที่แน่นอนของการสร้างความรู้นั้นและขอบเขตนั้นจะต้องสามารถใช้ได้ดีกับระบบผู้เชี่ยวชาญ ดังที่กล่าวมาแล้ว

เมื่อผู้พัฒนาระบบสามารถหาแหล่งและกำหนดขอบเขตของความรู้ได้แล้ว ขั้นตอนต่อไปคือ จำแนกโดเมนของปัญหา และปัญหาเฉพาะของงาน ในการเลือกปัญหาที่ถูกต้องบางที่อาจจะเป็นส่วนที่สำคัญที่สุดของการพัฒนาระบบ ในขั้นตอนนี้เราเข้าใจว่าเทคโนโลยีทางด้านนี้ยังมีข้อจำกัดอยู่มาก ถ้าหากว่าเลือกปัญหาไม่ถูกต้อง บางครั้งการพัฒนาระบบผู้เชี่ยวชาญจะเกิดปัญหาขึ้นโดยที่ใครก็ไม่สามารถช่วยแก้ไขได้ ดังนั้น ในการเลือกปัญหาจึงจำเป็นอย่างยิ่งที่จะต้องพิจารณาความสามารถของระบบผู้เชี่ยวชาญ เพราะในระบบใหญ่ ถ้าหากว่าการเลือกปัญหาผิด อาจจะทำให้ระบบทั้งระบบต้องล้มเหลวได้อีกเรื่องที่ถือว่ามีความสำคัญมากของขบวนการถอดความรู้ก็คือ ในการพัฒนาระบบต้องมีผู้เชี่ยวชาญที่พร้อมจะให้การช่วยเหลือ ดังที่กล่าวมาแล้วว่าระบบผู้เชี่ยวชาญเป็นระบบที่สร้างขึ้นมาเพื่อพยายามเลียนแบบการทำงานของผู้เชี่ยวชาญให้ความช่วยเหลือในการพัฒนา ระบบใหญ่นี้จะแตกต่างจากระบบเล็กซึ่งวิศวกรความรู้สามารถประมวลความรู้ด้วยตนเองและหาความรู้เพิ่มเติมจากแหล่งภายนอกได้ แต่ในระบบใหญ่แล้ว ความรู้ที่ใส่จะต้องเป็นความรู้ที่ได้มาจากผู้เชี่ยวชาญที่มีประสบการณ์ในความรู้นั้น ๆ อย่างดีและสามารถเข้าใจและรู้วิธีแก้ปัญหานั้น ๆ ได้เช่นกัน

4. การสร้างต้นแบบ

ในการสร้างต้นแบบ ผู้พัฒนาระบบควรเริ่มต้นจากกระดาษ เขียนแนวความคิดของความรู้ทั้งหมดที่เราจะสร้างโดยเริ่มต้นจาก

เป้าหมาย (Goal) ที่เด่นชัด เป้าหมายในที่นี้คือ จุดหมายปลายทางของระบบการใช้ คำปรึกษานี้จะเป็นเช่นไร หรือพูดง่าย ๆ คือคำตอบของการให้คำปรึกษานั่นเอง และคำตอบนี้จะมีอยู่หลาย ๆ คำตอบ ซึ่งระบบผู้เชี่ยวชาญจะเป็นผู้เลือกให้สอดคล้องกับลักษณะเฉพาะของปัญหา

กำหนดการไหลของไดอะแกรม (flow diagram) ของปัญหาทั้งหมาด ในขั้นนี้เป็นขั้นตอนของการแสดงความรู้ การแสดงความรู้ต้องจัดลำดับของขั้นตอนที่จำเป็นออกมา โนดแต่ละโนดจะเป็นแอตตริบิวต์ที่ต้องมีการอนุมาน ในกรณีที่ความรู้มีมากและโดเมนใหญ่ การหาค่าของแอตตริบิวต์ไม่จำเป็นที่จะต้องได้มาจากผู้ใช้ แอตตริบิวต์เหล่านี้อาจจะหาได้จากการอนุมานของเครื่องได้

ในการสร้างต้นแบบนั้นเป็นการแสดงความรู้เฉพาะส่วนขึ้นมา โดยการจำกัดโดเมนของความรู้ให้แคบลง การสร้างระบบต้นแบบนั้นไม่จัดประสงค์เพื่อหาความเป็นไปได้ของการสร้างระบบและหาหนทางในการแก้ปัญหาก่อนที่จะสร้างระบบจริง ระบบต้นแบบที่สร้างขึ้นนี้ จะต้องมีลักษณะการทำงานที่เหมือนระบบจริงที่จะพัฒนาต่อ แต่กำหนดให้ขอบเขตการแก้ปัญหาทำได้น้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กว่า ระบบต้นแบบนี้จะเป็นต้นแบบที่ใช้ในการทดสอบว่าการแก้ปัญหาที่ได้ทำการออกแบบมานั้นถูกต้องหรือไม่ และเพื่อเป็นแนวทางในการขยายระบบต่อไป

5. ขยาย ทดสอบ และปรับปรุงระบบ

การขยายระบบโดยการนำต้นแบบที่แน่ใจว่าถูกต้องแล้วมาทำการเพิ่มองค์ประกอบต่าง ๆ จนกระทั่งเป็นระบบที่สมบูรณ์ตามที่ได้มีการวางแผนไว้ โดยการเติมความรู้ในส่วนที่ยังขาดอยู่และตกแต่งระบบให้ดูประณีต เพิ่มส่วนที่ใช้ในการอธิบายส่วนต่าง ๆ การพัฒนาระบบผู้เชี่ยวชาญขนาดใหญ่ก่อนที่จะมีการขยายระบบต้นแบบนี้จะต้องมีการตรวจสอบโดยผู้เชี่ยวชาญและวิศวกรความรู้อย่างละเอียดด้วยการนำเงื่อนไขต่าง ๆ ที่ได้วางไว้ในการสร้างระบบต้นแบบมาทำการทดสอบ และตรวจโดยผู้เชี่ยวชาญเพื่อดูว่าเงื่อนไขต่าง ๆ ที่ทดสอบนั้นถูกต้องหรือไม่ ถ้าหากว่าระบบต้นแบบมีความคลาดเคลื่อนจากการวางระบบเอาไว้ ก็จะต้องวกกลับไปทำการออกแบบระบบต้นใหม่ สำหรับการทดสอบระบบต้นแบบ มีสิ่งที่พึงระลึกไว้เสมอว่า ระบบนี้ได้มีการจำลองระบบให้มีขอบเขตของการแก้ปัญหาที่เล็กลงกว่าระบบจริง ดังนั้นเงื่อนไขในการทดสอบบางอย่างที่ไม่ได้กำหนดไว้ในการสร้างระบบต้นแบบก็จะนำมาตรวจสอบไม่ได้ การประเมินผลของระบบเมื่อระบบสร้างแล้วยังต้องมีการประเมินผลด้วยว่าระบบที่ได้ออกมานี้เป็นไปตามความต้องการของผู้ออกแบบระบบหรือไม่ ในการตรวจสอบผู้ตรวจสอบจะต้องมีผู้เชี่ยวชาญที่มาช่วยในการพัฒนาระบบมาให้คำปรึกษาอย่างใกล้ชิด วิศวกรความรู้จะต้องตรวจสอบเงื่อนไขต่าง ๆ ของการอนุมานให้ครบถ้วน และผู้เชี่ยวชาญจะต้องตรวจสอบความรู้ทุกอย่างที่มีอยู่ในระบบว่า ตรงกับความเป็นจริงหรือไม่ ถ้าหากว่าเกิดความผิดพลาดขึ้น วิศวกรความรู้จะต้องเป็นผู้แก้ไขกฎหรือข้อมูลต่าง ๆ ในฐานความรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การประเมินความน่าเชื่อถือของข้อมูล

(Data Reliability)

3.1 ค่าความน่าเชื่อถือ (Reliability)

เริ่มแรกนั้นค่าความน่าเชื่อถือ(Reliability)มีการกล่าวถึงกันในด้านของวิศวกรรมการผลิต โดยจะปรากฏอยู่ในแขนงของการควบคุมคุณภาพการผลิตหรือที่เรียกว่า Quality Control (QC) โดย (Robertson 1971 : 137-140) ได้นิยามความหมายของคำว่าความน่าเชื่อถือไว้ว่า "ความน่าเชื่อถือหมายถึงค่าความสามารถที่ทำให้ผลิตภัณฑ์มีความถูกต้องสมบูรณ์ภายใต้เงื่อนไขที่ต้องการ และในช่วงเวลาที่กำหนด"ส่วน (Vaughn 1974 : 197-198) ได้นิยามความหมายของค่าความน่าเชื่อถือในลักษณะของค่าความน่าจะเป็น(Probability)โดยให้คำนิยามไว้ว่า "ความน่าเชื่อถือหมายถึงค่าความน่าจะเป็นที่จะเกิดความถูกต้องสมบูรณ์ของผลิตภัณฑ์ภายใต้เงื่อนไขที่ต้องการและในช่วงเวลาที่กำหนด"

จากคำนิยามความหมายค่าความน่าเชื่อถือของ Vaughn สามารถที่จะเขียนเป็นสมการของค่าความน่าเชื่อถือได้ดังนี้

$$R = 1 - P_f \tag{3.1}$$

เมื่อ

P_f คือ ค่าความน่าจะเป็นที่จะเกิดข้อผิดพลาดของผลิตภัณฑ์

โดยที่

$$P_f = \frac{N_f}{N_t} \tag{3.2}$$

เมื่อ

N_f คือ จำนวนผลิตภัณฑ์ที่เกิดการผิดพลาด และ

N_t คือ จำนวนผลิตภัณฑ์ทั้งหมดที่ใช้ทดสอบ (Items tested) ในช่วงเวลา t

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อค่านำสมการ 3.2 แทนลงในสมการ 3.1 จะได้

$$R = 1 - \frac{N_f}{N_t} \qquad (3.3)$$

3.2 ค่าความน่าเชื่อถือของข้อมูลในระบบฐานข้อมูล (Data Reliability in Database System)

จากทฤษฎีการหาค่าความน่าเชื่อที่กล่าวมาข้างต้นนั้น สามารถนำมาประยุกต์ใช้ในระบบฐานข้อมูลได้ โดยการแปลงจำนวนผลิตภัณฑ์ที่ใช้ในการทดสอบหาค่าความน่าเชื่อถือในช่วงเวลาที่กำหนด(N_t)เป็นจำนวนรายการข้อมูลทั้งหมดที่ใช้ทดสอบ(N) เพราะในการประเมินหาค่าความน่าเชื่อถือของข้อมูลในระบบฐานข้อมูลนั้นจะไม่ทำการตรวจสอบกับรายการข้อมูลในหน่วยของเวลา(t) แต่จะตรวจสอบกับข้อมูลทั้งหมดที่มีอยู่ในระบบฐานข้อมูล เพราะฉะนั้นจากสมการที่ 3.3 สามารถประยุกต์ให้อยู่ในรูปของสมการที่ใช้คำนวนหาค่าความน่าเชื่อของข้อมูลได้ดังนี้

$$R = 1 - \frac{N_f}{N} \qquad (3.4)$$

เมื่อ

N_f คือ จำนวนผลิตภัณฑ์ที่เกิดการผิดพลาด และ

N คือ จำนวนรายการข้อมูลทั้งหมดที่ใช้ทดสอบ

ค่าความน่าเชื่อถือของข้อมูลนั้นจะขึ้นอยู่กับข้อผิดพลาดที่เกิดขึ้นกับข้อมูลที่เก็บอยู่ในระบบฐานข้อมูล เพราะฉะนั้นการประเมินหาค่าความน่าเชื่อถือของข้อมูลก็คือการประเมินหารายการข้อมูลที่เกิดการผิดพลาดนั่นเอง

3.2.1 สาเหตุของการเกิดข้อผิดพลาดในฐานข้อมูล

ปัญหาที่อาจทำให้เกิดข้อผิดพลาดขึ้นภายในฐานข้อมูล (database corruption) เป็นไปได้จากหลายสาเหตุด้วยกัน ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ข้อผิดพลาดจากผู้เขียนโปรแกรม(programmer errors) ข้อผิดพลาดที่เกิดจากผู้เขียนโปรแกรมทำให้เกิดปัญหาด้านความถูกต้องสมบูรณ์ของข้อมูลได้ในหลายรูปแบบ เช่น เขียนโปรแกรมค่าเก็บไว้ผิดฟิลด์ ทำการคำนวณค่าต่าง ๆ ผิดพลาดคลาดเคลื่อน เชื่อมโยงความสัมพันธ์ระหว่างฟิลด์และไฟล์ผิดพลาด หรือใช้ตรรกในการเขียนโปรแกรมไม่ถูกต้อง เป็นต้น

2) ข้อผิดพลาดจากผู้ปฏิบัติงาน(operation errors) เจ้าหน้าที่ควบคุมการทำงานและผู้ออกแบบระบบอาจเป็นผู้ที่ก่อให้เกิดข้อผิดพลาดขึ้นได้หลายรูปแบบ ตัวอย่างเช่น ในกรณีการสั่งปิดระบบใช้งานในขณะที่ตัวระบบยังทำการประมวลผลไม่เรียบร้อย อาจทำให้เกิดภาวะปัญหาข้อมูลคลาดเคลื่อน (data inconsistency) เพราะยังทำงานไม่สมบูรณ์ครบวงจรของระบบจัดการฐานข้อมูลทั้งหมด

3) ข้อผิดพลาดจากข้อมูลเข้า (input errors) ดังคำกล่าวที่รู้จักกันดี "ขยะเข้ามาก็ได้ขยะออกไป" นั่นคือ สารสนเทศ (information) ที่จะได้ออกไปจากระบบก็จะต้องเกิดจากการใส่ข้อมูลเข้ามาอย่างถูกต้อง

4) ข้อผิดพลาดจากระบบฐานข้อมูล (DBMS errors) ระบบจัดการฐานข้อมูลก็คล้ายคลึงกับโปรแกรมซอฟต์แวร์อื่นๆทั่วไปคือ มีโอกาสที่จะมีข้อผิดพลาดทางตรรก (logic errors) ได้ หากเกิดข้อผิดพลาดประเภทนี้ขึ้น ข้อมูลในฐานข้อมูลก็อาจจะสูญหาย หรือคลาดเคลื่อนบิดเบือนไปจากความเป็นจริงได้ ตัวอย่างเช่น เมื่อมีการแก้ไขข้อมูลในแฟ้มข้อมูลและแฟ้มดัชนีแล้ว อาจเกิดปัญหาว่าผู้ใช้ไม่สามารถเรียกใช้เรคอร์ดที่ถูกแก้ไขแล้วได้ เพราะมีข้อผิดพลาดทางตรรกในการกำหนดดัชนีดังกล่าว

5) ข้อผิดพลาดจากขบวนการทำงาน (system or process failure) ความคลาดเคลื่อนของข้อมูลอาจเป็นผลลัพธ์จากข้อผิดพลาดของสื่อ (แผ่นดิสก์) หรือความล้มเหลวของกระบวนการทั้งหมด ขณะที่ผู้ป้อนข้อมูลกำลังใส่ข้อมูลอยู่และเกิดกระแสไฟฟ้ากระตุกหรือดับลงกะทันหัน ความสูญเสียหรือข้อผิดพลาดอาจจะเกิดขึ้นกับข้อมูลหรือเรคอร์ดที่กำลังถูกดำเนินการอยู่ ในบางกรณีก็ไม่อาจแน่ใจได้ว่าการกระทำ transaction สมบูรณ์หรือยังเมื่อเกิดภาวะระบบล้มเหลวขึ้น

6) ข้อผิดพลาดจากฮาร์ดแวร์ (hardware errors and disasters) ในหลายกรณีข้อผิดพลาดเกิดขึ้นนอกเหนือการควบคุมของผู้เขียนโปรแกรมหรือผู้ออกแบบระบบ เช่น มีข้อผิดพลาดเกิดขึ้นในส่วนที่ทำการส่งผ่านอ่านเขียนข้อมูลระหว่างตัวเครื่องกับแผ่น (transmission error) อาจทำให้ได้ผลลัพธ์ค่าข้อมูลในฐานข้อมูลผิดพลาดไป หรือเกิดข้อผิดพลาดในหน่วยความจำของเครื่องก็อาจทำให้เกิดปัญหาได้ นอกจากนี้อันตรายที่เกิดจากภัยธรรมชาติ (disaster) ก็อาจทำให้ฐานข้อมูลเสียหายได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) ข้อผิดพลาดโดยรวม (compounded errors) เป็นความผิดพลาดเกิดจากสาเหตุต่าง ๆ หลายสาเหตุ เกิดผสมผสานกันไป ตัวอย่างเช่น มีใบสั่งสินค้า สั่งสินค้าเข้ามา 10 ชิ้น เสมียนกรอกจำนวนผิดเป็น100 ชิ้น ทำให้สินค้าถูกถอนออกมาจากคลังมากเกินไป ทำให้ customer file คลาดเคลื่อน เครดิตของลูกค้าอาจถูกลดลงอย่างรวดเร็ว ไม่สามารถสั่งสินค้าครั้งต่อไปได้ เพราะเครดิตไม่พอหรือจำนวนสินค้าจะถูกสั่งเพิ่มเข้ามาเพื่อเก็บในคลังมากมายเกินต้องการ เพราะระบบเข้าใจผิดว่าสินค้าที่มีอยู่ในคลังลดลงไปเป็นจำนวนมาก เป็นต้น

3.2.2 วิธีการตรวจสอบความถูกต้องสมบูรณ์ของฐานข้อมูล

ความสำเร็จของระบบจัดการฐานข้อมูล นอกจากจะขึ้นอยู่กับรูปแบบที่นำเสนอต่อผู้ใช้ได้ตรงตามความต้องการของผู้ใช้แล้ว ความถูกต้องของข้อมูลก็เป็นสิ่งจำเป็นมาก ระบบฐานข้อมูลที่ดีจะต้องมีการกำหนดวิธีการตรวจสอบความถูกต้องสมบูรณ์ เทคนิคสำคัญในการตรวจสอบระบบแบ่งได้เป็น 2 ด้าน คือ

1) domain integrity check เป็นการเช็คความถูกต้องของการจัดฟิลด์สำหรับใส่ข้อมูลในเรคอร์ด กฎเกณฑ์การกำหนดฟิลด์นี้จะต้องกำหนดโดยผู้ออกแบบระบบ (เช่น กำหนดประเภทของแต่ละฟิลด์ว่าเป็นประเภทใด ตัวเลข ตัวอักษร หรืออื่น ๆ) การกำหนดช่วงของค่าที่เป็นไปได้ขนาดความยาว กำหนดรูปแบบของค่าภายในฟิลด์ (เช่น โทรศัพท์ใช้รูปแบบ 999-9999 เป็นต้น) ซอฟต์แวร์ที่ใช้สร้างระบบจัดการฐานข้อมูลส่วนใหญ่มีวิธีการตรวจสอบ domain integrity จัดไว้เป็นส่วนหนึ่งของโปรแกรมอยู่แล้ว แต่อย่างไรก็ตามผู้ออกแบบระบบจะต้องเป็นผู้ออกแบบข้อกำหนดลักษณะของแต่ละฟิลด์ต่าง ๆ ดังกล่าว

2) referential integrity check เป็นการเช็คความถูกต้องของข้อมูลก่อนที่จะสั่งดำเนินการใด ๆ ว่าเป็นไปตามข้อกำหนดกฎเกณฑ์หรือไม่ โดยผู้ออกแบบและเขียนโปรแกรมการทำงานของระบบจะต้องวางแผนวิธีการควบคุมดังกล่าวล่วงหน้า ก่อนลงมือเขียนโปรแกรมใช้งาน

3.3 วิธีการประเมินความน่าเชื่อถือของข้อมูล

เนื่องจากลักษณะของข้อมูลที่มีเก็บอยู่ในระบบฐานข้อมูลนั้นมีลักษณะที่แตกต่างกันมากมาย เพราะฉะนั้นในการประเมินค่าความน่าเชื่อถือของข้อมูลนั้นจะต้องแบ่งวิธีการประเมินออกลักษณะย่อย ๆ เพื่อให้ครอบคลุมกับลักษณะของข้อมูลที่มีอยู่ โดยในการทำวิจัยจะแยกวิธีการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเมินค่าความน่าเชื่อถือของข้อมูลออกตามลักษณะของข้อมูลได้เป็น 3 ลักษณะ ซึ่งในแต่ลักษณะจะมีรายละเอียดดังนี้

1) ความน่าเชื่อถือของข้อมูลในแต่ละรายการข้อมูล (Internal Reliability : R_i)

เป็นค่าความน่าเชื่อถือของข้อมูล ที่เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการ โดยที่ลักษณะพื้นฐานของรายการข้อมูลนี้จะเป็นมาตรฐานที่ยอมรับกันโดยทั่ว ๆ ไป

R_i เป็นค่าความน่าเชื่อถือที่ประเมินได้โดยการใช้ลักษณะพื้นฐานของข้อมูลแต่ละรายการเป็นตัวประเมินหา โดยที่ลักษณะพื้นฐานดังกล่าวจะขึ้นอยู่กับข้อมูลแต่ละชนิดแต่ละรายการและเป็นลักษณะที่บุคคลทั่วไปสามารถวินิจฉัยไปในแนวทางเดียวกันได้ ตัวอย่างเช่น รายการข้อมูลจำนวนนักศึกษา ในฐานข้อมูลทะเบียนนักศึกษาจะต้องเป็นค่าบวกจะเป็นลบไม่ได้หรือรายการข้อมูลที่เป็นวันที่จะต้องมีจำนวนวันไม่เกิน 31 และจำนวนเดือนไม่เกิน 12 เป็นต้น

2) ความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Relative Reliability : R_r)

เป็นค่าความน่าเชื่อถือที่ประเมินได้จากการเปรียบเทียบข้อมูลว่า เป็นจริงตามความสัมพันธ์และเงื่อนไขของระบบฐานข้อมูลที่ได้กำหนดไว้หรือไม่

R_r เป็นค่าความน่าเชื่อถือที่ได้จากการเปรียบเทียบข้อมูลกับข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล เพราะฉะนั้นในการประเมินความน่าเชื่อถือขั้นตอนนี้ จะต้องประเมินว่าข้อมูลทั้งหมดที่มีอยู่นั้น เป็นไปตามข้อกำหนดหรือเงื่อนไขที่มีอยู่หรือไม่ ตัวอย่างเช่น ถ้าในฐานข้อมูลทะเบียนนักศึกษา ได้มีการกำหนดเงื่อนไขว่าจะต้องไม่มีข้อมูลของนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองเทอมติดต่อกัน ฉะนั้นถ้าฐานข้อมูลมีความน่าเชื่อถือลักษณะนี้ ก็แสดงว่ารายการข้อมูลของนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองเทอมติดต่อกันเป็นรายการข้อมูลที่ผิด

3) ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความเป็นจริง (Absolute Reliability : R_a)

เป็นค่าความน่าเชื่อถือของข้อมูลที่ได้จากการเปรียบเทียบข้อมูลที่มีอยู่ในฐานข้อมูลกับข้อมูลตามความเป็นจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

R_3 เป็นค่าความน่าเชื่อถือที่ได้จากการสังเกตเห็น (Observation) หรือกล่าวอีกนัย คือค่าความน่าเชื่อถือที่ได้จากการดูรายการข้อมูลทุก ๆ รายการ และนำไปเปรียบเทียบกับข้อมูลที่มีอยู่จริง ซึ่งจะเห็นได้ว่าค่า R_3 หาได้ยากมาก แต่เมื่อสามารถหาได้แล้วจะเป็นเครื่องยืนยันได้อย่างดีว่าระบบฐานข้อมูลหรือระบบสารสนเทศนั้น ๆ มีความน่าเชื่อถือได้มากน้อยเพียงใด ตัวอย่างเช่น รายการข้อมูลรหัสและชื่อนักศึกษา ในฐานข้อมูลนักศึกษา จะต้องตรงกับรหัสและชื่อนักศึกษาที่มีอยู่จริง

ในบทต่อไปจะเป็นกล่าวถึงวิธีการประเมินค่าความน่าเชื่อถือของข้อมูลแต่ลักษณะที่กล่าวมาแล้วโดยการสร้างเป็นระบบผู้เชี่ยวชาญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลในแต่ละรายการข้อมูล

(Expert System for Internal Reliability)

ในบทนี้เป็นการกล่าวถึงวิธีการประเมินหาความน่าเชื่อถือของข้อมูลที่เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการ ซึ่งสามารถทำได้โดยการสร้างระบบผู้เชี่ยวชาญที่ประกอบไปด้วยฐานความรู้(knowledge base) โดยในฐานความรู้นี้จะประกอบไปด้วยเงื่อนไขหรือกฎ(rule) ต่าง ๆ ซึ่งเก็บรวบรวมไว้ด้วยกันในส่วนที่เรียกว่าฐานของกฎเกณฑ์(rule base) โดยจะเอาไว้ใช้ตรวจสอบความผิดพลาดข้อมูลในแต่ละรายการข้อมูลที่มีอยู่จริงในระบบฐานข้อมูล ผลลัพธ์ที่ได้จะเป็นจำนวนของรายการข้อมูลที่ผิดตามเงื่อนไขที่กำหนดไว้ในฐานของกฎเกณฑ์ หลังจากนั้นจึงนำเอาจำนวนของรายการข้อมูลที่ผิดตามเงื่อนไขที่กำหนดไว้นั้นไปคำนวนหาค่าเปอร์เซ็นต์ความน่าเชื่อถือของข้อมูลต่อไป โดยสามารถสรุปเป็นขั้นตอนได้ดังนี้

4.1 ส่วนประกอบของระบบผู้เชี่ยวชาญลักษณะ R_i

4.1.1 การสร้างฐานความรู้ในลักษณะของตารางกฎเกณฑ์ (rule base table)

เนื่องจากการวิจัยนี้เป็นการพยายามที่จะสร้างระบบผู้เชี่ยวชาญเพื่อประเมินหาความน่าเชื่อถือของข้อมูล โดยที่ระบบผู้เชี่ยวชาญที่สร้างขึ้นนี้จะไม่ยึดติดกับฐานข้อมูลใดฐานข้อมูลหนึ่ง เพราะฉะนั้นฐานของกฎเกณฑ์ที่ใช้ในระบบผู้เชี่ยวชาญนี้จะต้องมีลักษณะที่สามารถทำการแก้ไขปรับปรุง หรือทำการเปลี่ยนแปลงในกรณีที่มีการเปลี่ยนแปลงระบบโครงสร้างของฐานข้อมูลได้ จากการศึกษาพบว่าเราสามารถที่จะสร้างฐานของกฎเกณฑ์ให้อยู่ในลักษณะของตารางข้อมูลได้ โดยให้รายการข้อมูลแต่ละรายการในตารางข้อมูลที่สร้างขึ้นนั้นแทนเงื่อนไขแต่ละเงื่อนไขที่ใช้ตรวจสอบความผิดพลาดของข้อมูล โดยจะเรียกตารางข้อมูลที่สร้างขึ้นนี้ว่า ตารางกฎเกณฑ์(rule base table)

ก่อนที่จะเก็บเงื่อนไขไว้ในตารางกฎเกณฑ์นั้นจะต้องแปลงเงื่อนไขให้อยู่ในรูปของภาษาที่ใช้ในการสอบถามข้อมูลที่เป็นโครงสร้างหรือที่เราเรียกว่าภาษา SQL (Structured Query Language) เสียก่อนจึงจะทำการเก็บเงื่อนไขที่อยู่ในลักษณะของภาษา SQL นี้ลงไปเก็บไว้ในตารางกฎเกณฑ์ โดยในตารางกฎเกณฑ์ที่สร้างขึ้นมานี้จะประกอบด้วยฟิลด์ข้อมูลที่สามารถเก็บลักษณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของข้อมูลที่เป็นคำสั่งภาษา SQL พื้นฐาน คือ SELECT , FROM , WHERE ได้ สาเหตุที่ใช้คำสั่งภาษา SQL เพียง 3 คำสั่งเท่านั้นเป็นเพราะเงื่อนไขที่ใช้ในการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i นั้นจะเป็นเงื่อนไขที่เกิดขึ้นในแต่ละฟิลด์แต่ละตารางข้อมูลเท่านั้นเพราะฉะนั้นจึงไม่จำเป็นต้องใช้คำสั่งภาษา SQL ที่นอกเหนือจากนี้ โดยที่ในตารางกฎเกณฑ์ที่สร้างขึ้นนั้นต้องประกอบไปด้วยฟิลด์ข้อมูลดังต่อไปนี้

RuleNumber เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นหมายเลขเงื่อนไขแต่ละเงื่อนไข

TableName เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของตารางข้อมูลในระบบฐานข้อมูลที่ปรากฏอยู่ในเงื่อนไขแต่ละเงื่อนไข ซึ่งก็คือส่วนของคำสั่ง FROM ในภาษา SQL โดยชื่อตารางข้อมูลที่เก็บอยู่นี้จะต้องมีอยู่จริงในฐานข้อมูล

Condition เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นส่วน condition ของเงื่อนไข ซึ่งก็คือส่วนของคำสั่ง WHERE ในภาษา SQL โดยที่ในฟิลด์ข้อมูลนี้จะประกอบไปด้วยข้อมูลอยู่ 3 ส่วน คือ

1. Field Name เป็นข้อมูลของชื่อฟิลด์ข้อมูลที่ปรากฏอยู่ในเงื่อนไข โดยชื่อฟิลด์ข้อมูลนี้จะต้องเป็นชื่อฟิลด์ข้อมูลที่มีอยู่จริงในตารางข้อมูลที่ชื่อปรากฏอยู่ในฟิลด์ TableName ในเงื่อนไขเดียวกัน

2. Condition Operators เป็นข้อมูลของสัญญลักษณ์ของการเปรียบเทียบเงื่อนไข รวมถึงสัญญลักษณ์ทางตรรกะ โดยมีสัญญลักษณ์และความหมายดังตารางที่ 1

3. Value เป็นข้อมูลของค่าที่ใช้เปรียบเทียบในเงื่อนไข โดย value นี้จะประกอบไปด้วยข้อมูล 3 ลักษณะคือ

3.1 ข้อมูลชนิดตัวอักษร (Character) ในการใช้ค่าที่เป็นชนิดตัวอักษรจะต้องมีการใช้ภายใต้เครื่องหมาย " " เท่านั้น

3.2 ข้อมูลชนิดตัวเลข (Numeric) ในการใช้ค่าที่เป็นชนิดตัวเลขในการใช้ไม่ต้องอยู่ภายใต้เครื่องหมายใด ๆ

3.3 ข้อมูลชนิดวันที่และเวลา (Date&Time) ในการใช้ค่าที่เป็นชนิดวันที่และเวลาในการใช้ต้องอยู่ภายใต้เครื่องหมาย # # และในการกำหนดรูปแบบของข้อมูลชนิดวันที่และเวลามีดังนี้

Gerneral Date	มีรูปแบบเป็น 19/06/1995 5:34:23 PM.
Long Date	มีรูปแบบเป็น Sunday, 19 June, 1995
Medium Date	มีรูปแบบเป็น 19-Jun-95
Short Date	มีรูปแบบเป็น 19/06/1995
Long Time	มีรูปแบบเป็น 5:34:23 PM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Medium Time มีรูปแบบเป็น 5:34 PM

Short Time มีรูปแบบเป็น 17:34

ตารางที่ 1

แสดงสัญญลักษณ์ ชื่อเรียก และความหมายของเครื่องหมายเงื่อนไขและเครื่องหมายทางตรรกะ

สัญญลักษณ์	ชื่อเรียก	ความหมาย
=	เท่ากับ (Equal to)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่าเท่ากับข้อมูลที่อยู่ทางขวาของเครื่องหมาย
<>	ไม่เท่ากับ (Not equal to)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่าไม่เท่ากับข้อมูลที่อยู่ทางขวาของเครื่องหมาย
>	มากกว่า (Greater than)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่ามากกว่าข้อมูลที่อยู่ทางขวาของเครื่องหมาย
<	น้อยกว่า (Less than)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่าน้อยกว่าข้อมูลที่อยู่ทางขวาของเครื่องหมาย
>=	มากกว่าหรือเท่ากับ (Greater than or equal to)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่ามากกว่าหรือเท่ากับข้อมูลที่อยู่ทางขวาของเครื่องหมาย
<=	น้อยกว่าหรือเท่ากับ (Less than or equal to)	หมายถึงข้อมูลที่อยู่ทางซ้ายของเครื่องหมายจะต้องมีค่าน้อยกว่าหรือเท่ากับข้อมูลที่อยู่ทางขวาของเครื่องหมาย
AND	และ (and)	ใช้เชื่อมเงื่อนไข 2 เงื่อนไขเข้าด้วยกัน จะให้ผลเป็นจริง ถ้าทั้งสองเงื่อนไขเป็นจริง
OR	หรือ (or)	ใช้เชื่อมเงื่อนไข 2 เงื่อนไขเข้าด้วยกัน จะให้ผลเป็นจริง ถ้าเงื่อนไขใดเงื่อนไขหนึ่งเป็นจริง
NOT	ไม่ (not)	ใช้เปลี่ยนค่าทางตรรกะของเงื่อนไขเป็นตรงกันข้าม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเลือกว่าจะใช้ค่าชนิดวันที่และเวลาในรูปแบบใดนั้นขึ้นอยู่กับการกำหนดรูปแบบของฟิลด์ข้อมูลชนิดวันที่และเวลาที่มีการกำหนดไว้แล้วในฐานข้อมูล

และในการเลือกว่าจะใช้ค่าชนิดใดนั้นจะต้องคำนึงว่าค่านั้นจะต้องเป็นข้อมูลชนิดเดียวกับฟิลด์ข้อมูลที่นำมาใช้เปรียบเทียบ เช่นถ้าฟิลด์ข้อมูลที่นำมาสร้างเงื่อนไขเป็นฟิลด์ข้อมูลชนิดตัวอักษร ค่าที่ใช้เปรียบเทียบจะต้องเป็นชนิดตัวอักษรเช่นเดียวกัน โดยสามารถยกตัวอย่างการเขียน condition ได้ดังนี้

condition 1 คือ Name = "สมชาย"

โดยที่ ฟิลด์ Name เป็นฟิลด์ข้อมูลชนิดตัวอักษร

condition 2 คือ Salary > 6000

โดยที่ ฟิลด์ salary เป็นฟิลด์ข้อมูลชนิดตัวเลข

condition 3 คือ Born Date <> #15/05/1995#

โดยที่ ฟิลด์ Born Date เป็นฟิลด์ข้อมูลชนิดวันที่ที่รูปแบบเป็น Short Date

4.1.2 การสร้างเครื่องอนุมาน (inference engine)

ในระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i นี้จะเป็นการอนุมานแบบเดินหน้า (forward chaining inference) ที่สร้างด้วยภาษา Visual Basic เพราะภาษา Visual Basic สามารถที่จะเชื่อมโยงเข้าถึงกับระบบฐานข้อมูลที่สร้างด้วย Microsoft Access version 2.0 ได้ โดยในการทำงานของเครื่องอนุมานนั้นจะทำงานโดยอาศัยหมายเลขของเงื่อนไขที่เก็บอยู่ในฟิลด์ Rule Number ในตารางฐานของกฎเกณฑ์เป็นหลัก โดยจะเลือกเงื่อนไขทุก ๆ เงื่อนไขตามหมายเลขเงื่อนไขจากน้อยไปมาก ที่เป็นเช่นนี้เพราะว่าระบบผู้เชี่ยวชาญที่สร้างขึ้นนี้จะต้องนำเงื่อนไขทุก ๆ เงื่อนไขที่เก็บอยู่ในตารางกฎเกณฑ์ไปตรวจสอบหาข้อมูลที่ผิดพลาดจึงจะสามารถประเมินค่าความน่าเชื่อถือของข้อมูลออกมาได้

4.1.3 ส่วนดึงความรู้ (knowledge acquisition facility)

ในส่วนดึงความรู้นี้จะสร้างด้วยภาษา Visual Basic เช่นเดียวกับเครื่องอนุมาน โดยจะเป็นส่วนที่ไปดึงเอาเงื่อนไขต่าง ๆ ที่เก็บไว้ในตารางกฎเกณฑ์มาแปลงเป็นคำสั่งภาษา SQL เพื่อนำไปค้นหารายการข้อมูลที่ผิดพลาดในระบบฐานข้อมูลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 ส่วนติดต่อกับผู้ใช้ (user interface)

ในส่วนนี้จะสร้างโดยใช้ในลักษณะของ Form ในโปรแกรม Microsoft Access โดยในส่วนที่เชื่อมต่อกับผู้ใช้ในการประเมินความน่าเชื่อถือของข้อมูลในลักษณะ R_i นี้จะแบ่งออกเป็น 2 ส่วนคือ

1. ส่วนที่เชื่อมต่อกับผู้เชี่ยวชาญของระบบฐานข้อมูล เพื่อที่จะใช้ในการเพิ่ม แก้ไข หรือลบเงื่อนไขที่เป็นส่วนของเงื่อนไขในฐานความรู้

2. ส่วนที่เชื่อมต่อกับผู้ใช้ระบบผู้เชี่ยวชาญนี้ เพื่อที่จะสั่งให้ระบบทำการประเมินหาค่าความน่าเชื่อถือของข้อมูลตามเงื่อนไขที่สร้างไว้โดยผู้เชี่ยวชาญในส่วนที่ 1

4.2 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญ

4.2.1 ทำการรวบรวมเงื่อนไขของข้อมูลชนิด R_i ทั้งหมดที่มีอยู่ในฐานข้อมูล

หลังจากที่ทำการสร้างตารางกฎเกณฑ์ขึ้นมาแล้ว ขั้นตอนต่อไปก็คือทำการรวบรวมเงื่อนไขหรือข้อกำหนดของข้อมูลที่เป็นชนิด R_i ทั้งหมดที่อยู่ในฐานข้อมูล ตัวอย่างเงื่อนไขเช่น

เงื่อนไขที่ 1 ในตารางข้อมูล CLASS จำนวนนักศึกษาจะต้องมากกว่าหรือเท่ากับ 0

Condition 1 CLASS.CLASS_SIZE > 0

เงื่อนไขที่ 2 ในตารางข้อมูล CLASS รหัสภาคการศึกษาจะต้องมีค่าระหว่าง 01 ถึง 03

Condition 2 CLASS.SEMESTER >= "01" AND CLASS.SEMESTER <= "03"

เงื่อนไขที่ 3 ในตารางข้อมูล STUDENT_BIO1 รหัสเพศจะต้องเป็น "1" หรือ "2" เท่านั้น

Condition 3 STUDENT_BIO1.SEX = "1" OR STUDENT_BIO1.SEX = "2"

เงื่อนไขที่ 4 ในตารางข้อมูล STUDENT_BIO1 เกรดเฉลี่ยสะสมจะมีค่ามากกว่า 0.0 และน้อยกว่าหรือเท่ากับ 4.0

Condition 4 STUDENT_BIO1.AVERAGE_GPA >= 0.0 AND
STUDENT_BIO1.AVERAGE_GPA <= 4.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เงื่อนไขที่ 5 ในตารางข้อมูล COURSE_SECTION จำนวนผู้ลงเรียนได้มากที่สุดจะต้องมากกว่าหรือเท่ากับจำนวนคนน้อยที่สุดที่จะเปิด

Condition 5 COURSE_SECTION.SECTION_SIZE_MAX >= COURSE_SECTION.SECTION_SIZE_MIN

ในการกำหนดเงื่อนไขให้ครอบคลุมทุก ๆ เงื่อนไขในฐานข้อมูลนั้น สามารถทำได้โดยการพิจารณาทีละฟิลด์ข้อมูลในแต่ละตารางข้อมูลว่าข้อมูลที่ปรากฏอยู่ในฟิลด์ข้อมูลเหล่านั้นมีโอกาสที่จะเกิดความผิดพลาดได้หรือไม่ ถ้าเป็นฟิลด์ข้อมูลมีโอกาสที่จะเกิดความผิดพลาดได้ แล้วสามารถเขียนเป็นเงื่อนไขตามลักษณะของข้อมูลที่ปรากฏอยู่ได้ก็จะเก็บรวบรวมเงื่อนไขนี้ไว้ โดยจะทำเช่นนี้กับทุก ๆ ฟิลด์ข้อมูลและทุก ๆ ตารางข้อมูลที่มีอยู่ในฐานข้อมูลที่จะทำการประเมินหาค่าความน่าเชื่อถือ

ตารางที่ 2

แสดง inverse condition operators ของ condition operators แต่ละตัว

condition operators	Inverse condition operators
=	<>
<>	=
>	<=
<	>=
>=	<
<=	>
AND	OR
OR	AND
NOT	ไม่ใช้เครื่องหมาย NOT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 แปลงเงื่อนไขที่ได้จากฐานข้อมูลเป็นเงื่อนไขที่ใช้ทดสอบความผิดพลาด

เนื่องจากเงื่อนไขที่ได้จากขั้นตอนที่ 1 นั้นเป็นเงื่อนไขของข้อมูลตามลักษณะที่ปรากฏอยู่จริง เพราะฉะนั้นจะทำให้เป็นเงื่อนไขที่ใช้ทดสอบความผิดพลาดได้นั้นจะต้องเปลี่ยนแปลงเงื่อนไขเป็นตรงกันข้าม(inverse condition) โดยในการเปลี่ยนแปลงนั้นจะทำการเปลี่ยนที่เครื่องหมายของเงื่อนไข(condition operator) ดังตารางที่ 2

จากเงื่อนไขในขั้นตอนที่ 1 สามารถเปลี่ยนแปลงเป็นเงื่อนไขเพื่อหาความผิดพลาดได้ดังนี้

เงื่อนไขที่ 1

Condition 1 CLASS.CLASS_SIZE > 0

Inverse Condition 1 CLASS.CLASS_SIZE <= 0

เงื่อนไขที่ 2

Condition 2 CLASS.SEMESTER >= "01" AND CLASS.SEMESTER <= "03"

Inverse Condition 2 CLASS.SEMESTER < "01" OR CLASS.SEMESTER > "03"

เงื่อนไขที่ 3

Condition 3 STUDENT_BIO1.SEX = "1" OR STUDENT_BIO1.SEX = "2"

Inverse Condition 3 STUDENT_BIO1.SEX <> "1" AND STUDENT_BIO1.SEX <> "2"

เงื่อนไขที่ 4

Condition 4 STUDENT_BIO1.AVERAGE_GPA >= 0.0 AND
STUDENT_BIO1.AVERAGE_GPA <= 4.0

Inverse Condition 4 STUDENT_BIO1.AVERAGE_GPA < 0.0 OR
STUDENT_BIO1.AVERAGE_GPA > 4.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เงื่อนไขที่ 5

Condition 5	COURSE_SECTION.SECTION_SIZE_MAX >= COURSE_SECTION.SECTION_SIZE_MIN
Inverse Condition 5	COURSE_SECTION.SECTION_SIZE_MAX < COURSE_SECTION.SECTION_SIZE_MIN

4.2.3 การแทนเงื่อนไขที่ใช้ตรวจสอบความผิดพลาดลงไปในตารางกฎเกณฑ์

ขั้นตอนนี้เป็นการนำเอาเงื่อนไขที่ใช้ตรวจสอบความผิดพลาดของข้อมูลที่ได้จากขั้นตอนที่ 2 มาเก็บไว้ในตารางกฎเกณฑ์ แต่เงื่อนไขที่ใช้ตรวจสอบความผิดพลาดที่ได้จากขั้นตอนที่ 2 นั้นเป็นเงื่อนไขที่อยู่ในรูปของภาษาธรรมชาติ เพราะฉะนั้นในการที่นำเอาเงื่อนไขที่ได้เหล่านี้ไปเก็บในตารางกฎเกณฑ์นั้น จะต้องทำการแปลงเงื่อนไขให้อยู่ในรูปของภาษา SQL เสียก่อน แล้วจึงจะทำการแปลงจากภาษา SQL ไปเก็บไว้ในแต่ละฟิลด์ข้อมูลในตารางกฎเกณฑ์อีกทีหนึ่ง ตัวอย่างเช่น

เงื่อนไขที่ 1

Inverse Condition CLASS.CLASS_SIZE <= 0

คำสั่งภาษา SQL

```
SELECT    *
FROM      CLASS
WHERE     CLASS_SIZE <= 0;
```

ตารางที่ 3

แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 1

RuleNumber	TableName	Condition
001	CLASS	CLASS_SIZE <= 0

เงื่อนไขที่ 2

Inverse Condition CLASS.SEMESTER < "01" OR CLASS.SEMESTER > "03"

คำสั่งภาษา SQL

```
SELECT    *
FROM      CLASS
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
WHERE     CLASS.SEMESTER < "01" OR
          CLASS.SEMESTER > "03";
```

ตารางที่ 4

แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 2

RuleNumber	TableName	Condition
002	CLASS	SEMESTER < "01" OR SEMESTER > "03"

เงื่อนไขที่ 3

Inverse Condition STUDENT_BIO1.SEX <> "1" AND STUDENT_BIO1.SEX <> "2"

คำสั่งภาษา SQL

```
SELECT    *
FROM      STUDENT_BIO1
WHERE     SEX <> "1" AND SEX <> "2";
```

ตารางที่ 5

แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 3

RuleNumber	TableName	Condition
001	STUDENT_BIO1	SEX <> "1" AND SEX <> "2"

เงื่อนไขที่ 4

Inverse Condition

```
STUDENT_BIO1.AVERAGE_GPA < 0.0   OR
STUDENT_BIO1.AVERAGE_GPA > 4.0
```

คำสั่งภาษา SQL

```
SELECT    *
FROM      STUDENT_BIO1
WHERE     AVERAGE_GPA < 0.0 OR AVERAGE_GPA > 4.0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 6

แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 4

RuleNumber	TableName	Condition
004	STUDENT_BIO1	AVERAGE_GPA < 0.0 OR AVERAGE_GPA > 4.0

เงื่อนไขที่ 5

Inverse Condition COURSE_SECTION.SECTION_SIZE_MAX <
COURSE_SECTION.SECTION_SIZE_MIN

คำสั่งภาษา SQL

```
SELECT    *
FROM      COURSE_SECTION
WHERE     SECTION_SIZE_MAX < SECTION_SIZE_MIN;
```

ตารางที่ 7

แสดงการเก็บเงื่อนไขในรูปของตารางกฎเกณฑ์ของเงื่อนไขที่ 5

RuleNumber	TableName	Condition
005	COURSE_SECTION	SECTION_SIZE_MAX < SECTION_SIZE_MIN

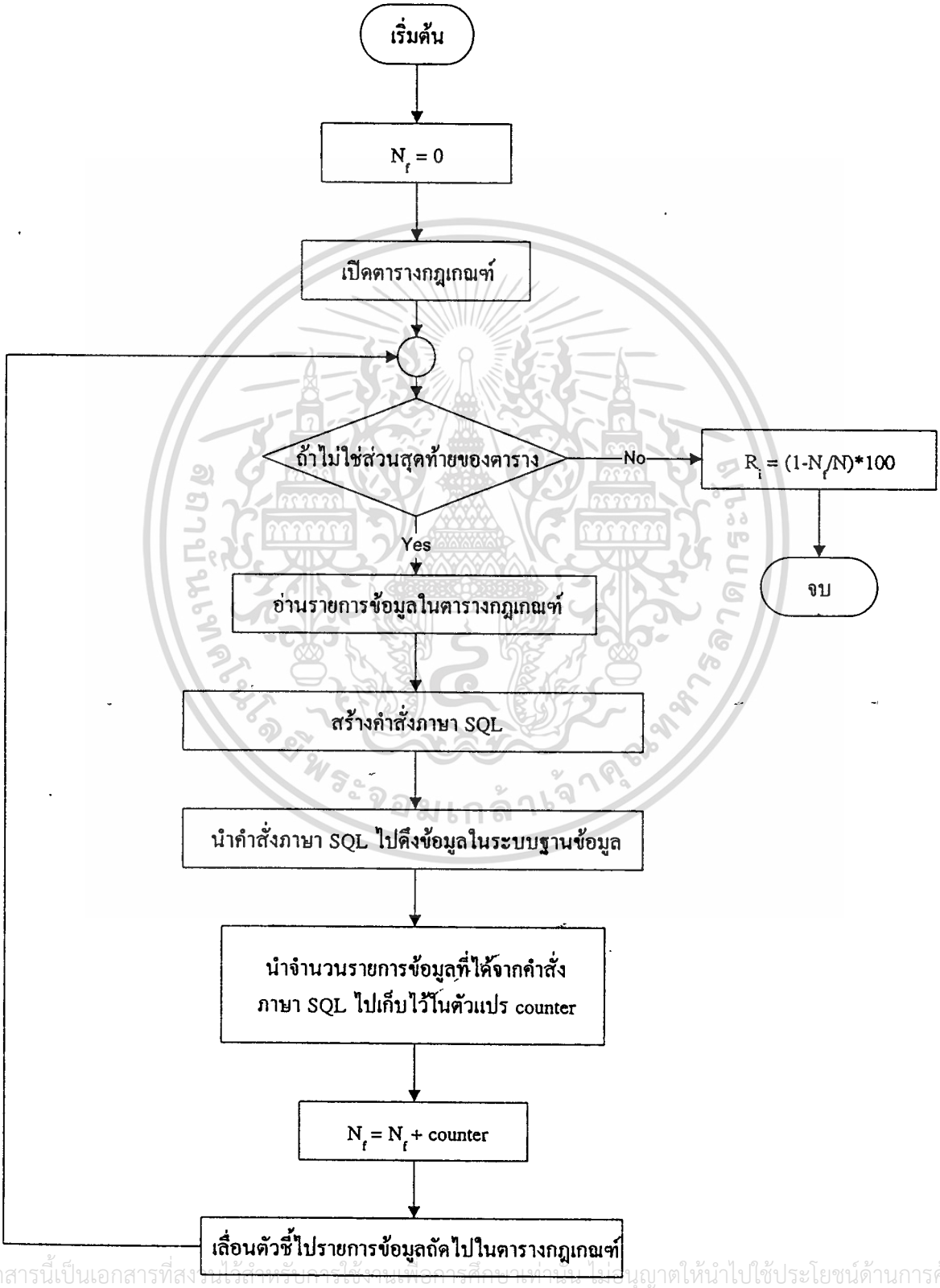
4.2.4 นำเงื่อนไขที่เก็บไว้ในตารางกฎเกณฑ์ไปเปรียบเทียบค้นหาข้อมูลที่ผิดพลาด

เมื่อมีการเก็บเงื่อนไขทั้งหมดลงไปในตารางกฎเกณฑ์แล้ว เมื่อต้องการที่จะประเมินความน่าเชื่อถือของข้อมูลจะทำได้โดยการนำเอาเงื่อนไขแต่ละเงื่อนไขที่อยู่ในลักษณะของรายการข้อมูลแต่ละรายการในตารางกฎเกณฑ์ ไปเปรียบเทียบกับรายการข้อมูลที่มีอยู่จริงในฐานข้อมูล ซึ่งสามารถทำได้ด้วยการสร้างโปรแกรมขึ้นมาช่วยเพื่อที่จะนำเงื่อนไขที่เก็บไว้ในตารางกฎเกณฑ์แปลงเป็นคำสั่งภาษา SQL แล้วจึงนำเอาคำสั่งภาษา SQL ที่ได้นี้ไปค้นหารายการข้อมูลที่มีอยู่ในระบบฐานข้อมูล โดยจำนวนรายการข้อมูลที่ได้โดยคำสั่งภาษา SQL ก็คือจำนวนรายการข้อมูลที่เกิดการผิดพลาดนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 6

แสดงผังการทำงานของระบบการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะทำเช่นนี้กับทุก ๆ เงื่อนไขที่มีเก็บอยู่ในตารางกฎเกณฑ์ และจะเก็บจำนวนรายการข้อมูลทั้งหมดที่ได้จากคำสั่งภาษา SQL ในทุก ๆ เงื่อนไขไว้เพื่อนำไปใช้ในการคำนวณหาค่าความน่าเชื่อถือของข้อมูลในลักษณะ R_i ต่อไป

4.2.5 คำนวณหาค่าความน่าเชื่อถือ

หลังจากนำเอาเงื่อนไขที่เก็บไว้ในตารางกฎเกณฑ์ไปเปรียบเทียบค้นหาข้อมูลที่ผิดพลาดในฐานข้อมูลแล้ว เราจะได้จำนวนของรายการข้อมูลออกมา การคำนวณหาค่าความน่าเชื่อถือนั้นสามารถทำได้โดยการนำเอาจำนวนรายการข้อมูลที่ผิดพลาด(N_f)ที่ได้มานั้นไปคำนวณหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดของข้อมูล ได้จากสมการที่ 3.4

ค่าความน่าเชื่อถือที่คำนวณได้นั้นจะเป็นค่าของเปอร์เซ็นต์ความน่าเชื่อถือ ถ้าค่าเปอร์เซ็นต์ความน่าเชื่อที่คำนวณได้มีค่ามากก็สามารถอนุมานได้ว่าข้อมูลที่นำมาคำนวณหาค่าความน่าเชื่อถือนั้นมีความถูกต้องสูง แต่ในทางกลับกันถ้าค่าเปอร์เซ็นต์ความน่าเชื่อที่คำนวณได้มีค่าน้อยก็สามารถอนุมานได้ว่าข้อมูลที่นำมาคำนวณหาค่าความน่าเชื่อถือนั้นไม่ค่อยมีความถูกต้องเท่าที่ควร

4.3 ผลการทดลอง

ในการทดลองจะกระทำกับฐานข้อมูลนักศึกษา โดยจะแบ่งการทดลองออกเป็น 3 ครั้ง ซึ่งในแต่ละครั้งจะใช้จำนวนรายการข้อมูลที่ไม่เท่ากันเพื่อทำการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลชนิด R_i ด้วยวิธีการประเมินโดยใช้ระบบผู้เชี่ยวชาญที่สร้างขึ้น กับวิธีการประเมินโดยใช้เจ้าหน้าที่ที่เกี่ยวข้องกับฐานข้อมูลนี้ โดยในแต่ละครั้งจะประกอบไปด้วยเงื่อนไขในตารางกฎเกณฑ์ดังตารางที่ 8

4.3.1 ข้อกำหนดในการทดลอง

1. เครื่องคอมพิวเตอร์ที่ใช้เป็นเครื่องคอมพิวเตอร์ระดับบุคคล(personal computer) ซึ่งมีลักษณะรายละเอียดดังนี้

   1.1 มี CPU ขนาด 486 DX2-66

   1.2 หน่วยความจำภายใน(RAM) ขนาด 8 MB.

   1.3 หน่วยความจำ Cache ขนาด 256 KB.

2. ใช้โปรแกรม Microsoft Access version 2.0 เป็นโปรแกรมจัดการฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ใช้บุคลากรที่เกี่ยวข้องกับระบบฐานข้อมูลนี้ 2 ท่าน โดยทั้งสองท่านนี้เป็นเจ้าหน้าที่ของศูนย์คอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ และเป็นผู้ที่ออกแบบและสร้างระบบฐานข้อมูลนักศึกษาที่ใช้เป็นตัวอย่างในการทดลอง

ตารางที่ 8

แสดงตารางกฎเกณฑ์ที่ใช้ในการทดลองการประเมินความน่าเชื่อถือของข้อมูลชนิด R_i

RuleNumber	TableName	Condition
001	CLASS	YEAR > 2538
002	CLASS	MEETING_TIME_BEGIN >= MEETING_TIME_END
003	CLASS	CLASSE_SIZE <= 0
004	CLASS	SEMESTER < "1" OR SEMESTER > "3"
005	COURSE	COURSE_BEGIN_DATE >= COURSE_END_DATE
006	COURSE	COURSE_CRD < 0.0
007	COURSE	COURSE_CRD_LE + COURSE_CRD_LAB <> COURSE_CRD
008	COURSE_SECTION	YEAR > 2538
009	COURSE_SECTION	SEMESTER < "1" OR SEMESTER > "3"
010	COURSE_SECTION	SECTION_SIZE_MAX < SECTION_SIZE_MIN
011	COURSE_SECTION	COURSE_CRD_LE + COURSE_CRD_LAB <> COURSE_CRD
012	DEPARTMENT	DEPT_BEGIN_DATE >= DEPT_END_DATE
013	DIVISION	DIV_BEGIN_DATE >= DIV_END_DATE
014	REGISTRATION	YEAR > 2538
015	REGISTRATION	SEMESTER < "1" OR SEMESTER > "3"
016	REGISTRATION	GRADE <> "A" AND GRADE <> "B" AND GRADE <> "C" AND GRADE <> "D" AND GRADE <> "F" AND GRADE <> "U" AND GRADE <> "I" AND GRADE <> "A" AND
017	STUDENT_BIO1	SEX <> "1" AND SEX <> "2"
018	STUDENT_BIO1	AVERAGE_GPA < 0.0 OR AVERAGE > 4.0
019	STUDENT_BIO2	HIGHT <= 0
020	STUDENT_BIO2	WEIGTH <= 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 ผลการทดลองที่ได้

การทดลองครั้งที่ 1 ใช้รายการข้อมูลทั้งหมด 300 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 3 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่ และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญ จะได้ผลการทดลองดังตารางที่ 9

ตารางที่ 9

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 1

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	300	300
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	3	3
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	3	3
4. ค่าความน่าเชื่อถือของข้อมูล	99%	99%
5. เปอร์เซ็นต์ความถูกต้อง	100%	100%

การทดลองครั้งที่ 2 ใช้รายการข้อมูลทั้งหมด 600 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 10 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญ จะได้ผลการทดลองดังตารางที่ 10

การทดลองครั้งที่ 3 ใช้รายการข้อมูลทั้งหมด 1000 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 13 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญจะได้ผลการทดลองดังตารางที่ 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 10

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 2

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	600	600
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	10	10
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	9	10
4. ค่าความน่าเชื่อถือของข้อมูล	98.5%	98.3%
5. เปอร์เซ็นต์ความถูกต้อง	99.80%	100%

ตารางที่ 11

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 3

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	1000	1000
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	13	13
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	11	13
4. ค่าความน่าเชื่อถือของข้อมูล	98.9%	98.7%
5. เปอร์เซ็นต์ความถูกต้อง	99.80%	100%

4.3.3 สรุปผลการทดลอง

จากผลการทดลองสรุปได้ว่า การประเมินความน่าเชื่อถือของข้อมูลในลักษณะ R_i โดยการใช้ระบบผู้เชี่ยวชาญนี้ สามารถที่จะประเมินหาค่าความน่าเชื่อถือของข้อมูลได้ถูกต้องกว่าการประเมินความน่าเชื่อถือของข้อมูลโดยใช้เจ้าหน้าที่ และจะใช้เวลาในการประเมินที่น้อยกว่าวิธีที่ใช้เจ้าหน้าที่มาก ซึ่งจะเห็นได้ว่ากรณีที่ระบบฐานข้อมูลมีขนาดใหญ่และมีจำนวนรายการข้อมูลมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประเมินความน่าเชื่อถือของข้อมูลโดยการใช้เจ้าหน้าที่ไม่สามารถที่จะทำได้เลยเพราะจะไม่คุ้มกับเวลาที่เสียไป ในกรณีเช่นนี้จึงเหมาะที่จะใช้การประเมินความน่าเชื่อถือของข้อมูลโดยระบบผู้เชี่ยวชาญ แต่ทั้งนี้ทั้งนั้นการประเมินค่าความน่าเชื่อถือของข้อมูลโดยใช้ระบบผู้เชี่ยวชาญจะให้ค่าความถูกต้องสมบูรณ์หรือไม่จะต้องขึ้นอยู่กับการกำหนดเงื่อนไขของระบบฐานข้อมูลที่มีอยู่ทั้งหมดให้กับระบบผู้เชี่ยวชาญโดยครบถ้วน มิฉะนั้นจะทำให้ค่าความน่าเชื่อถือของข้อมูลที่ประเมินได้นั้นมีความคลาดเคลื่อนไปจากความเป็นจริงได้

ในกรณีที่มีการเปลี่ยนแปลงเกี่ยวกับโครงสร้างของระบบฐานข้อมูลในส่วนที่กระทบกระเทือนต่อเงื่อนไขที่ใช้ในระบบผู้เชี่ยวชาญนั้น จะต้องทำการปรับปรุงในส่วนของเงื่อนไขในระบบผู้เชี่ยวชาญให้มีความถูกต้องสมบูรณ์อยู่เสมอจึงจะทำให้ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลให้ค่าที่ถูกต้องตามความเป็นจริงมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

ระบบผู้เชี่ยวชาญสำหรับความน่าเชื่อถือของข้อมูล ที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล

(Expert system for Relative reliability)

ในการประเมินความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูลนั้น สามารถทำได้โดยการสร้างระบบผู้เชี่ยวชาญที่ประกอบไปด้วยฐานความรู้ โดยในฐานความรู้นี้จะประกอบไปด้วยเงื่อนไขต่าง ๆ ที่ใช้ตรวจสอบความผิดพลาดข้อมูลในแต่ละรายการข้อมูลที่มีอยู่จริงในระบบฐานข้อมูลที่ต้องการจะทำการประเมินหาค่าความเชื่อถือ หรือกล่าวอีกนัยหนึ่งก็คือพยายามสร้างฐานของกฎเกณฑ์ให้เป็นไปตามเงื่อนไขหรือข้อกำหนดของฐานข้อมูลที่มีอยู่และนำไปตรวจสอบกับรายการข้อมูลที่มีอยู่จริงในฐานข้อมูลนั้นเอง โดยผลที่ได้จะเป็นจำนวนของรายการข้อมูลที่ผิดไปจากเงื่อนไขที่กำหนดไว้ในกฎเกณฑ์ หลังจากนั้นจึงนำเอาจำนวนของรายการข้อมูลที่ผิดไปจากเงื่อนไขที่กำหนดไว้นั้นไปคำนวนหาค่าเปอร์เซ็นต์ความน่าเชื่อถือของข้อมูลต่อไป โดยสามารถแยกออกเป็น 2 วิธี คือ การสร้างฐานของกฎเกณฑ์ในลักษณะของตารางข้อมูล และการสร้างฐานของกฎเกณฑ์ในลักษณะของโปรแกรม โดยในวิธีการสร้างฐานของกฎเกณฑ์ในลักษณะของตารางข้อมูลนั้นจะทำได้โดยการนำเงื่อนไขหรือกฎไปเก็บไว้ในลักษณะของรายการข้อมูลเช่นเดียวกับวิธีการประเมินความน่าเชื่อถือของข้อมูลชนิด R_i ส่วนวิธีการสร้างฐานของกฎเกณฑ์ในลักษณะของโปรแกรมนั้นจะกระทำกับส่วนของเงื่อนไขที่มีความซับซ้อนจนไม่สามารถที่กำหนดไว้ในลักษณะของตารางข้อมูลได้เท่านั้น ซึ่งรายละเอียดทั้ง 2 วิธีมีดังนี้

5.1 ส่วนประกอบของระบบผู้เชี่ยวชาญลักษณะ R_r

5.1.1 การสร้างฐานความรู้ในลักษณะของตารางกฎเกณฑ์ (rule base table)

ในเนื้อหาบทนี้เป็นการกล่าวถึงการประเมินความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล เพราะฉะนั้นในการที่จะประเมินหาข้อมูลที่ผิดพลาดได้นั้น จะต้องทำการสร้าง ฐานของกฎเกณฑ์ ขึ้นมาเพื่อตรวจหาความผิดพลาดของข้อมูลในลักษณะนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยฐานของกฎเกณฑ์ที่สร้างขึ้นนั้นจะต้องครอบคลุมเงื่อนไขทั้งหมดที่ใช้ตรวจสอบความผิดพลาดของข้อมูลในแต่ละฟิลด์ข้อมูล แต่ละรายการข้อมูล และแต่ละตารางข้อมูล

เนื่องจากการวิจัยนี้เป็นการพยายามที่จะสร้างระบบผู้เชี่ยวชาญเพื่อประเมินหาความน่าเชื่อถือของข้อมูล โดยที่ระบบผู้เชี่ยวชาญที่สร้างขึ้นนี้จะไม่ยึดติดกับฐานข้อมูลใดฐานข้อมูลหนึ่ง เพราะฉะนั้นฐานของกฎเกณฑ์ที่ใช้ในระบบผู้เชี่ยวชาญนี้จะต้องมีลักษณะที่สามารถทำการแก้ไขปรับปรุง หรือทำการเปลี่ยนแปลงในกรณีที่มีการเปลี่ยนแปลงโครงสร้างของระบบฐานข้อมูลได้ จากการศึกษาพบว่า เงื่อนไขที่ประกอบเป็นฐานของกฎเกณฑ์นั้นโดยส่วนใหญ่สามารถที่จะสร้างให้อยู่ในลักษณะของตารางข้อมูลได้ โดยให้รายการข้อมูลแต่ละรายการ(หรือมากหนึ่งรายการข้อมูล)ในตารางกฎเกณฑ์ที่สร้างขึ้นนั้นจะแทนเงื่อนไขที่ใช้ตรวจสอบความผิดพลาดของข้อมูลแต่ละเงื่อนไข เช่นเดียวกับตารางกฎเกณฑ์ใน R_i แต่ต่างกันที่การประเมินความน่าเชื่อของข้อมูลแบบ R_f นั้นเงื่อนไขในตารางกฎเกณฑ์แต่ละเงื่อนไขอาจจะเกิดขึ้นจากความสัมพันธ์กันระหว่างตารางข้อมูลหลาย ๆ ตารางข้อมูล เพราะฉะนั้นตารางกฎเกณฑ์ที่สร้างขึ้นนั้นจะต้องเก็บรายละเอียดของเงื่อนไขทั้งหมดที่เกิดขึ้นได้

ก่อนที่จะเก็บเงื่อนไขไว้ในตารางกฎเกณฑ์นั้นจะต้องแปลงเงื่อนไขในอยู่ในรูปของภาษาที่ใช้ในการสอบถามข้อมูลที่เป็นโครงสร้างหรือที่เราเรียกว่าภาษา SQL เสียก่อนจึงจะทำการเก็บเงื่อนไขที่อยู่ในลักษณะของภาษา SQL นี้ลงไปเก็บไว้ในตารางกฎเกณฑ์ โดยในตารางกฎเกณฑ์ที่สร้างขึ้นมานี้จะประกอบด้วยฟิลด์ข้อมูลที่สามารถเก็บลักษณะของข้อมูลที่เป็นคำสั่งภาษา SQL ได้ ตัวอย่างเช่นคำสั่ง SELECT , FROM , WHERE , GROUP BY , HAVING เป็นต้น

ตารางกฎเกณฑ์ที่สร้างขึ้นมานี้จะประกอบด้วยตารางข้อมูล 2 ตารางที่จะค่อยทำหน้าที่ร่วมกัน คือ ตารางข้อมูลในส่วนที่ใช้เก็บความสัมพันธ์ต่าง ๆ ของเงื่อนไขซึ่งจะเรียกว่า ตารางความสัมพันธ์(relation table) ส่วนอีกตารางหนึ่งคือตารางข้อมูลที่ใช้เก็บค่า condition ของเงื่อนไขซึ่งจะเรียกว่า ตารางเงื่อนไข(condition table) โดยตารางข้อมูลทั้ง 2 จะมีรายละเอียดดังต่อไปนี้

1. ตารางความสัมพันธ์ (relation table)

เนื่องจากเงื่อนไขที่ใช้ตรวจสอบหาข้อมูลที่ผิดพลาดใน R_f นั้นอาจเกิดขึ้นจากความสัมพันธ์ระหว่างรายการข้อมูลจากตารางข้อมูลหลาย ๆ ตารางข้อมูลก็ได้ เพราะฉะนั้นในการสร้างตารางกฎเกณฑ์จะต้องสร้างตารางข้อมูลที่ใช้เก็บรายละเอียดความสัมพันธ์ระหว่างตารางข้อมูลทั้งหมดในระบบฐานข้อมูล โดยเราจะเรียกตารางข้อมูลที่เก็บรายละเอียดของความสัมพันธ์นี้ว่า ตารางความสัมพันธ์ ซึ่งในตารางความสัมพันธ์จะเป็นที่เก็บชื่อของฟิลด์ข้อมูลที่ทำหน้าที่เชื่อมโยงความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัมพันธ์ระหว่างตารางข้อมูลต่าง ๆ ที่สัมพันธ์กันซึ่งเราเรียกว่า foreign key โดยที่ตารางความสัมพันธ์จะประกอบไปด้วยฟิลด์ข้อมูลดังต่อไปนี้

1. Relation Number เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นหมายเลขความสัมพันธ์แต่ละความสัมพันธ์

2. Primary Table Name เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของตารางข้อมูลที่เป็นต้นทางของความสัมพันธ์

3. Secondary Table Name เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของตารางข้อมูลที่เป็นปลายทางของความสัมพันธ์

4. Primary Foreign Key เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของฟิลด์ข้อมูลที่ทำหน้าที่เป็นฟิลด์เชื่อมโยง(foreign key) ในตารางข้อมูลที่เป็นต้นทางของความสัมพันธ์

5. Secondary Foreign Key เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของฟิลด์ข้อมูลที่ทำหน้าที่เป็นฟิลด์เชื่อมโยงในตารางข้อมูลที่เป็นปลายทางของความสัมพันธ์

ตารางที่ 12
แสดงตัวอย่างตารางความสัมพันธ์

Relation Number	Primary Table Name	Secondary Table Name	Primary Foreign Key	Secondary Foreign Key
001	STUDENT_BIO1	REGISTRATION	STU_CODE	STU_CODE
002	STUDENT_BIO1	FACULTY	FAC_CODE	FAC_CODE
003	REGISTRATION	CUMULATIVE	STU_CODE	STU_CODE
004	FACULTY	COURSE	FAC_CODE	FAC_CODE

2. ตารางเงื่อนไข (Condition Table)

ในการสร้างตารางกฎเกณฑ์นั้น หลังจากที่เราสร้างตารางความสัมพันธ์ขึ้นมาแล้ว เราจะต้องสร้างตารางข้อมูลขึ้นมาอีกหนึ่งตาราง เพื่อใช้เก็บข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูลดังที่กล่าวมาแล้วข้างต้น โดยที่ตารางเงื่อนไขนี้จะประกอบไปด้วยฟิลด์ข้อมูลที่จำเป็นดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Rule Number เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นหมายเลขเงื่อนไขแต่ละเงื่อนไข

2. Primary Table Name เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของตารางข้อมูลต้นทางที่มีความสัมพันธ์กันอยู่ในเงื่อนไขแต่ละเงื่อนไข ซึ่งก็คือส่วนหนึ่งของคำสั่ง FROM ในภาษา SQL โดยชื่อตารางข้อมูลนี้จะต้องมีอยู่จริงในฐานข้อมูล

3. Secondary Table Name เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อของตารางปลายทางที่มีความสัมพันธ์กันอยู่ในเงื่อนไขแต่ละเงื่อนไข ซึ่งก็คืออีกส่วนหนึ่งของคำสั่ง FROM ในภาษา SQL ด้วยเช่นกัน โดยชื่อตารางข้อมูลนี้จะต้องมีอยู่จริงในฐานข้อมูลด้วยเช่นเดียวกับฟิลด์ Primary Table Name แต่ในกรณีที่เงื่อนไขได้ระบุถึงตารางข้อมูลเพียงตารางเดียว เราใส่ค่า "null" ลงในฟิลด์ข้อมูลนี้แทนชื่อตารางข้อมูล

4. Relation Number เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นหมายเลขความสัมพันธ์ เพื่อใช้เชื่อมโยงกับกับตารางข้อมูลความสัมพันธ์

5. Condition เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นส่วน condition ของเงื่อนไข ซึ่งก็คือส่วนของคำสั่ง WHERE ในภาษา SQL นั่นเอง โดยในฟิลด์ข้อมูลนี้จะประกอบไปด้วยข้อมูลอยู่ 3 ส่วน คือ

5.1 Field Name เป็นข้อมูลของชื่อฟิลด์ข้อมูลที่ปรากฏอยู่ในเงื่อนไข โดยชื่อฟิลด์ข้อมูลนี้จะต้องเป็นชื่อฟิลด์ข้อมูลที่มีอยู่จริงในตารางข้อมูลที่ชื่อปรากฏอยู่ในฟิลด์ Primary Table Name หรือ Secondary Table Name ในเงื่อนไขเดียวกัน และในกรณีที่ชื่อของฟิลด์ข้อมูลเหมือนกันอยู่ในตารางข้อมูลมากกว่าหนึ่งตารางที่สัมพันธ์กันและอยู่ในเงื่อนไขเดียวกัน การเขียนเงื่อนไขเพื่ออ้างอิงถึงชื่อฟิลด์ข้อมูลนั้นจะทำได้โดยการใช้ชื่อตารางข้อมูลและขั้นด้วยเครื่องหมาย "." จากนั้นจึงตามด้วยชื่อฟิลด์ข้อมูล ตัวอย่างเช่น การอ้างถึงข้อมูลที่เป็นชื่อนักศึกษาในตารางนักศึกษาที่ 1 และตารางนักศึกษาที่ 2 สามารถเงื่อนไขอ้างอิงเป็น Student1.Name และ Student2.Name เป็นต้น

5.2 Condition Operators เป็นข้อมูลของสัญญลักษณ์ของการเปรียบเทียบเงื่อนไข โดยมีสัญญลักษณ์ดังตารางที่ 1

5.3 Value เป็นข้อมูลของค่าที่ใช้เปรียบเทียบในเงื่อนไข โดยค่านี้จะประกอบไปด้วยข้อมูลลักษณะต่าง ๆ ดังนี้คือ

5.3.1 ข้อมูลชนิดตัวอักษร (Character) ในการใช้ค่าที่เป็นชนิดตัวอักษรจะต้องมีการใช้ภายใต้เครื่องหมาย " " เท่านั้น

5.3.2 ข้อมูลชนิดตัวเลข (Numeric) ในการใช้ค่าที่เป็นชนิดตัวเลขในการใช้ไม่ต้องอยู่ภายใต้เครื่องหมายใด ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.3 ข้อมูลชนิดวันที่และเวลา (Date&Time) ในการใช้ค่าที่เป็นชนิดวันที่และเวลาในการใช้ต้องอยู่ภายใต้เครื่องหมาย # # เช่นเดียวกับวิธีการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i

ในการเลือกว่าจะใช้ค่าชนิดวันที่และเวลาในรูปแบบใดขึ้นอยู่กับการกำหนดรูปแบบของฟิลด์ข้อมูลชนิดวันที่และเวลาที่มีการกำหนดไว้แล้วในฐานข้อมูล

และในการเลือกว่าจะใช้ค่าชนิดใดนั้นจะต้องคำนึงว่าค่าจะต้องเป็นข้อมูลชนิดเดียวกับฟิลด์ข้อมูลที่นำมาใช้เปรียบเทียบ เช่นเดียวกับการสร้างเงื่อนไขในการประเมินค่าความน่าเชื่อชนิด R_i ในบทที่ผ่านมา

6. Group By เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นชื่อฟิลด์ของข้อมูลที่อยู่ในตารางข้อมูลที่มีชื่อปรากฏอยู่ในฟิลด์ Primary table name หรือ Secondary table name โดยข้อมูลที่เป็นชื่อฟิลด์ที่เก็บอยู่ในฟิลด์ Group By นั้นจะต้องเป็นชื่อฟิลด์ข้อมูลที่ต้องการแบ่งกลุ่มหรือจัดหมวดหมู่ให้ข้อมูลที่เหมือนกันไปอยู่ในกลุ่มเดียวกัน ตัวอย่างเช่น ถ้ามีเงื่อนไขที่ต้องการแบ่งกลุ่มข้อมูลนักศึกษาออกเป็นกลุ่มตามเพศ ฉะนั้นชื่อฟิลด์ข้อมูลที่จะเก็บในฟิลด์ Group By จะเป็นฟิลด์ข้อมูล Sex เป็นต้น ในกรณีที่เงื่อนไขที่สร้างขึ้นนั้นไม่มีการแบ่งกลุ่มของข้อมูลเราจะใส่ค่า "null" ลงไปในฟิลด์ Group By นี้แทน

7. Having เป็น ฟิลด์ข้อมูลที่ใช้เก็บข้อมูลที่เป็นเงื่อนไข(condition)ของกระทำกับกลุ่มของข้อมูลที่แบ่งตามฟิลด์ข้อมูลที่มีชื่อปรากฏในฟิลด์ Group By โดยการกระทำนั้นได้แก่ การรวมข้อมูล(Sum) , การนับรายการข้อมูล(Count) , การหาค่าน้อยที่สุด(Min) , การหาค่ามากที่สุด(Max) เป็นต้น โดยการเขียนเงื่อนไขนั้นจะใช้ลักษณะการเขียนและเครื่องหมายเงื่อนไข(condition operator) เช่นเดียวกับการเขียนเงื่อนไขในฟิลด์ Condition แต่มีข้อแม้ว่าถ้าข้อมูลที่อยู่ฟิลด์ Group By เป็น "null" แล้วข้อมูลในฟิลด์ Having จะต้องเป็น "null" ด้วย

8. Connect Condition เป็นฟิลด์ข้อมูลที่ใช้เก็บข้อมูลเงื่อนไขทางตรรกะที่ใช้เชื่อมต่อกับรายการข้อมูลเงื่อนไขถัดไปที่เกี่ยวเนื่องสัมพันธ์กันเช่น AND , OR , NOT เป็นต้น ซึ่งข้อมูลในฟิลด์ Connect Condition นี้จะใช้ในกรณีที่ไม่สามารถที่จะเก็บเงื่อนไขหนึ่งเงื่อนไขไว้ในหนึ่งรายการข้อมูลได้ แต่ในกรณีที่สามารถเก็บเงื่อนไขไว้ในหนึ่งรายการข้อมูลได้นั้นเราใส่ค่า "null" ลงไปเก็บไว้ในฟิลด์ Connect Condition นี้แทน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.2 การสร้างฐานความรู้ในลักษณะของโปรแกรม

จากการวิจัยพบว่า ในบางครั้งเงื่อนไขที่ใช้สร้างเป็นฐานของกฎเกณฑ์ของระบบผู้เชี่ยวชาญนั้นก็มีความซับซ้อนจนไม่สามารถที่จะกำหนดให้อยู่ลักษณะของตารางข้อมูลได้ ตัวอย่างเช่น ถ้ามีเงื่อนไขของระบบฐานข้อมูลนักศึกษาว่า "จะต้องไม่มีนักศึกษาที่มีเกรดเฉลี่ยสะสมที่ผ่านมา 2 เทอมติดต่อกันน้อยกว่า 2.00 ลงทะเบียนเรียนในเทอมนี้" จะเห็นว่าเงื่อนไขนี้ไม่สามารถที่จะเขียนในอยู่ในลักษณะของรายการข้อมูลเงื่อนไขดังวิธีที่กล่าวมาข้างต้นได้ เพราะฉะนั้นจึงกำหนดเงื่อนไขนี้ในลักษณะของโปรแกรมควบคุมระบบฐานข้อมูล โดยที่แต่ละเงื่อนไขจะแยกเขียนไว้เป็นโมดุล(module)ของโปรแกรมที่แยกออกจากกัน เพื่อง่ายต่อการแก้ไขในกรณีที่เกิดการผิดพลาดในแต่ละเงื่อนไข ซึ่งโปรแกรมที่สร้างขึ้นนี้จะต้องเป็นภาษาโปรแกรมที่สามารถเข้าถึงระบบฐานข้อมูลที่จะทำการประเมินความน่าเชื่อถือได้

5.1.3 การสร้างเครื่องอนุมาน (inference engine)

ในระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_r นี้จะเป็นการอนุมานแบบเดินหน้า (forward chaining inference) ที่สร้างด้วยภาษา Visual Basic เพราะภาษา Visual Basic สามารถที่จะเชื่อมโยงเข้าถึงกับระบบฐานข้อมูลที่สร้างด้วย Microsoft Access version 2.0 ได้ โดยในการทำงานของเครื่องอนุมานนั้นจะทำงานโดยอาศัยหมายเลขของเงื่อนไขที่เก็บอยู่ในฟิลด์ Rule Number ในตารางฐานของกฎเกณฑ์เป็นหลัก โดยจะเลือกเงื่อนไขทุก ๆ เงื่อนไขตามหมายเลขเงื่อนไขจากน้อยไปมาก ที่เป็นเช่นนี้เพราะว่าระบบผู้เชี่ยวชาญที่สร้างขึ้นนี้จะต้องนำเงื่อนไขทุก ๆ เงื่อนไขที่เก็บอยู่ในตารางกฎเกณฑ์ไปตรวจสอบหาข้อมูลที่ผิดพลาดจึงจะสามารถประเมินค่าความน่าเชื่อถือของข้อมูลออกมาได้

5.1.4 ส่วนดึงความรู้ (knowledge acquisition facility)

ในส่วนดึงความรู้นี้จะสร้างด้วยภาษา Visual Basic เช่นเดียวกับเครื่องอนุมาน โดยจะเป็นส่วนที่ไปดึงเอาเงื่อนไขต่าง ๆ ที่เก็บไว้ในตารางความสัมพันธ์ และตารางเงื่อนไขมาแปลงเป็นคำสั่งภาษา SQL เพื่อนำไปค้นหารายการข้อมูลที่ผิดพลาดในระบบฐานข้อมูลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.5 ส่วนติดต่อกับผู้ใช้ (user interface)

ในส่วนนี้จะสร้างโดยใช้ในลักษณะของ Form ในโปรแกรม Microsoft Access โดยในส่วนที่เชื่อมต่อกับผู้ใช้ในการประเมินความน่าเชื่อถือของข้อมูลในลักษณะ R_r นี้จะแบ่งออกเป็น 2 ส่วนคือ

1. ส่วนที่เชื่อมต่อกับผู้เชี่ยวชาญของระบบฐานข้อมูล เพื่อที่จะใช้ในการเพิ่ม แก้ไข หรือลบเงื่อนไขที่เป็นส่วนของเงื่อนไขในฐานความรู้

2. ส่วนที่เชื่อมต่อกับผู้ใช้ระบบผู้เชี่ยวชาญนี้ เพื่อที่จะสั่งให้ระบบทำการประเมินหาค่าความน่าเชื่อถือของข้อมูลตามเงื่อนไขที่สร้างไว้โดยผู้เชี่ยวชาญในส่วนที่ 1

5.2 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญลักษณะ R_r

5.2.1 ทำการรวบรวมเงื่อนไขของข้อมูลชนิด R_r ทั้งหมดที่มีอยู่ในฐานข้อมูล

หลังจากที่ทำการสร้างตารางกฎเกณฑ์ขึ้นมาแล้ว ขั้นตอนต่อไปก็คือ ทำการรวบรวมเงื่อนไขหรือข้อกำหนดของข้อมูลที่เป็นชนิด R_r ทั้งหมดที่อยู่ในฐานข้อมูล ตัวอย่างเงื่อนไขเช่น

เงื่อนไข : "นักศึกษาคณะวิทยาศาสตร์ทุกคนจะลงทะเบียนแต่ละเทอมได้ไม่เกิน 24 หน่วยกิต"

ในการรวบรวมเงื่อนไขให้ครอบคลุมทุก ๆ เงื่อนไขในฐานข้อมูลนั้น สามารถทำได้โดยการพิจารณาทีละฟิลด์ข้อมูลในแต่ละตารางข้อมูลที่มีความสัมพันธ์กัน จากนั้นจึงนำไปพิจารณารวมกับเงื่อนไขหรือข้อกำหนดที่มีอยู่ในระบบฐานข้อมูลนั้น ท้ายที่สุดคือพิจารณาอยู่ว่าเงื่อนไขที่มีอยู่นั้นสามารถที่จะกำหนดให้อยู่ในลักษณะของรายการข้อมูลในตารางกฎเกณฑ์ได้หรือไม่

5.2.2 แปลงเงื่อนไขที่ได้จากฐานข้อมูลเป็นเงื่อนไขที่ใช้ทดสอบความผิดพลาดของข้อมูล

เนื่องจากเงื่อนไขที่ได้จากขั้นตอนที่ 1 นั้นเป็นเงื่อนไขของข้อมูลตามลักษณะที่ปรากฏอยู่จริง เพราะฉะนั้นจะทำให้เป็นเงื่อนไขที่ใช้ทดสอบความผิดพลาดของข้อมูลได้นั้นจะต้องเปลี่ยนแปลงเงื่อนไขเป็นตรงกันข้ามเช่นเดียวกับวิธีการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_i โดยในการเปลี่ยนแปลงนั้นจะทำการเปลี่ยนที่เครื่องหมายของเงื่อนไข(condition operators) ดังตารางที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2 ตัวอย่างเช่นจากเงื่อนไขในขั้นตอนที่ 1 สามารถเขียนเป็นเงื่อนไขที่ใช้ทดสอบความผิดพลาดของข้อมูลได้ดังนี้

เงื่อนไข : "ค้นหานักศึกษาคณะวิทยาศาสตร์ที่ลงลงเบียนเรียนในแต่ละเทอมมากกว่า 24 หน่วยกิต"

5.2.3 แทนเงื่อนไขของระบบผู้เชี่ยวชาญให้อยู่ในลักษณะของตารางกฎเกณฑ์

ขั้นตอนนี้เป็นการนำเอาเงื่อนที่ใช้ตรวจสอบความผิดพลาดของข้อมูลที่ได้จากขั้นตอนที่ 2 มาเก็บไว้ในตารางกฎเกณฑ์ แต่เงื่อนไขที่ใช้ตรวจสอบความผิดพลาดที่ได้จากขั้นตอนที่ 2 นั้นเป็นเงื่อนไขที่อยู่ในรูปของภาษาธรรมชาติ เพราะฉะนั้นในการที่นำเอาเงื่อนไขที่ได้เหล่านี้ไปเก็บในตารางกฎเกณฑ์นั้น จะต้องทำการแปลงเงื่อนไขให้อยู่ในรูปของภาษา SQL เสียก่อน แล้วจึงจะทำการแปลงจากภาษา SQL ไปเก็บไว้ในแต่ละฟิลด์ข้อมูลในตารางกฎเกณฑ์อีกทีหนึ่ง โดยจะแยกเก็บเป็น 2 ส่วนคือ ส่วนที่เป็น condition ของเงื่อนไขจะเก็บไว้ในตารางเงื่อนไขและส่วนที่เป็นความสัมพันธ์ระหว่างตารางข้อมูลในเงื่อนไขนั้นจะเก็บไว้ในตารางความสัมพันธ์ ตัวอย่างเช่นเงื่อนไขในขั้นตอนที่ 2 สามารถแทนเป็นตารางกฎเกณฑ์ได้ดังตารางที่ 13 และ ตารางที่ 14

5.2.4 นำเงื่อนไขที่เก็บไว้ในตารางกฎเกณฑ์ไปเปรียบเทียบค้นหาข้อมูลที่ผิดพลาด

เมื่อมีการเก็บเงื่อนไขทั้งหมดลงไปในตารางกฎเกณฑ์แล้ว เมื่อต้องการประเมินความน่าเชื่อถือของข้อมูล จะต้องมีการนำเอาเงื่อนไขแต่ละเงื่อนไขที่อยู่ในลักษณะของรายการข้อมูลแต่ละรายการในตารางกฎเกณฑ์ ซึ่งได้แก่ตารางความสัมพันธ์และตารางเงื่อนไขไปเปรียบเทียบกับรายการข้อมูลที่มีอยู่จริงในฐานข้อมูล ในการประเมินความสามารถทำได้โดยใช้วิธีการนำเอาข้อมูลแต่ละรายการข้อมูลที่เป็นเงื่อนไขจากตารางความสัมพันธ์และตารางเงื่อนไขมาสร้างเป็นคำสั่งภาษา SQL เมื่อได้คำสั่งภาษา SQL แล้ว(หนึ่งคำสั่งภาษา SQL จะเท่ากับหนึ่งเงื่อนไขแต่อาจจะมาจากหลายรายการข้อมูลก็ได้) จะนำคำสั่งภาษา SQL แต่ละคำสั่งไปตรวจสอบหาข้อมูลที่ผิดพลาดในระบบฐานข้อมูล โดยจำนวนรายการข้อมูลที่เป็นไปตามเงื่อนไขนั้นก็คือจำนวนรายการข้อมูลที่ผิดพลาด ซึ่งเราสามารถนำเอาจำนวนรายการข้อมูลเหล่านี้ไปเก็บไว้ในตัวแปร Counter ทำเช่นนี้กับทุก ๆ เงื่อนไขที่มีอยู่ในตารางกฎเกณฑ์ก็จะได้ผลลัพธ์ที่เป็นจำนวนรายการข้อมูลที่ผิดพลาดทั้งหมด (N_r)ออกมา โดยสามารถอธิบายขั้นตอนการทำงานได้ดังแผนผังในรูปภาพที่ 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 13

แสดงตัวอย่างตารางเงื่อนไขที่จัดเก็บส่วนของเงื่อนไขในขั้นตอนที่ 3

Rule Number	Primary Table Name	Secondary Table Name	Relation Number	Condition	Group By	Having	Connect Condition
001	STEDENT_BIO1	FACULTY	002	FACULTY.FAC_CODE = "4"	NULL	NULL	AND
002	STUDENT_BIO1	CUMULATIVE	001	CUMULATIVE.ATMP_SEM > 24	NULL	NULL	NULL

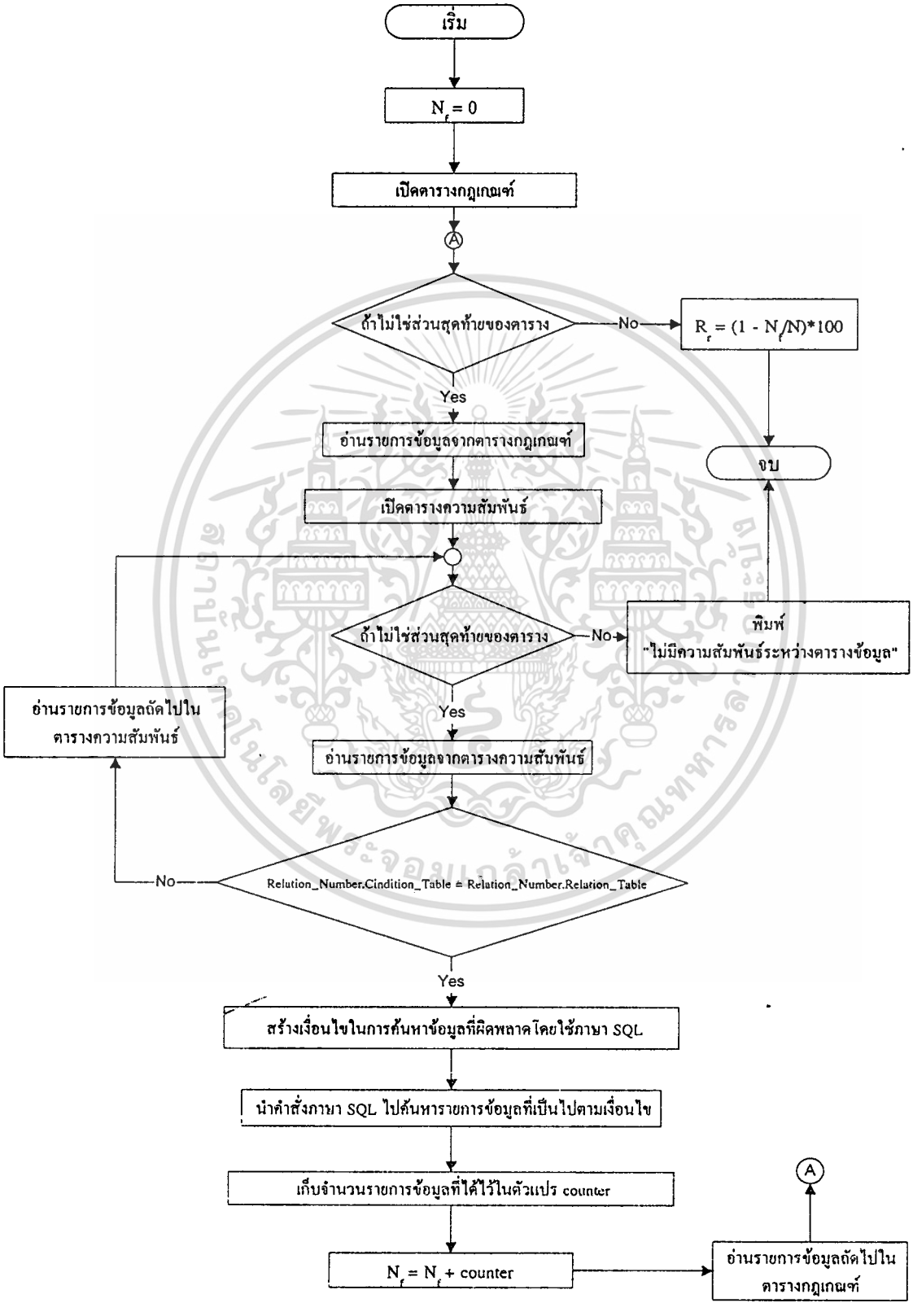
ตารางที่ 14

แสดงตัวอย่างตารางความสัมพันธ์ที่เก็บส่วนของความสัมพันธ์ในเงื่อนไขในขั้นตอนที่ 3

Relation Number	Primary Table Name	Secondary Table Name	Primary Foreign Key	Secondary Foreign Key
001	STUDENT_BIO1	CUMULATIVE	STU_CODE	STU_CODE
002	STUDENT_BIO1	FACULTY	FAC_CODE	FAC_CODE

ภาพที่ 7

แสดงขั้นตอนการทำงานของระบบการประเมินค่าความน่าเชื่อถือของข้อมูลลักษณะ R_r



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.5 คำนวณหาค่าความน่าเชื่อถือของข้อมูล

หลังจากนำเอาเงื่อนไขที่เก็บไว้ในตารางกฎเกณฑ์ไปเปรียบเทียบค้นหาข้อมูลที่ผิดพลาดในฐานข้อมูลแล้ว เราจะได้จำนวนรายการข้อมูลที่ผิดพลาดทั้งหมดออกมา การคำนวณหาค่าความน่าเชื่อถือนั้นสามารถทำได้โดยการนำเอาจำนวนรายการข้อมูลที่ผิดพลาดทั้งหมดที่ได้มานั้นไปคำนวณหาค่าความน่าเชื่อถือของข้อมูล ได้จากสมการที่ 3.6

ค่าความน่าเชื่อถือที่คำนวณได้นั้นจะเป็นค่าของเปอร์เซ็นต์ความน่าเชื่อถือ ถ้าค่าเปอร์เซ็นต์ความน่าเชื่อที่คำนวณได้มีค่ามากก็สามารถอนุมานได้ว่าข้อมูลที่นำมาคำนวณหาค่าความน่าเชื่อถือนั้นมีความถูกต้องสูง แต่ในทางกลับกันถ้าค่าเปอร์เซ็นต์ความน่าเชื่อที่คำนวณได้มีค่าน้อยก็สามารถอนุมานได้ว่าข้อมูลที่นำมาคำนวณหาค่าความน่าเชื่อถือนั้นไม่ค่อยมีความถูกต้องเท่าที่ควร

จะเห็นได้ว่าการสร้างฐานของกฎเกณฑ์ในลักษณะของโปรแกรมนั้นจะทำให้เป็นการยากในการแก้ไขเปลี่ยนแปลงในกรณีที่เกิดการเปลี่ยนแปลงทางโครงสร้างของระบบฐานข้อมูล แต่สามารถแก้ปัญหาได้ด้วยการแยกเขียนโปรแกรมออกเป็นโมดูลได้ โดยในแต่ละโมดูลนั้นจะใช้แทนลักษณะเงื่อนไขแต่ละเงื่อนไข จากนั้นจึงนำโมดูลเหล่านี้ไปตรวจสอบหาข้อมูลที่ผิดพลาดตามเงื่อนไขที่กำหนดไว้ เมื่อได้จำนวนรายการข้อมูลที่ผิดพลาดแล้วจะนำไปคำนวณหาค่าความน่าเชื่อถือของข้อมูลจากสมการ 3.6 อย่างไรก็ตามผู้ที่จะสร้างเงื่อนไขลักษณะนี้ได้จะต้องมีความรู้เกี่ยวกับภาษาของโปรแกรมที่จะเขียนเป็นอย่างดีซึ่งไม่เหมือนกับวิธีการสร้างฐานของกฎเกณฑ์ในลักษณะของตารางข้อมูลเงื่อนไขที่กล่าวมาข้างต้นผู้ที่มีความรู้เพียงโครงสร้างของระบบฐานข้อมูลและโครงสร้างของภาษา SQL ก็สามารถสร้างระบบฐานของกฎเกณฑ์ได้ แต่ทั้งนี้ทั้งนั้นก็จะต้องมีการสร้างฐานของกฎเกณฑ์ลักษณะนี้ขึ้นมาเพื่อให้ได้มาซึ่งระบบผู้เชี่ยวชาญที่มีเงื่อนไขครบถ้วนสมบูรณ์จึงจะทำให้ได้ค่าความน่าเชื่อถือของข้อมูลที่มีความถูกต้องมากที่สุด

5.3 ผลการทดลอง

ในการทดลองจะกระทำกับฐานข้อมูลนักศึกษา โดยจะแบ่งการทดลองออกเป็น 3 ครั้ง ซึ่งในแต่ละครั้งจะใช้จำนวนรายการข้อมูลที่ไม่เท่ากันเพื่อทำการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลชนิด R_r ด้วยวิธีการประเมินโดยใช้ระบบผู้เชี่ยวชาญที่สร้างขึ้น กับวิธีการประเมินโดยใช้เจ้าหน้าที่ที่เกี่ยวข้องกับฐานข้อมูลนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.1 ข้อกำหนดในการทดลอง

1. เครื่องคอมพิวเตอร์ที่ใช้เป็นเครื่องคอมพิวเตอร์ระดับบุคคล(personal computer) ซึ่งมีลักษณะรายละเอียดดังนี้

1.1 มี CPU ขนาด 486 DX2-66

1.2 หน่วยความจำภายใน(RAM) ขนาด 8 MB.

1.3 หน่วยความจำ Cache ขนาด 256 KB.

2. ใช้โปรแกรม Microsoft Access version 2.0 เป็นโปรแกรมจัดการฐานข้อมูล

3. ใช้บุคลากรที่เกี่ยวข้องกับระบบฐานข้อมูลนี้ 2 ท่าน โดยทั้งสองท่านนี้เป็นเจ้าหน้าที่ของศูนย์คอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ และเป็นผู้ที่ออกแบบและสร้างระบบฐานข้อมูลนักศึกษาที่ใช้เป็นตัวอย่างในการทดลอง

5.3.2 ขั้นตอนในการทดลอง

ในการทดลองจะมีการกำหนดเงื่อนไขบางเงื่อนไขให้กับระบบผู้เชี่ยวชาญเพื่อทดสอบการประเมินความน่าเชื่อถือของข้อมูลในลักษณะ R_c โดยจะมีเงื่อนไขดังต่อไปนี้

เงื่อนไขที่ 1 "ค้นหานักศึกษาคณะวิทยาศาสตร์ที่ลงทะเบียนเรียนในแต่ละเทอมมากกว่า 24 หน่วยกิต"

เงื่อนไขที่ 2 "ค้นหานักศึกษาที่มีสถานภาพเป็น จบการศึกษา ลาออก พ้นสภาพการเป็นนักศึกษา แต่ยังมีการลงทะเบียนเรียนอยู่"

เงื่อนไขที่ 3 "ค้นหานักศึกษาคณะวิทยาศาสตร์ที่จบการศึกษาโดยมีจำนวนหน่วยกิตน้อยกว่า 144 หน่วยกิต"

จากเงื่อนไขที่กำหนดไว้ข้างต้นสามารถเขียนเป็นคำสั่งภาษา SQL ในโปรแกรม Microsoft Access เวอร์ชั่น 2.0 ได้ดังนี้

เงื่อนไขที่ 1

```
SELECT      STUDENT_BIO1.STU_CODE
FROM        (STUDENT_BIO1 INNER JOIN FACULTY ON
            STUDENT_BIO.FAC_CODE = FACULTY.FAC_CODE)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        INNER JOIN CUMULATIVE ON
        STUDENT_BIO1.STU_CODE = CUMULATIVE.STU_CODE)
WHERE   (FACULTY.FAC_CODE = "4") AND (CUMULATIVE.ATMP_SEM > 24);
```

ตารางที่ 15

แสดงความสัมพันธ์ของเงื่อนไขที่ใช้ในการทดลอง

Relation Number	Primary Table Name	Secondary Table Name	Primary Foreign Key	Secondary Foreign Key
001	CLASS	COURSE	COURSE_CODE	COURSE_CODE
002	COURSE	CLASS	COURSE_CODE	COURSE_CODE
003	COURSE	DEPARTMENT	DEPT_CODE	DEPT_CODE
004	COURSE	FACULTY	FAC_CODE	FAC_CODE
005	COURSE	REGISTRATION	COURSE_CODE	COURSE_CODE
006	COURSE_SECTION	REGISTRATION	COURSE_CODE	COURSE_CODE
007	CUMULATIVE	REGISTRATION	STU_CODE	STU_CODE
008	DEPARTMENT	COURSE	DEPT_CODE	DEPT_CODE
009	DEPARTMENT	STUDENT_BIO1	DEPT_CODE	DEPT_CODE
010	DIVISION	STUDENT_BIO1	DIV_CODE	DIV_CODE
011	FACULTY	COURSE	FAC_CODE	FAC_CODE
012	FACULTY	STUDENT_BIO1	FAC_CODE	FAC_CODE
013	REGISTRATION	COURSE	COURSE_CODE	COURSE_CODE
014	REGISTRATION	COURSE_SECTION	COURSE_CODE	COURSE_CODE
015	REGISTRATION	STUDENT_BIO1	STU_CODE	STU_CODE
016	STUDENT_BIO1	CUMULATIVE	STU_CODE	STU_CODE
017	STUDENT_BIO1	DEPARTMENT	DEPT_CODE	DEPT_CODE
018	STUDENT_BIO1	DIVISION	DIV_CODE	DIV_CODE
019	STUDENT_BIO1	FACULTY	FAC_CODE	FAC_CODE
020	STUDENT_BIO1	REGISTRATION	STU_CODE	STU_CODE
021	STUDENT_BIO1	STUDENT_BIO2	STU_CODE	STU_CODE
022	STUDENT_BIO1	STUDENT_TYPE	STU_TYPE_CODE	STU_TYPE_CODE
023	STUDENT_BIO1	STUDENT_STATUS	STU_STATUS	STU_ST_CODE
024	STUDENT_BIO2	STUDENT_BIO1	STU_CODE	STU_CODE
025	STUDENT_TYPE	STUDENT_BIO1	STU_TYPE_CODE	STU_TYPE_CODE
026	STUDENT_STATUS	STUDENT_BIO1	STU_ST_CODE	STU_STATUS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 16

แสดงตารางเงื่อนไขที่จัดเก็บส่วนของเงื่อนไขในการทดลอง

Rule Number	Primary Table Name	Secondary Table Name	Relation Number	Condition	Group By	Having	Connect Condition
001	STEDENT_BIO1	FACULTY	019	FACULTY.FACCODE = "4"	NULL	NULL	AND
002	STUDENT_BIO1	CUMULATIVE	016	CUMULATIVE.ATMP_SEM > 24	NULL	NULL	NULL
003	STUDENT_BIO1	STUDENT_STATUS	023	STUDENT_STATUS.STU_ST_CODE = "1" OR STUDENT_STATUS.STU_ST_CODE = "2" OR STUDENT_STATUS.STU_ST_CODE = "3"	NULL	NULL	AND
004	STUDENT_BIO1	REGISTRATION	020	NULL	NULL	NULL	NULL
005	STUDENT_BIO1	FACULTY	019	NULL	STUDENT_BIO1.STU_CODE , FACULTY.FAC_CODE	FACULTY.FAC_CODE = "4"	AND
006	STUDENT_BIO1	CUMULATIVE	016	NULL	CUMULATIVE.STU_CODE	Max(CUMULATIVE.ERN_CUM) > 144	AND
007	STUDENT_BIO1	STUDENT_STATUS	023	NULL	STUDENT_STATUS.STU_ST_CODE	STUDENT_STATUS.STU_ST_CODE = "1"	NULL

เงื่อนไขที่ 2

```
SELECT    STUDENT_BIO1.STU_CODE
FROM      (STUDENT_BIO1 INNER JOIN REGISTRATION ON
          STUDENT_BIO1.STU_CODE = REGISTRATION.STU_CODE)
          INNER JOIN STUDENT_STATUS ON
          STUDENT_BIO1.STU_STATUS = STUDENT_STATUS.STU_ST_CODE
WHERE     (STUDENT_STATUS.STU_ST_CODE = "1" OR
          STUDENT_STATUS.STU_ST_CODE = "2" OR
          STUDENT_STATUS.STU_ST_CODE = "3" );
```

เงื่อนไขที่ 3

```
SELECT    STUDENT_BIO1.STU_CODE
FROM      (((STUDENT_BIO1 INNER JOIN FACULTY ON
          STUDENT_BIO1.FAC_CODE = FACULTY.FAC_CODE)
          INNER JOIN CUMULATIVE ON
          STUDENT_BIO1.STU_CODE =CUMULATIVE.STU_CODE)
          INNER JOIN STUDENT_STATUS ON
          STUDENT_BIO1.STU_STATUS = STUDENT_STATUS.STU_ST_CODE)
GROUP BY  STUDENT_BIO1.STU_CODE, FACULTY.FAC_CODE,
          CUMULATIVE.STU_CODE, STUDENT_STATUS.STU_ST_CODE
HAVING    (FACULTY.FAC_CODE = "4") AND
          (Max(CUMULATIVE.ERN_CUM) < 144 ) AND
          (STUDENT_STATUS.STU_ST_CODE = "1");
```

เมื่อได้คำสั่งภาษา SQL แล้วจะไปเก็บในตารางความสัมพันธ์ และตารางเงื่อนไข ตัวอย่างแสดงในตารางที่ 15 และตารางที่ 16

5.3.3 ผลการทดลองที่ได้

การทดลองครั้งที่ 1 ใช้รายการข้อมูลทั้งหมด 300 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 3 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่ และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญดัง ได้ผลลัพธ์ดังตารางที่ 17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 17

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 1

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	300	300
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	3	3
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	1	3
4. ค่าความน่าเชื่อถือของข้อมูล	99.67%	99%
5. เปอร์เซ็นต์ความถูกต้อง	100%	100%

การทดลองครั้งที่ 2 ใช้รายการข้อมูลทั้งหมด 600 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 10 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญ ได้ผลลัพธ์ดังตารางที่ 18

ตารางที่ 18

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 2

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	600	600
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	10	10
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	หาไม่พบ	10
4. ค่าความน่าเชื่อถือของข้อมูล	ไม่สามารถคำนวณได้	98.3%
5. เปอร์เซ็นต์ความถูกต้อง	ไม่สามารถคำนวณได้	100%

หมายเหตุ ผลของการทดสองที่เจ้าหน้าที่ค้นหาจำนวนรายการข้อมูลที่ผิดพลาดและได้ผลลัพธ์เป็น "หาไม่พบ" หมายถึงเจ้าหน้าที่ได้ทำการค้นหาจนกระทั่งถึงรายการข้อมูลสุดท้ายแต่ไม่พบรายการข้อมูลที่ผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองครั้งที่ 3 ใช้รายการข้อมูลทั้งหมด 1000 รายการ และจำลองรายการข้อมูลที่เกิดความผิดพลาดไว้ทั้งสิ้น 13 รายการ โดยได้ผลการเปรียบเทียบการประเมินความน่าเชื่อถือของข้อมูลระหว่างวิธีการประเมินโดยใช้เจ้าหน้าที่และวิธีประเมินโดยใช้ระบบผู้เชี่ยวชาญ ได้ผลลัพธ์ดังตารางที่ 19

ตารางที่ 19

แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการประเมินความน่าเชื่อถือของข้อมูลด้วยวิธีการใช้เจ้าหน้าที่และวิธีการใช้ระบบผู้เชี่ยวชาญ จากการทดลองครั้งที่ 3

รายการทดสอบ	เจ้าหน้าที่	ระบบผู้เชี่ยวชาญ
1. จำนวนรายการข้อมูลทั้งหมด	1000	1000
2. จำนวนรายการข้อมูลที่ผิดพลาดตามความเป็นจริง	13	13
3. จำนวนรายการข้อมูลที่ผิดพลาดที่หาได้	ไม่สามารถหาได้	13
4. ค่าความน่าเชื่อถือของข้อมูล	ไม่สามารถคำนวณได้	98.7%
5. เปอร์เซ็นต์ความถูกต้อง	ไม่สามารถคำนวณได้	100%

หมายเหตุ ผลของการทดลองที่เจ้าหน้าที่ค้นหาจำนวนรายการข้อมูลที่ผิดพลาดและได้ผลลัพธ์เป็น "ไม่สามารถหาได้" หมายถึงเจ้าหน้าที่ไม่สามารถที่จะทำการค้นหาข้อมูลที่ผิดพลาดได้เพราะข้อมูลทั้งหมดมีจำนวนมากเกินความสามารถของเจ้าหน้าที่

5.3.4 สรุปผลการทดลอง

จากผลการทดลองสรุปได้ว่า การประเมินความน่าเชื่อถือของข้อมูลในลักษณะ R_r โดยการใช้ระบบผู้เชี่ยวชาญนี้ สามารถที่จะประเมินหาค่าความน่าเชื่อถือของข้อมูลได้ถูกต้องกว่าการประเมินความน่าเชื่อถือของข้อมูลโดยใช้เจ้าหน้าที่ และจะใช้เวลาในการประเมินที่น้อยกว่าวิธีที่ใช้เจ้าหน้าที่มาก ซึ่งจะเห็นได้ว่ากรณีที่ระบบฐานข้อมูลมีขนาดใหญ่และมีจำนวนรายการข้อมูลมาก การประเมินความน่าเชื่อถือของข้อมูลโดยการใช้เจ้าหน้าที่ไม่สามารถที่จะทำได้เลย(ดังการทดลองครั้งที่ 2 และครั้งที่ 3)เพราะจะไม่คุ้มกับเวลาที่เสียไป ในกรณีเช่นนี้จึงเหมาะที่จะใช้การประเมินความน่าเชื่อถือของข้อมูลโดยระบบผู้เชี่ยาชาญ แต่ทั้งนี้ทั้งนั้นการประเมินค่าความน่าเชื่อถือของข้อมูลโดยใช้ระบบผู้เชี่ยวชาญจะให้ค่าความถูกต้องสมบูรณ์หรือไม่จะต้องขึ้นอยู่กับการกำหนดเงื่อนไขของระบบฐานข้อมูลที่มีอยู่ทั้งหมดให้กับระบบผู้เชี่ยวชาญโดยครบถ้วน มิฉะนั้นจะทำให้ค่าความน่าเชื่อถือของข้อมูลที่ประเมินได้นั้นมีความคลาดเคลื่อนไปจากความเป็นจริงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในกรณีที่ได้มีการเปลี่ยนแปลงเกี่ยวกับโครงสร้างของระบบฐานข้อมูลในส่วนที่กระทบกระเทือนต่อเงื่อนไขที่ใช้ในระบบผู้เชี่ยวชาญนั้น จะต้องทำการปรับปรุงในส่วนของเงื่อนไขในระบบผู้เชี่ยวชาญให้มีความถูกต้องสมบูรณ์อยู่เสมอจึงจะทำให้ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลให้ค่าที่ถูกต้องตามความเป็นจริงมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

ระบบผู้เชี่ยวชาญเพื่อการประเมินค่าความน่าเชื่อถือของข้อมูลตามความเป็นจริง

(Expert System for Absolute Reliability)

การประเมินค่าความน่าเชื่อถือของข้อมูลตามความเป็นจริงนั้นจะเป็นการประเมินว่าข้อมูลทุก ๆ รายการข้อมูลที่มีอยู่มีความผิดพลาดไปจากความเป็นจริงหรือไม่ โดยในการศึกษาพบว่าในระบบฐานข้อมูลนั้นจะประกอบไปด้วยข้อมูลที่ลักษณะแตกต่างกันมากมายไม่ว่าจะเป็นข้อมูลชนิดตัวเลข(numeric), ตัวอักษร(character) ,ชนิดวันที่(date) ฯลฯ ซึ่งการเกิดความผิดพลาดของข้อมูลนั้นสามารถเกิดขึ้นกับข้อมูลทุก ๆ ชนิด เพราะฉะนั้นระบบการประเมินหาค่าน่าเชื่อถือของข้อมูลโดยคอมพิวเตอร์นั้นจะทำได้โดยต้องสร้างให้ระบบคอมพิวเตอร์สามารถรู้จักและเข้าใจกับข้อมูลทุกๆ ข้อมูลหรือรู้จักคำทุกคำค่าทุกค่าที่ประกอบกันเป็นข้อมูล อนึ่งอาจกล่าวได้ว่าการที่จะให้ระบบคอมพิวเตอร์ตรวจสอบความผิดพลาดของข้อมูลได้นั้นจะต้องทำให้ระบบคอมพิวเตอร์สามารถรู้จักและเข้าใจถึงภาษาธรรมชาติได้ ซึ่งภาษาธรรมชาติได้แก่ ข้อมูลที่เป็นตัวอักษรและตัวเลขภาษาไทย ตัวอักษรและตัวเลขภาษาอังกฤษ เป็นต้น

เมื่อทำให้ระบบคอมพิวเตอร์สามารถเข้าใจถึงภาษาธรรมชาติได้แล้วก็ใช่ว่าจะนำระบบคอมพิวเตอร์ไปประเมินหาข้อมูลที่ผิดพลาดได้เพราะการที่ระบบคอมพิวเตอร์สามารถเข้าใจภาษาธรรมชาติได้นั้น ก็เป็นเพียงแค่สามารถใช้ระบบคอมพิวเตอร์ตรวจสอบได้เฉพาะข้อมูลที่ผิดไปจากไวยกรณ์ของแต่ละภาษาเท่านั้น แต่ไม่สามารถที่จะตรวจสอบได้ว่าข้อมูลนั้นเป็นข้อมูลที่ผิดพลาดได้หรือไม่ ตัวอย่างเช่น ถ้ามีการเก็บข้อมูลรายชื่อนักศึกษาคนหนึ่งเป็น “สมชาย” แต่ข้อมูลรายชื่อของนักศึกษาตามความเป็นจริงนั้นคือ “สมหมาย” ซึ่งเห็นว่าระบบคอมพิวเตอร์ไม่มีทางที่จะตรวจสอบได้เลยว่าข้อมูลรายชื่อที่ถูกต้องของนักศึกษาคนนี้คือ “สมชาย” หรือ “สมหมาย” เพราะทั้งสองรายชื่อนั้นถูกต้องตามไวยกรณ์ของภาษาทุกอย่าง ซึ่งจะเห็นแล้วว่าที่พยายามจะทำให้ระบบคอมพิวเตอร์สามารถเข้าใจถึงภาษาธรรมชาติได้นั้นก็ไม่สามารถที่จะประเมินหารายการข้อมูลที่ผิดพลาดได้

จากการวิจัยพบว่าวิธีที่เหมาะสมและได้ผลที่สุดในจะประเมินหาค่าความน่าเชื่อถือชนิดนี้คือให้เจ้าที่ที่เกี่ยวข้องนำรายการข้อมูลไปตรวจสอบกับข้อมูลที่มีอยู่จริงไม่ว่าจะเป็นข้อมูลที่เก็บอยู่ลักษณะของเอกสารหรือข้อมูลตามความเป็นจริง ซึ่งการทำเช่นนี้ถ้าเป็นการกระทำกับฐานข้อมูลที่มีขนาดเล็ก(มีจำนวนรายการข้อมูลไม่มากนัก)ก็สามารถทำได้โดยไม่ยากนัก แต่ถ้าเป็นการกระทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กับฐานข้อมูลที่มีขนาดใหญ่มาก ๆ (มีจำนวนรายการข้อมูลมากมาย) เราไม่สามารถที่จะนำเอารายการข้อมูลทั้งหมดมาตรวจสอบได้ ซึ่งปัญหาเหล่านี้เราสามารถแก้ไขได้โดยการสุ่มเอารายการเฉพาะข้อมูลบางรายการเท่านั้นมาตรวจสอบ โดยการสุ่มรายการข้อมูลนั้นจะอ้างอิงตามเทคนิคการสุ่มตัวอย่างของ ดร.อนันต์ ศรีโสภา ดังรายละเอียดในหัวข้อถัดไป

6.1 เทคนิคการสุ่มตัวอย่าง

การสุ่มตัวอย่างเป็นวิธีการคัดเลือกสมาชิกจากประชากรที่เราสนใจ เพื่อเก็บรวบรวมข้อมูลจากตัวอย่างนั้นไปใช้ในการกะประมาณค่าเฉลี่ยของประชากรทั้งหมดที่ต้องการ อนึ่งกล่าวได้ว่าการสุ่มตัวอย่างมีจุดมุ่งหมายเพื่อต้องการหาข้อสรุปเกี่ยวกับประชากรโดยอาศัยข้อมูลจากตัวอย่างที่สุ่มได้จากประชากรนั้น การที่จะทำให้ตัวอย่างที่สุ่มมาเป็นตัวแทนของประชากรได้ดีเพียงไรนั้นย่อมขึ้นอยู่กับขนาดของตัวอย่าง เทคนิคการสุ่มและเครื่องมือในการเก็บข้อมูล เพราะฉะนั้นในหัวข้อต่อไปนี้จะเป็นกล่าวถึง การสุ่มตัวอย่างอย่างไรจึงจะทำให้ได้ข้อมูลที่เป็นตัวแทนที่ดีของประชากรทั้งหมด

6.1.1 คำศัพท์เฉพาะ

สมาชิก เป็น element ของประชากรที่เราต้องการจะวัด เช่น เด็กชายดำเป็นสมาชิกของประชากรนักเรียนชั้นประถมปีที่ 4 เราต้องการจะวัดว่าเด็กชายดำมีผลสัมฤทธิ์ในวิชาเลขคณิตเท่าใดเป็นต้น

ประชากร (Population) ประกอบด้วยสมาชิกต่าง ๆ ที่เราต้องการจะสุ่มตัวอย่างซึ่งแบ่งออกเป็น 2 ชนิด คือ finite population และ infinite population สำหรับ finite population เป็นประชากรที่มีสมาชิกจำนวนจำกัด เราสามารถทำการนับได้ให้ครบถ้วน เช่น ประชากรนักเรียนชั้นมัธยมปีที่ 1 โรงเรียนสวนกุหลาบวิทยาลัย เป็นต้น ส่วน infinite population นั้นเป็นประชากรที่มีจำนวนสมาชิกไม่จำกัด เราไม่สามารถจะนับจำนวนสมาชิกให้ครบถ้วนได้ เช่น ประชากรของเม็ดทรายที่ชายหาดบางแสน ประชากรของการดีใจของชายคนหนึ่งตั้งแต่เกิดจนตาย ฯลฯ เป็นต้น อย่างไรก็ดีประชากรที่เราต้องการจะสุ่มตัวอย่างหรือเรียกอีกอย่างหนึ่งว่า target population นั้น จำเป็นจะต้องนิยามขอบเขตให้ชัดเจนเสียก่อนที่จะทำการสุ่มตัวอย่าง

หน่วยการสุ่ม (Sampling units) ก่อนที่จะทำการสุ่มตัวอย่าง จะต้องแบ่งประชากรออกเป็นส่วน ๆ ซึ่งเรียกว่าหน่วยการสุ่ม หรือ sampling units หน่วยการสุ่มประกอบด้วยสมาชิกหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวหรือมากกว่าก็ได้ แต่สมาชิกใดเมื่ออยู่ในหน่วยใดแล้วจะไปปรากฏที่หน่วยอื่นอีกไม่ได้ หน่วยการสุ่มอาจจะเป็นนักเรียนแต่ละคน หรือห้องเรียนแต่ละห้องหรือโรงเรียนแต่ละแห่งก็ได้ ฯลฯ เป็นต้น

ตัวอย่าง (sample) เป็นกลุ่มของหน่วยการสุ่ม (sampling unit) ที่เลือกสุ่มมาจากบัญชีรายชื่อสมาชิกทั้งหมดของประชากร โดนเราจะเก็บรวบรวมข้อมูลจากสมาชิกในตัวอย่าง เพื่อนำมาใช้อธิบายประชากรที่เราสนใจ เช่นในการเลือกตัวอย่างของผู้ที่มีสิทธิออกเสียงลงคะแนนจากบัญชีรายชื่อผู้ที่มีสิทธิออกเสียงเลือกตั้งทั้งหมดมาจำนวนหนึ่ง แล้วถามทัศนะของแต่ละคนในตัวอย่างที่เลือกมานี้ว่า ชอบผู้สมัครรับเลือกตั้งคนไหนมากที่สุด จากข้อมูลที่เก็บได้จากตัวอย่างนี้เราก็หาข้อสรุปได้ว่าประชาชนทั้งหมดในหมู่บ้านนี้มีทัศนะของใครมากที่สุด เช่นนี้เป็นต้น

6.1.2 ขั้นตอนในการสุ่มตัวอย่าง

เนื่องจากการสุ่มตัวอย่างมีตั้งแต่ง่ายถึงยาก และตั้งแต่ขนาดเล็กจนถึงขนาดใหญ่ แต่อย่างไรก็ตามเราต้องการให้ตัวอย่างที่สุ่มได้นั้นสามารถเป็นตัวแทนของประชากร และเป็นไปอย่างประหยัด การที่จะบรรลุจุดมุ่งหมายนี้ได้จำเป็นจะต้องคำนึงถึงขั้นตอนในการสุ่มตัวอย่างที่สำคัญดังต่อไปนี้

1) การกำหนดจุดมุ่งหมายในการสุ่มตัวอย่าง ก่อนเริ่มทำการสุ่มตัวอย่าง ผู้สุ่มจะต้องกำหนดจุดมุ่งหมายในการสุ่มตัวอย่างให้ชัดเจน ว่าต้องการจะสุ่มอะไรบ้าง สุ่มจำนวนมากน้อยเท่าใด จะสุ่มอย่างไร สุ่มที่ไหนและจะนำข้อมูลที่ได้ไปใช้ทำอะไร

2) การนิยามประชากรที่จะทำการสุ่มตัวอย่าง เนื่องจากประชากรเป็นที่รวมของสมาชิกทั้งหมดที่เราต้องการจะสุ่มตัวอย่าง จำเป็นจะต้องนิยามขอบเขตของประชากรให้ชัดเจน เพื่อช่วยให้ผู้แจงนับ(enumerator) สามารถทำการตัดสินใจได้ทันทีว่าอันไหนเป็นสมาชิกของประชากรและอันไหนไม่เป็น นอกจากนี้ควรจะต้องทราบสภาพธรรมชาติของประชากรเป็นอย่างไร มีขอบเขตแค่ไหน ประกอบด้วยสมาชิกกี่ชนิด มีจำนวนจำกัดหรือไม่จำกัด มีหน่วยการวัดเป็นอย่างไร เป็นต้น

3) การระบุระดับความถูกต้องในการจัดเก็บข้อมูล เนื่องจากการจัดเก็บข้อมูลจากตัวอย่างมักจะมีความคลาดเคลื่อนที่เกิดจากการวัด (measurement error) และความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง(sampling error) รวมอยู่ด้วยเสมอ การแก้ปัญหานี้อาจเพิ่มขนาดตัวอย่างให้มากขึ้น ใช้เครื่องมือการวัดที่มีคุณภาพดีขึ้น

4) การเลือกตัวอย่าง ในขั้นนี้ผู้สุ่มตัวอย่างจะต้องระบุวิธีในการเลือกตัวอย่างให้เหมาะสมกับธรรมชาติของประชากรที่จะสุ่ม ซึ่งมีวิธีหรือเทคนิคในการสุ่มตัวอย่างหลายวิธีด้วยกัน อย่างไรก็

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดีไม่ว่าจะใช้เทคนิคการสุ่มตัวอย่างแบบใดก็ตาม ควรจะต้องประมาณขนาดของตัวอย่าง (sample size) อย่างคร่าว ๆ ไว้ล่วงหน้าโดยอาศัยระดับความถูกต้องในการจัดเก็บข้อมูลเป็นเกณฑ์ในการคำนวณ นอกจากนี้ควรจะต้องคำนึงถึงค่าใช้จ่ายและเวลาที่จะต้องใช้ในการสุ่มว่ามีมากน้อยแตกต่างกันอย่างไรอีกด้วย

5) ทำการวิเคราะห์และสรุปข้อมูลที่สุ่มได้ หลังจากได้ทำการตรวจสอบความถูกต้องของข้อมูลที่เก็บได้แล้ว ขั้นต่อไปก็นำข้อมูลเหล่านั้นมาคำนวณการประมาณว่า พารามิเตอร์ ของประชากรจากค่าสถิติต่าง ๆ ที่ได้จากข้อมูลเหล่านี้เป็นเท่าใด อนึ่งในการสรุปผลการประมาณในตัวแปรที่สำคัญของประชากร ควรจะได้ระบุขอบเขตความคลาดเคลื่อนในการประมาณซึ่งอยู่ในลักษณะของความน่าจะเป็น (probability)

6.1.3 การสุ่มตัวอย่างเพื่อหาค่าสัดส่วนของประชากรที่มีการแจกแจงแบบไบโนเมียล (sampling for proportion in binomial distribution)

บางครั้งประชากรที่เราสนใจอาจประกอบด้วยหน่วย(unit) ซึ่งแบ่งออกได้เป็นสองพวกหรือสองกลุ่ม อาทิเช่น ลักษณะของรายการข้อมูลที่สุ่มออกมาจากระบบฐานข้อมูลเพื่อตรวจความถูกต้อง โดยผลที่ต้องการจะมีอยู่สองแบบเท่านั้น คือ "ผิด" และ "ไม่ผิด" เป็นต้น เราจะเรียกลักษณะของประชากรประเภทนี้ว่า ประชากรที่มีการแจกแจงแบบไบโนเมียล(bimomial distribution) โดยจะกล่าวได้ว่าทุกหน่วย (unit) ในประชากรจะต้องตกอยู่ในกลุ่ม C หรือไม่ก็อยู่ในกลุ่ม C

ถ้าสมมติค่าของ y_i ในประชากรให้เป็นดังนี้

y_i เป็น 1 ถ้าหน่วยนี้ตกอยู่ในกลุ่ม C

y_i เป็น 0 ถ้าหน่วยนี้ตกอยู่ในกลุ่ม C

โดยที่

A เป็น จำนวนของประชากรทั้งหมด(population total)ที่อยู่ในกลุ่ม C จะได้

$$A = \sum_{i=1}^{N} y_i \qquad (6.1)$$

P เป็น ค่าสัดส่วน(proportion)ของหน่วยประชากรที่อยู่ในกลุ่ม C จากจำนวนประชากรทั้งหมด

$$P = \sum_{i=1}^{N} \frac{y_i}{N} = \frac{A}{N} \qquad (6.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และให้ p เป็น ค่าสัดส่วนของหน่วยประชากรที่อยู่ในกลุ่ม C ที่หาได้จากกลุ่มของตัวอย่าง

$$p = \sum_{i=1}^{n} \frac{y_i}{n} = \frac{a}{n} \qquad (6.3)$$

$\hat{A}$ เป็น ค่ากะประมาณของ A

$$\hat{A} = Np \qquad (6.4)$$

เมื่อ

a คือ จำนวนหน่วยที่อยู่ในกลุ่ม C ของตัวอย่าง

n คือ จำนวนหน่วยทุกหน่วยของตัวอย่าง

N คือ จำนวนหน่วยทุกหน่วยของประชากร

ในการกะประมาณค่าสัดส่วนของประชากรจากค่าสัดส่วนที่หาได้จากกลุ่มของตัวอย่างนั้น ย่อมจะต้องมีความคลาดเคลื่อนเกิดขึ้น เพราะฉะนั้นจึงให้ค่า S เป็นค่าความคลาดเคลื่อนมาตรฐาน (standard error) โดยคำนวณได้จาก

ค่าความคลาดเคลื่อนของการกะประมาณค่าสัดส่วนของประชากร คือ

$$S = \sqrt{\frac{N - n}{(n - 1)N} pq} \qquad (6.5)$$

เมื่อ

$$q = 1 - p \qquad (6.6)$$

6.1.4 การเลือกขนาดตัวอย่างสำหรับการกะประมาณค่าสัดส่วนของประชากร

การตัดสินใจประการแรกในการวางแผนสุ่มตัวอย่างได้แก่การกำหนดขนาดของตัวอย่าง หรือ sample size เพราะว่าถ้าใช้ขนาดตัวอย่างใหญ่เกินไป ก็ทำให้สิ้นเปลืองแรงงาน เงินและเวลา แต่ถ้าใช้ขนาดตัวอย่างเล็กเกินไปก็ทำให้ผลลัพธ์ที่ได้ขาดความเที่ยงตรงและความน่าเชื่อถือ อย่างไรก็ดีเราไม่มีความพอใจกับการตัดสินใจเกี่ยวกับขนาดของตัวอย่างมากนัก ทั้งนี้เนื่องจากไม่มีข้อมูลเพียงพอที่จะช่วยให้ได้ขนาดตัวอย่างที่ดีที่สุด โดยที่ทฤษฎีการสุ่มตัวอย่าง เป็นเพียงส่วนช่วยให้เราสามารถกำหนดขนาดตัวอย่างได้อย่างมีเหตุมีผลมากขึ้นเท่านั้น

เนื่องจากการกำหนดความถูกต้องในการกะประมาณนั้นขึ้นอยู่กับจำนวนความคลาดเคลื่อนที่ได้จากการสุ่ม บางครั้งเป็นเรื่องยากที่จะตัดสินใจให้มีความเคลื่อนเท่าใดได้ เพราะผลที่ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการสุ่มตัวอย่างอาจจะนำไปใช้ในจุดมุ่งหมายที่แตกต่างกัน สมมติเราถามนายแพทย์คนหนึ่งว่าทำไมท่านจึงต้องการให้การกะประมาณจำนวนเปอร์เซนต์ของคนที่มีเลือดหมู่ O มีความคลาดเคลื่อนได้ 5 เปอร์เซนต์ ทำไมไม่เอา 4 เปอร์เซนต์ หรือ 6 เปอร์เซนต์ เขาอาจจะตอบว่าการแบ่งกลุ่มเลือดอาจนำมาใช้ในการจัดจำแนกคนออกตามหมู่เลือด เพื่อประโยชน์ทางการแพทย์บางอย่างเท่านั้น เขามีความมั่นใจว่าคนในหมู่บ้านนี้มีสัดส่วนของการจำแนกอยู่ระหว่าง 35 เปอร์เซนต์ ถึง 50 เปอร์เซนต์ ดังนั้น การที่มีความคลาดเคลื่อนในการกะประมาณ 5 เปอร์เซนต์ จึงไม่ทำให้การจัดจำแนกผิดพลาดไปจากความเป็นจริงเท่าใดนัก

ในการคำนวณหาขนาดของตัวอย่าง(n) นั้นจะต้องมีการกำหนดช่วงของความคลาดเคลื่อน(B)ของการกะประมาณ โดยช่วงความคลาดเคลื่อนการกะประมาณจะเท่ากับสองเท่าของค่าความคลาดเคลื่อนของการกะประมาณสัดส่วน(S) โดยเขียนเป็นสมการได้ดังนี้

$$B = 2S \tag{6.7}$$

โดยที่

$$S = \sqrt{\frac{N - n}{(n - 1)N}\ pq} \tag{6.8}$$

จะได้

$$B = 2\sqrt{\frac{N - n}{(n - 1)N}\ pq} \tag{6.9}$$

$$n = \frac{Npq}{(N - 1)D + pq} \tag{6.10}$$

เมื่อ

$$D = \frac{B^2}{N} \tag{6.11}$$

โปรดสังเกตว่าในทางปฏิบัตินั้นเรามักจะไม่ทราบค่า p เพราะฉะนั้นในการคำนวณบางทีก็ใช้ค่า p ซึ่งได้จากการสำรวจที่มีมาในอดีตหรือการสำรวจในลักษณะที่คล้ายกัน แต่ถ้าไม่ทราบจริง ๆ ก็อาจสมมติให้ p = 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.2 การประเมินความน่าเชื่อของข้อมูลด้วยการสุ่มตัวอย่าง

จากที่กล่าวแล้วว่าการประเมินความน่าเชื่อถือของข้อมูลตามความเป็นจริง (R_a) จะทำได้โดยการหาจำนวนรายการข้อมูลที่ผิดพลาดไปจากความเป็นจริง ซึ่งจะทำอย่างนั้นได้ต้องมีการตรวจสอบรายการข้อมูลทุก ๆ รายการในตารางข้อมูล และตรวจสอบทุก ๆ ตารางข้อมูลในระบบฐานข้อมูล โดยจะเห็นว่าเป็นการยุ่งยากและสิ้นเปลืองเวลาและแรงงานมากมายในการตรวจสอบกับระบบฐานข้อมูลที่มีจำนวนรายการข้อมูลมาก ๆ เพราะฉะนั้นในการตรวจสอบเราจะประยุกต์ใช้ทฤษฎีการสุ่มตัวอย่างที่กล่าวมาแล้วข้างต้น เพื่อสุ่มตัวอย่างบางรายการข้อมูลในแต่ละตารางข้อมูลเท่านั้นมาทำการตรวจสอบ โดยจะทำการตรวจสอบทุก ๆ ตารางข้อมูล โดยจะแยกแต่ละตารางข้อมูลออกจากกัน ซึ่งจะเห็นว่าถ้าเราทำเช่นนี้จะเกิดคำถามขึ้นมาว่า "แล้วเราจะสุ่มเอารายการข้อมูลออกมากี่รายการในแต่ละตารางข้อมูล เพื่อที่ทำให้เราสามารถกะประมาณจำนวนของรายการข้อมูลทั้งหมดที่เกิดความผิดพลาดได้ถูกต้องที่สุด" โดยคำตอบของคำถามนั้นจะขึ้นอยู่กับปัจจัยเพียง 2 ประการเท่านั้น คือ

1. เราจะต้องทราบถึงจำนวนรายการข้อมูลทั้งหมดที่มีอยู่ในแต่ละรายการ

2. เราต้องการให้ค่ากะประมาณของจำนวนรายการข้อมูลที่เกิดความผิดพลาดนั้นมีความคลาดเคลื่อนไปจากความเป็นจริงมากน้อยเพียงใด

ถ้าเราทราบถึงปัจจัยทั้งสองหัวข้อนี้เราก็สามารถที่จะคำนวณการกะประมาณของรายการข้อมูลที่เกิดความผิดพลาดทั้งหมดได้

6.3 การสร้างระบบผู้เชี่ยวชาญ

จากที่กล่าวมาแล้วข้างต้น ถ้าจะสร้างระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลลักษณะนี้จะต้องทราบถึงตัวแปรดังต่อไปนี้

N คือ จำนวนรายการข้อมูลทั้งหมด

B คือ ช่วงความคลาดเคลื่อนในการกะประมาณจำนวนรายการข้อมูลที่เกิดความผิดพลาดทั้งหมด

n คือ จำนวนรายการข้อมูลที่สุ่มมาเป็นตัวอย่าง โดยคำนวนได้จากสมการที่ 6.10

a คือ จำนวนรายการข้อมูลที่เกิดความผิดพลาดในตัวอย่าง

จะเห็นว่าจำนวนตัวแปรทั้งหมดที่กล่าวถึงนั้น จะมีตัวแปร 2 ตัวแปรที่ระบบผู้เชี่ยวชาญต้องการจากผู้ใช้ระบบ คือ B และ a ที่เป็นเช่นนี้ก็เพราะว่า B เป็นความต้องการของผู้ใช้ที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ให้จำนวนรายการข้อมูลที่เกิดความผิดพลาดที่ได้จากการกะประมาณคลาดเคลื่อนไปจากความเป็นจริงมากน้อยเพียงใด และ a เป็นจำนวนรายการข้อมูลที่เกิดความผิดพลาดในกลุ่มตัวอย่างที่ได้จากการตรวจสอบของผู้ใช้เอง เพราะฉะนั้นเมื่อระบบผู้เชี่ยวชาญทราบถึงตัวแปรทั้งสองนี้แล้วก็สามารถที่จะคำนวณหาจำนวนรายการข้อมูลที่เกิดความผิดพลาดทั้งหมดได้ดังสมการต่อไปนี้

$$N_f = N * \frac{a}{n} \qquad (6.12)$$

6.4 ขั้นตอนการทำงานของระบบผู้เชี่ยวชาญ

1) รับค่าความคลาดเคลื่อนในการกะประมาณ(B)จากผู้ใช้ ในกรณีที่ไม่ทราบช่วงความคลาดเคลื่อนจากผู้ใช้ จะกำหนดให้ช่วงความคลาดเคลื่อนเท่ากับ 0.05

2) ทำการหาจำนวนรายการข้อมูลทั้งหมดในแต่ละตารางข้อมูล

3) คำนวณหาจำนวนรายการข้อมูลที่เหมาะสมสำหรับจะสุ่มมาเป็นตัวอย่าง โดยแยกตามแต่ละตารางข้อมูล จากสมการที่ 6.10 ซึ่งในการคำนวณหาจำนวนรายการข้อมูลที่เหมาะสมของตัวอย่างจะใช้ค่าสัดส่วนของรายการข้อมูลที่ผิดพลาดต่อจำนวนรายการข้อมูลทั้งหมด(p)ในเท่ากับ 0.5 แต่ในการคำนวณหาจำนวนรายการข้อมูลที่เหมาะสมของตัวอย่างครั้งต่อไปจะใช้ค่า p ที่คำนวณได้จากการประเมินค่าก่อนหน้านั้น โดยที่ $p = N_f/N$

4) ทำการสุ่มตัวอย่างรายการข้อมูลในแต่ละตารางข้อมูล ตามจำนวนที่คำนวณได้ในข้อ 3 โดยในการสุ่มตัวอย่างนั้นจะเป็นการสุ่มตัวอย่างแบบไม่มีรายการข้อมูลที่ซ้ำกัน

5) เจ้าหน้าที่ที่เกี่ยวข้องกับระบบฐานข้อมูลนำรายการข้อมูลที่สุ่มได้ในข้อ 4 ไปตรวจสอบกับข้อมูลจริง และเก็บผลลัพธ์ที่เป็นจำนวนรายการที่ผิดพลาดไว้

6) รับผลที่เป็นจำนวนรายการข้อมูลที่เกิดความผิดพลาด(a)ในตัวอย่างที่นำไปตรวจสอบกับข้อมูลจริงโดยเจ้าหน้าที่ที่ทำการตรวจสอบในข้อ 5

7) คำนวณหาค่ากะประมาณจำนวนรายการข้อมูลที่เกิดความผิดพลาดทั้งหมด(N_f)ในแต่ละตารางข้อมูล โดยใช้สมการที่ 6.12

8) นำจำนวนรายการข้อมูลที่เกิดความผิดพลาดที่ได้ในแต่ละตารางข้อมูลมารวมกันเพื่อคำนวณหาค่าความน่าเชื่อถือของข้อมูล R_s จากสมการที่ 3.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.5 ผลการทดลอง

ในการทดลองจะทำการทดลองกับระบบฐานข้อมูลนักศึกษาโดยจะทำการทดลองกับตารางข้อมูลเพียงตารางเดียวคือตารางข้อมูล STUDENT_BIO1 ซึ่งการทดลองจะจัดแบ่งออกเป็น 4 กลุ่ม ในแต่ละกลุ่มจะมีการกำหนดจำนวนรายการข้อมูลทั้งหมดและช่วงความคาดเคลื่อนไว้ไม่เท่ากัน รวมถึงการจำลองรายการข้อมูลที่เกิดการผิดพลาดในลักษณะ R_u ขึ้นมาเพื่อนำไปคำนวณหาค่าความน่าเชื่อถือของข้อมูลและเปรียบเทียบกับค่าความน่าเชื่อถือของข้อมูลที่ได้จากการทดลอง โดยการทดลองในแต่ละกลุ่มนั้นจะแบ่งการทดลองออกเป็น 3 ครั้ง และนำค่าความน่าเชื่อถือของข้อมูลทั้ง 3 ครั้งมาหาค่าความน่าเชื่อถือของข้อมูลโดยเฉลี่ย หลังจากนั้นจึงคำนวณหาเปอร์เซ็นต์ความผิดพลาดระหว่างค่าความน่าเชื่อถือของข้อมูลตามความเป็นจริงกับค่าความน่าเชื่อถือของข้อมูลที่ได้จากการทดลอง

6.5.1 ผลการทดลองที่ได้

1) การทดลองกลุ่ม ก.

ในการทดลองกลุ่มนี้จะกำหนดจำนวนรายการข้อมูลทั้งหมดในตารางข้อมูล STUDENT_BIO1 ไว้เป็นจำนวน 1000 รายการ โดยมีการจำลองรายการข้อมูลที่เกิดการผิดพลาดในลักษณะ R_u ไว้ทั้งสิ้น 50 รายการ และกำหนดช่วงความคลาดเคลื่อนเท่ากับ 0.05 ซึ่งผลการทดลองที่ได้ดังตารางที่ 20

2) การทดลองกลุ่ม ข.

ในการทดลองกลุ่มนี้จะกำหนดจำนวนรายการข้อมูลทั้งหมดในตารางข้อมูล STUDENT_BIO1 ไว้เป็นจำนวน 1000 รายการ โดยมีการจำลองรายการข้อมูลที่เกิดการผิดพลาดในลักษณะ R_u ไว้ทั้งสิ้น 50 รายการ และกำหนดช่วงความคลาดเคลื่อนเท่ากับ 0.1 ซึ่งผลการทดลองที่ได้ดังตารางที่ 21

3) การทดลองกลุ่ม ค.

ในการทดลองกลุ่มนี้จะกำหนดจำนวนรายการข้อมูลทั้งหมดในตารางข้อมูล STUDENT_BIO1 ไว้เป็นจำนวน 2000 รายการ โดยมีการจำลองรายการข้อมูลที่เกิดการผิดพลาดในลักษณะ R_u ไว้ทั้งสิ้น 110 รายการ และกำหนดช่วงความคลาดเคลื่อนเท่ากับ 0.05 ซึ่งผลการทดลองที่ได้ดังตารางที่ 22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) การทดลองกลุ่ม ง.

ในการทดลองกลุ่มนี้จะกำหนดจำนวนรายการข้อมูลทั้งหมดในตารางข้อมูล STUDENT_BIO1 ไว้เป็นจำนวน 2000 รายการ โดยมีการจำลองรายการข้อมูลที่เกิดการผิดพลาดในลักษณะ R_a ไว้ทั้งสิ้น 110 รายการ และกำหนดช่วงความคลาดเคลื่อนเท่ากับ 0.1 ซึ่งผลการทดลองที่ได้ดังตารางที่ 23

ตารางที่ 20

แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ก. จากจำนวนรายการข้อมูลทั้งหมด 1000 รายการ ในช่วงความคลาดเคลื่อน 0.05

รายการที่ใช้ทดสอบ	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1. ค่าสัดส่วนของรายการข้อมูลที่ผิดพลาดต่อจำนวนรายการข้อมูลทั้งหมด(p)	0.5	0.06	0.06
2. จำนวนรายการข้อมูลที่ใช้สุ่มตัวอย่าง(n)	286	78	88
3. จำนวนรายการข้อมูลที่เกิดการผิดพลาดที่ได้จากการสุ่มตัวอย่าง(a)	16	5	3
4. จำนวนรายการข้อมูลที่เกิดการผิดพลาดทั้งหมดที่ได้จากการกะประมาณ(N_f)	56	64	34
5. ค่าความน่าเชื่อถือของข้อมูล(R_a)	94.40%	93.60%	96.58%
ค่าความน่าเชื่อถือของข้อมูลจากการทดลองทั้ง 3 ครั้ง		94.86%	
ค่าความน่าเชื่อถือของข้อมูลจากความเป็นจริงที่จำลองไว้		95%	
เปอร์เซ็นความผิดพลาด		0.154%	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 21

แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ข.
จากจำนวนรายการข้อมูลทั้งหมด 1000 รายการ ในช่วงความคลาดเคลื่อน 0.1

รายการที่ใช้ทดสอบ	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1. ค่าสัดส่วนของรายการข้อมูลที่ผิดพลาดต่อจำนวนรายการข้อมูลทั้งหมด(p)	0.5	0.05	0.05
2. จำนวนรายการข้อมูลที่ใช้สุ่มตัวอย่าง(n)	91	20	18
3. จำนวนรายการข้อมูลที่เกิดการผิดพลาดที่ได้จากการสุ่มตัวอย่าง(a)	5	1	1
4. จำนวนรายการข้อมูลที่เกิดการผิดพลาดทั้งหมดที่ได้จากการกะประมาณ(N_f)	55	49	54
5. ค่าความน่าเชื่อถือของข้อมูล(R_a)	94.50%	95.09%	94.55%
ค่าความน่าเชื่อถือของข้อมูลจากการทดลองทั้ง 3 ครั้ง		94.71%	
ค่าความน่าเชื่อถือของข้อมูลจากความเป็นจริงที่จำลองไว้		95%	
เปอร์เซ็นความผิดพลาด		0.30%	

ตารางที่ 22

แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_a ในกลุ่ม ค.
จากจำนวนรายการข้อมูลทั้งหมด 2000 รายการ ในช่วงความคลาดเคลื่อน 0.05

รายการที่ใช้ทดสอบ	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1. ค่าสัดส่วนของรายการข้อมูลที่ผิดพลาดต่อจำนวนรายการข้อมูลทั้งหมด(p)	0.5	0.06	0.06
2. จำนวนรายการข้อมูลที่ใช้สุ่มตัวอย่าง(n)	333	90	80
3. จำนวนรายการข้อมูลที่เกิดการผิดพลาดที่ได้จากการสุ่มตัวอย่าง(a)	21	5	4
4. จำนวนรายการข้อมูลที่เกิดการผิดพลาดทั้งหมดที่ได้จากการกะประมาณ(N_f)	126	111	99
5. ค่าความน่าเชื่อถือของข้อมูล(R_a)	93.70%	94.46%	95.03%
ค่าความน่าเชื่อถือของข้อมูลจากการทดลองทั้ง 3 ครั้ง		94.40%	
ค่าความน่าเชื่อถือของข้อมูลจากความเป็นจริงที่จำลองไว้		94.50%	
เปอร์เซ็นความผิดพลาด		0.11%	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 23

แสดงผลการทดลองการประเมินความน่าเชื่อถือของข้อมูลลักษณะ R_3 ในกลุ่ม ง. จากจำนวนรายการข้อมูลทั้งหมด 2000 รายการ ในช่วงความคลาดเคลื่อน 0.1

รายการที่ใช้ทดสอบ	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1. ค่าสัดส่วนของรายการข้อมูลที่ผิดพลาดต่อจำนวนรายการข้อมูลทั้งหมด(p)	0.5	0.06	0.09
2. จำนวนรายการข้อมูลที่ใช้สุ่มตัวอย่าง(n)	95	23	31
3. จำนวนรายการข้อมูลที่เกิดการผิดพลาดที่ได้จากการสุ่มตัวอย่าง(a)	6	2	1
4. จำนวนรายการข้อมูลที่เกิดการผิดพลาดทั้งหมดที่ได้จากการกะประมาณ(N_f)	126	171	65
5. ค่าความน่าเชื่อถือของข้อมูล(R_3)	93.70%	91.43%	96.76%
ค่าความน่าเชื่อถือของข้อมูลจากการทดลองทั้ง 3 ครั้ง		93.96%	
ค่าความน่าเชื่อถือของข้อมูลจากความเป็นจริงที่จำลองไว้		94.50%	
เปอร์เซ็นความผิดพลาด		0.566%	

6.5.2 สรุปผลการทดลอง

จากผลการทดลองทั้ง 4 กลุ่มสามารถสรุปได้ว่า ความถูกต้องของค่าความน่าเชื่อของข้อมูลที่คำนวณได้จากการกะประมาณจะขึ้นอยู่กับจำนวนรายการข้อมูลที่นำมาเป็นกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างที่มีจำนวนรายการข้อมูลมากก็จะทำให้การกะประมาณค่าความน่าเชื่อถือของข้อมูลมีความถูกต้องมากตามไปด้วย แต่ในทางปฏิบัตินั้นถ้าจำนวนรายการข้อมูลในกลุ่มตัวอย่างมีมากก็จะทำให้เสียเวลาและค่าใช้จ่ายในการสำรวจหาข้อมูลที่ผิดพลาดตามความเป็นจริง เพราะฉะนั้นในทางปฏิบัติจะต้องกำหนดจำนวนรายการข้อมูลที่ประกอบกันเป็นกลุ่มตัวอย่างให้มีจำนวนที่เหมาะสม ทั้งนี้ทั้งนั้นจำนวนรายการข้อมูลที่เป็นกลุ่มตัวอย่างจะขึ้นอยู่กับช่วงความคลาดเคลื่อน(B) แต่ถ้ากำหนดช่วงความคลาดเคลื่อนน้อย ๆ ก็จะทำให้มีรายการข้อมูลที่เป็นตัวอย่างมีจำนวนมาก แต่ในทางกลับกัน ถ้ากำหนดช่วงความคลาดเคลื่อนมาก ๆ ก็จะทำให้มีรายการข้อมูลที่เป็นตัวอย่างมีจำนวนน้อยซึ่งก็จะทำให้ค่าความน่าเชื่อถือของข้อมูลที่ได้จากการกะประมาณมีความถูกต้องลดลงด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อ้างอิง

[1] วิลาส วูวงศ์, บุญเจริญ ศิริเนาวกุล, ระบบผู้เชี่ยวชาญ (กรุงเทพมหานคร : สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, 2535).

[2] ครรชิต ไมตรี, ระบบผู้เชี่ยวชาญ (กรุงเทพมหานคร : สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2535).

[3] จรณิต แก้วกังวาล, การออกแบบและจัดการฐานข้อมูล (กรุงเทพมหานคร : บริษัท ซีเอ็ดยูเคชั่น จำกัด, 2537), หน้า 198-200.

[4] อนันต์ ศรีโสภา, เทคนิคการสุ่มตัวอย่าง (กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2534), หน้า 24-81.

[5] อำพล ธรรมเจริญ, ทฤษฎีความน่าจะเป็นและสถิติ (ชลบุรี : มหาวิทยาลัยศรีนครินทรวิโรฒ บางแสน).

[6] Nachman Agmon, Niv-Ahituv, "Assessing Data Reliability in an Information System", Journal of Management Information System, Vol. 4, No. 2, 1987, pp. 41-44.

[7] Robertson. A. G., Quality Control and Riliability (London : Thomas Nelson, 1971).

[8] Vaughn. R. C., Quality Control (Iowa : Iowa State University Press, 1974).

[9] Anderson. T.. , Lee. P. A., Fault Tolerance (Englewood Cliffs. N.J. : Prentice Hall, 1981).

[10] Jame P. Ignizio, Introduction to Expert Sytems (New York : McGraw-Hill, 1991).

[11] Joseph Giarratano. Ph.D, Gary Riley, Expert Systems Principles and Programming (Boston : PWS-KENT Publishing Company).

[12] Date, C. J., An Introduction to Database Systems , 6th Edition (United States of America : Addison-Wesley Publishing Company, 1995).

[13] Nijssen, G. M., Halpin, T. A., Conceptual Schema and Relational Database Design A Fact Oriented Approach (Australia : Prentice Hall of Australia Pty Ltd.).

[14] Viescas, Jonh L., Running Microsoft Access 2 : for Windows (Washington : Microsoft Press, 1994).

[15] Jamil, H.M. ; Sadri, F , "Recognizing Credible Experts in Inaccurate Database", Methodology for Intelligent Systems. 8th Internation Symposium, ISMIS'94 Proceeding pp. 46-55 (Germany : Springer-Verlag, 1994).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

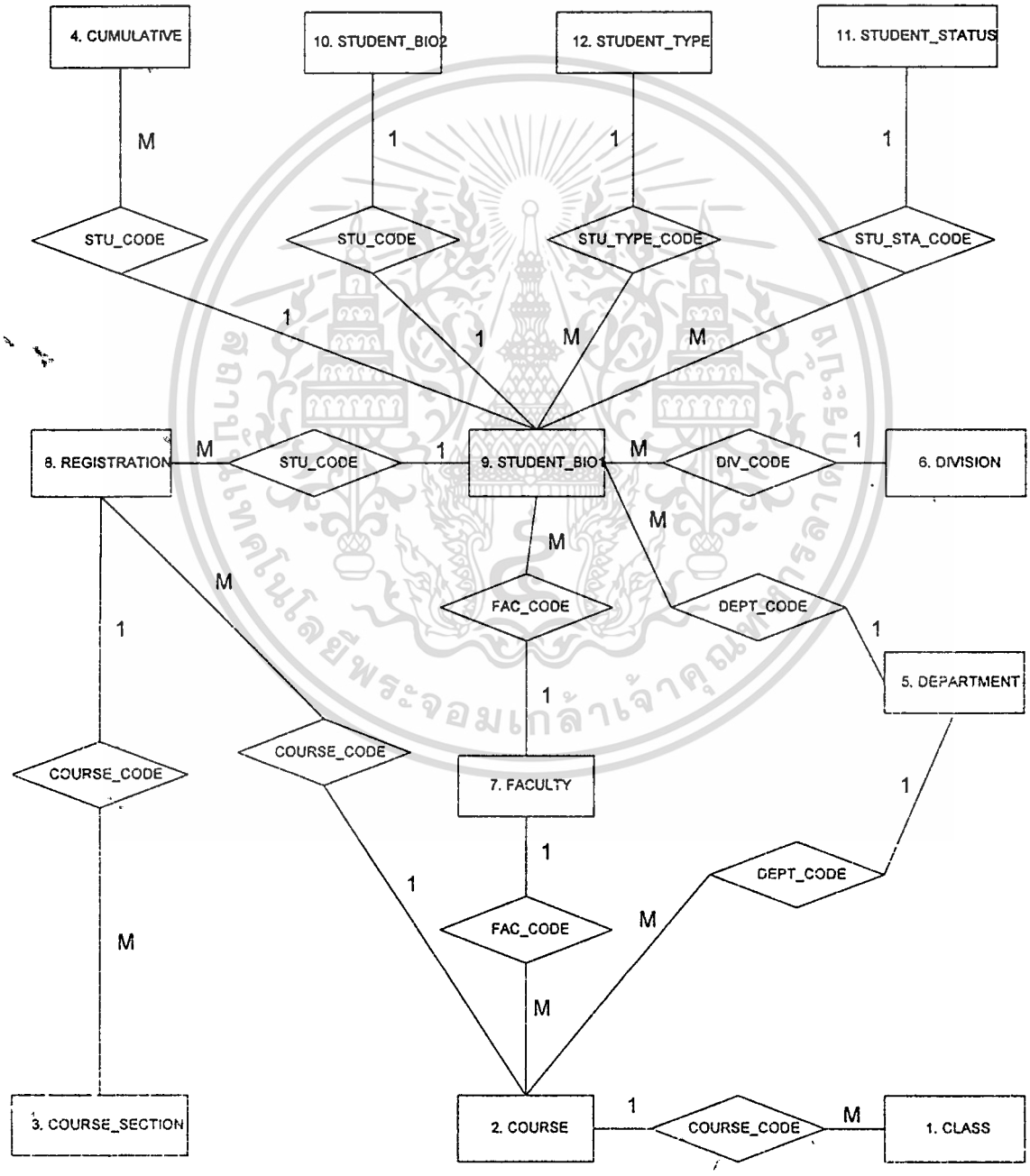
โครงสร้างฐานข้อมูลนักศึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

โครงสร้างฐานข้อมูลนักศึกษา

ENTITY/RELATIONSHIP DIAGRAM



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE DESCRIPTION

1. Table CLASS

เป็นตารางที่เก็บรายละเอียดของวิชาที่เปิดสอนอยู่ ใครสอนแต่ละวิชา สอนที่ไหน เริ่มเรียน/เลิกเรียนเมื่อไหร เพื่อใช้ในงานตารางสอน ตารางสอบ ตารางการใช้ห้อง

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสวิชา	COURSE_CODE	CHAR(07)
ตอน	SECTION	CHAR(02)
รหัสอาจารย์ผู้สอน	INSTRUCTOR_CODE	CHAR(06)
รหัสห้อง	ROOM_CODE	CHAR(04)
รหัสตึก	BUILDING_CODE	CHAR(02)
ปีการศึกษา	YEAR	CHAR(04)
ภาคการศึกษา	SEMESTER	CHAR(01)
วันเรียน	MEETING_DAY	CHAR(01)
เวลาเริ่มเรียน	MEETING_TIME_BEGIN	DATE
เวลาเลิกเรียน	MEETING_TIME_END	DATE
จำนวนนักศึกษา	CLASS_SIZE	NUMBER(3)

2. Table COURSE_SECTION

เป็นตารางที่ใช้ในการเก็บรายละเอียดของวิชาที่เปิดให้ลงทะเบียน และข้อกำหนดในการลงทะเบียน

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสวิชา	COURSE_CODE	CHAR(07)
ตอน	SECTION	CHAR(02)
ปีการศึกษา	YEAR	CHAR(04)
ภาคการศึกษา	SEMESTER	CHAR(01)
จำนวนผู้ลงเรียนได้มากที่สุด	SECTION_SIZE_MAX	NUMBER(4)
จำนวนคนน้อยสุดที่จะเปิด	SECTION_SIZE_MIN	NUMBER(4)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. Table COURSE

ใช้เก็บรายละเอียดวิชาต่าง ๆ ที่มีการเปิดสอน ใช้ในการลงทะเบียน งานตารางสอน

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสวิชา	COURSE_CODE	CHAR(07)
รหัสคณะ	FAC_CODE	CHAR(01)
รหัสภาควิชา	DEPT_CODE	CHAR(02)
ชื่อวิชา (ไทย)	COURSE_NAME_THAI	CHAR(40)
ชื่อวิชา (อังกฤษ)	COURSE_NAME_ENG	CHAR(30)
วันที่เริ่มใช้วิชา	COURSE_BEGIN_DATE	DATE
วันที่เลิกใช้วิชา	COURSE_END_DATE	DATE
จำนวนหน่วยกิต	COURSE_CRD	NUMBER(3)
หน่วยกิตบรรยาย	COURSE_CRD_LEC	NUMBER(3)
หน่วยกิตปฏิบัติ	COURSE_CRD_LAB	NUMBER(3)
จำนวนชั่วโมงบรรยาย	COURSE_HR_LEC	NUMBER(3)
จำนวนชั่วโมงปฏิบัติ	COURSE_HR_LAB	NUMBER(3)

4. Table DEPARTMENT

เป็นตารางที่เก็บรายละเอียดของภาควิชา เพื่อเป็นข้อมูลอ้างอิงกับรหัสภาควิชาในตารางอื่น

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสภาควิชา	DEPT_CODE	CHAR(02)
ชื่อภาควิชา (อังกฤษ)	DEPT_NAME_ENG	CHAR(40)
ชื่อภาควิชา (ภาษาไทย)	DEPT_NAME_THAI	CHAR(80)
วันที่เริ่มก่อตั้งภาควิชา	DEPT_BEGIN_DATE	DATE
วันที่ยกเลิกภาควิชา	DEPT_END_DATE	DATE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. Table CUMULATIVE

เป็นตารางข้อมูลที่เก็บจำนวนหน่วยกิตสะสมและเกรดการศึกษาของนักศึกษาในแต่ละเทอม

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสนักศึกษา	STU_CODE	CHAR(10)
ปีการศึกษา	YEAR	CHAR(04)
ภาคการศึกษา	SEMESTER	CHAR(01)
เกรดเฉลี่ยเทอมปัจจุบัน	INDEX_SEM	NUMBER(4)
เกรดเฉลี่ยสะสม	INDEX_CUM	NUMBER(4)
หน่วยกิตที่ลงทะเบียนเทอมปัจจุบัน	ATMP_SEM	NUMBER(4)
หน่วยกิตที่ลงทะเบียนสะสม	ATMP_CUM	NUMBER(4)
หน่วยกิตที่สอบได้เทอมปัจจุบัน	ERN_SEM	NUMBER(4)
หน่วยกิตที่สอบได้สะสม	ERN_CUM	NUMBER(4)

6. Table DIVISION

เป็นตารางที่เก็บรายละเอียดของสาขาวิชารหัสหลักสูตรที่เปิดสอน วันก่อตั้ง และสังกัด

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสสาขา	DIV_CODE	CHAR(03)
ชื่อสาขา (ไทย)	DIV_NAME_THAI	CHAR(40)
ชื่อสาขา (อังกฤษ)	DIV_NAME_ENG	CHAR(80)
รหัสหลักสูตร	CURR_CODE	CHAR(06)
วันที่เริ่มตั้งสาขา	DIV_BEGIN_DATE	DATE
วันที่ปิดสาขา	DIV_END_DATE	DATE
รหัสคณะ	FAC_CODE	CHAR(01)
รหัสภาควิชา	DEPT_CODE	CHAR(02)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. Table FACULTY

เป็นตารางที่เก็บรายละเอียดของคณะเพื่อเป็นข้อมูลอ้างอิงกับรหัสคณะในตารางอื่น ๆ

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสคณะ	FAC_CODE	CHAR(01)
ชื่อคณะ (ไทย)	FAC_NAME_THAI	CHAR(40)
ชื่อคณะ (อังกฤษ)	FAC_NAME_ENG	CHAR(80)

8. Table REGISTRATION

เป็นตารางเก็บข้อมูลการลงทะเบียนของนักศึกษา แยกตามรายวิชาที่ลงทะเบียนรวมทั้งจัดเก็บระดับคะแนนที่ได้เอาไว้ด้วย

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสประจำตัวนักศึกษา	STU_CODE	CHAR(10)
รหัสวิชา	COURSE_CODE	CHAR(07)
ตอน	SECTION	CHAR(02)
ปีการศึกษา	YEAR	CHAR(04)
ภาคการศึกษา	SEMESTER	CHAR(01)
ระดับคะแนนที่ได้	GRADE	CHAR(02)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. Table STUDENT_STATUS

เป็นตารางที่ใช้ในการเก็บรายละเอียดของสถานภาพการศึกษาของนักศึกษา

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสสถานภาพนักศึกษา	STU_ST_CODE	CHAR(01)
คำอธิบายสถานภาพ	STU_ST_DESC	CHAR(50)

10. Table STUDENT_BIO1

เป็นตารางข้อมูลที่ใช้เก็บข้อมูลรายละเอียดของนักศึกษาส่วนที่มีการเรียกใช้บ่อย

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสประจำตัวนักศึกษา	STU_CODE	CHAR(10)
คำนำหน้าชื่อ	STU_PRE_NAME	CHAR(02)
ชื่อ (ไทย)	STU_FIRST_NAME_THAI	CHAR(20)
นามสกุล (ไทย)	STU_LAST_NAME_THAI	CHAR(30)
ชื่อ (อังกฤษ)	STU_FIRST_NAME_ENG	CHAR(20)
นามสกุล (อังกฤษ)	STU_LAST_NAME_ENG	CHAR(30)
รหัสเพศ	SEX	CHAR(01)
รหัสระดับที่ศึกษา	LEVEL_CODE	CHAR(01)
รหัสหลักสูตร	CURR_CODE	CHAR(06)
รหัสคณะ	FAC_CODE	CHAR(01)
รหัสภาควิชา	DEPT_CODE	CHAR(02)
รหัสสาขาวิชา	DIV_CODE	CHAR(03)
ห้อง	CLASS	CHAR(02)
เกรดเฉลี่ยสะสม	AVERAGE_GPA	CHAR(04)
ประเภทนักศึกษา	STU_TYPE	CHAR(01)
สถานภาพนักศึกษา	STU_STATUS	CHAR(01)
รหัสอาจารย์ที่ปรึกษา	INSTRUCTOR_CODE	CHAR(06)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11. Table STUDENT_BIO2

เป็นตารางข้อมูลที่ใช้เก็บรายละเอียดข้อมูลของนักศึกษา(นอกเหนือจาก STUDENT_BIO1 เพราะมีอัตราการใช้น้อยกว่า)

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสประจำตัวนักศึกษา	STU_CODE	CHAR(10)
รหัสเชื้อชาติ	RACE	CHAR(01)
รหัสสัญชาติ	NATIONALITY	CHAR(01)
รหัสศาสนา	RELIGION	CHAR(01)
วัน/เดือน/ปี ที่เกิด	BIRTH_DATE	DATE
รหัสจังหวัดที่เกิด	BRON_PRO_CODE	CHAR(02)
ส่วนสูง (เซนติเมตร)	HIGHT	REAL
น้ำหนัก (กิโลกรัม)	WEIGTH	REAL
รหัสกลุ่มเลือด	BLOOD_GROUP_CODE	CHAR(01)
ที่อยู่ปัจจุบัน	ADDRESS	CHAR(80)
ชื่อบิดา	F_FIRST_NAME	CHAR(20)
นามสกุลบิดา	F_LAST_NAME	CHAR(30)
ชื่อมารดา	M_FIRST_NAME	CHAR(20)
นามสกุลมารดา	M_LAST_NAME	CHAR(30)
ชื่อผู้ปกครอง	G_FIRST_NAME	CHAR(20)
นามสกุลผู้ปกครอง	G_LAST_NAME	CHAR(30)
รหัสอาชีพบิดา	F_OCCUPATION_CODE	CHAR(02)
รหัสอาชีพมารดา	M_OCCUPATION_CODE	CHAR(02)
รหัสอัตราเงินเดือนบิดา	F_INCOME_CODE	CHAR(02)
รหัสอัตราเงินเดือนมารดา	M_INCOME_CODE	CHAR(02)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12. Table STUDENT_TYPE

เป็นตารางข้อมูลที่ใช้เก็บรายละเอียดประเภทของนักศึกษา

FIELD DESCRIPTION		
THAI	ENGLISH	TYPE
รหัสประเภทนักศึกษา	STU_TYPE_CODE	CHAR(1)
รายละเอียดประเภทนักศึกษา	STU_TYPE_DESC	CHAR(60)
ตัวย่อประเภทนักศึกษา	STU_TYPE_SHT	CHAR(2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

บทความที่ส่งตีพิมพ์ในวารสาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมาคมคอมพิวเตอร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์
THE COMPUTER ASSOCIATION OF THAILAND UNDER THE ROYAL PATRONAGE OF HIS MAJESTY THE KING

พญาไท พลาซ่า ชั้น 6 128 ถนนพญาไท เขตราชเทวี กรุงเทพฯ 10400 โทร. 216-5860-1 โทรสาร 216-5847
PAYATHAI PLAZA 6TH FLOOR, 128 PHYATHAI ROAD, RAJTAVEE, BANGKOK 10400 TEL. 216-5860-1 FAX. 216-5847

ที่ สคพท. 1/103

21 กันยายน 2538

เรื่อง รับรองบทความ

เรียน คุณพีระศักดิ์ เสรีกุล
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ตามที่ท่านได้เสนอบทความ เรื่อง "ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลในระบบสารสนเทศ (An Expert System for Assessing Data Reliability in an Information Systems)" นั้น ทางบรรณาธิการวารสารคอมพิวเตอร์ได้พิจารณาแล้วเห็นสมควรให้บทความของท่าน ได้ลงตีพิมพ์เพื่อเผยแพร่ใน "วารสารคอมพิวเตอร์"

จึงเรียนมาเพื่อทราบ

ขอแสดงความนับถือ

(นายสำนวน หิรัญวงษ์)
บรรณาธิการ วารสารคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบผู้เชี่ยวชาญเพื่อการประเมินความน่าเชื่อถือของข้อมูลในระบบสารสนเทศ

An Expert System for Assessing Data Reliability in an Information Systems

พีระศักดิ์ เสรีกุล* ดร.เอื้อน ปิ่นเงิน**

บทคัดย่อ

สารสนเทศที่มีคุณค่า จะต้องได้มาจากข้อมูลที่เชื่อถือได้ อย่างไรก็ตามวิธีการประเมินค่าความน่าเชื่อถือของข้อมูล ก็ยังมิได้รับการยอมรับกันอย่างกว้างขวางในระบบสารสนเทศ บทความฉบับนี้จะกล่าวถึง การนำเอาการประเมินความน่าเชื่อถือที่ใช้อยู่ในระบบการควบคุมคุณภาพ(Quality Control) มาประยุกต์ใช้กับระบบสารสนเทศ โดยจะกล่าวถึง วิธีการในการประเมินความน่าเชื่อถือของข้อมูล 3 ลักษณะ คือ Internal reliability ความน่าเชื่อถือของข้อมูลที่เป็นไปตามลักษณะพื้นฐานที่ยอมรับกันโดยทั่วไป ของรายการข้อมูลแต่ละรายการ Relative reliability ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความสัมพันธ์ และข้อกำหนดของระบบฐานข้อมูล Absolute reliability ความน่าเชื่อถือของข้อมูลเมื่อเปรียบเทียบกับข้อมูลตามความเป็นจริง

จากนั้นจะกล่าวถึง การนำเอาระบบผู้เชี่ยวชาญ (Expert system) มาประยุกต์ใช้ในการประเมินความน่าเชื่อถือ โดยจะอ้างอิงตามลักษณะการประเมินความน่าเชื่อถือทั้ง 3 ลักษณะที่กล่าวมาแล้ว

ABSTRACT

Information is valuable if it derives from reliable data. However measurements for data reliability have not been widely established in the area of information systems (IS). This paper attempts to draw some concepts of reliability from the field of quality control and to apply them to IS. The paper develops three measurements for data reliability. *Internal reliability* : Reliability whose assessment is based on commonly accepted criteria about the characteristics of the data items. *Relative reliability* :Reliability of the data in view of the user requirements. *Absolute reliability* : Reliability which is measured by performing direct comparisons between the content of the database and reality.

Also saying in this paper is applying Expert System in evaluating the reliability referring to the three measurements.

---

* นักศึกษาปริญญาโท สาขาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

** อาจารย์ประจำ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1. บทนำ

ปัจจุบันนี้ระบบสารสนเทศได้เข้ามามีส่วนสนับสนุนให้การปฏิบัติงานในระบบต่าง ๆ ให้มีประสิทธิภาพมากยิ่งขึ้น ไม่ว่าจะเป็นการปฏิบัติงานในภาครัฐ ภาคเอกชน แม้กระทั่งในภาคการศึกษา โดยเฉพาะอย่างยิ่ง ในองค์กรที่ใช้ระบบสารสนเทศที่ได้มาจากการประมวลผลด้วยคอมพิวเตอร์ด้วยแล้ว จะเห็นถึงความสำคัญของระบบสารสนเทศเป็นอย่างสูง หรือจะกล่าวได้ว่าองค์กรที่จะประสบความสำเร็จจะต้องประกอบไปด้วยระบบสารสนเทศที่ถูกต้องสมบูรณ์และมีความน่าเชื่อถือสูง เพราะฉะนั้นการที่สามารถประเมินความน่าเชื่อถือของข้อมูลในระบบสารสนเทศในขณะใดขณะหนึ่งได้นั้น ก็จะทำให้ทราบว่าสมควรที่จะทำการตรวจสอบ และปรับปรุงระบบสารสนเทศที่มีอยู่หรือไม่ เพื่อทำให้ได้มาซึ่งระบบสารสนเทศที่ถูกต้องและสมบูรณ์ที่สุด

2. การประเมินหาค่าความน่าเชื่อถือ

การศึกษาเกี่ยวกับความน่าเชื่อถือ ได้นำเอาทฤษฎีการหาความน่าเชื่อถือ (Reliability) ที่มีการใช้กันในด้านของการควบคุมคุณภาพ (Quality control) มาประยุกต์ใช้ โดยในระบบการควบคุมคุณภาพนั้นได้กล่าวถึงความน่าเชื่อถือในลักษณะของระยะเวลาระหว่างการผิดพลาด (Mean time between failures :MTBF) และ ระยะเวลาในการผิดพลาด (Mean Time to failures :MTTF) แต่ในการนำเอาทฤษฎีการหาความน่าเชื่อถือ มาประยุกต์ใช้กับระบบสารสนเทศนั้น จะต้องมีการเปลี่ยนแปลงความน่าเชื่อถือในลักษณะของระยะเวลาให้เป็นจำนวนรายการข้อมูล ตัวอย่างเช่น ระยะเวลาระหว่างการผิดพลาดก็ประยุกต์ใช้เป็นจำนวนรายการที่ใช้ทดสอบความผิดพลาด และ ระยะเวลาในการผิดพลาดประยุกต์เป็นจำนวนรายการที่เกิดการผิดพลาด ซึ่งสามารถสรุปออกเป็นตัวแปรและสมการได้ดังนี้

ให้ Nt คือ จำนวนรายการที่ใช้ทดสอบ (Items tested) ในระยะเวลา t

Nf คือ จำนวนรายการที่เกิดการผิดพลาด และ

Pf คือ ค่าความน่าจะเป็นในการผิดพลาดซึ่งสามารถหาได้จากสมการ

$$Pf = Nf/Nt \quad \text{และ}$$

$$R = (1 - Pf) \times 100$$

เพราะฉะนั้นจะได้

$$R = (1 - Nf/Nt) \times 100$$

เมื่อค่า R คือ ค่าความน่าเชื่อถือ

ความน่าเชื่อถือของข้อมูลในระบบสารสนเทศนั้น สามารถจำแนกออกได้เป็น 3 ลักษณะ ซึ่งในแต่ลักษณะจะมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1 ความน่าเชื่อถือของข้อมูลในแต่ละรายการข้อมูล (Internal reliability)

เป็นค่าความน่าเชื่อถือของข้อมูล ที่เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการ โดยที่ลักษณะพื้นฐานของรายการข้อมูลนี้จะเป็นมาตรฐานที่ยอมรับกันโดยทั่ว ๆ ไป

สัญลักษณ์ : Ri

ตัวอย่าง : รายการข้อมูลจำนวนนักศึกษา ในฐานข้อมูลทะเบียนนักศึกษาจะต้องเป็นค่าบวก จะเป็นลบไม่ได้ หรือรายการข้อมูลที่เป็นวันที่จะต้องมีจำนวนวันไม่เกิน 31 และจำนวนเดือนไม่เกิน 12 เป็นต้น

อธิบาย : Ri เป็นค่าความน่าเชื่อถือที่ประเมินได้โดยการใช้ลักษณะพื้นฐานของ ข้อมูลแต่ละรายการเป็นตัวประเมินหา โดยที่ลักษณะพื้นฐานดังกล่าวจะขึ้นอยู่กับ ข้อมูลแต่ละชนิดแต่ละรายการ และเป็นลักษณะที่บุคคลทั่วไปสามารถวินิจฉัยไปในแนวทางเดียวกันได้

2.2 ความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Relative reliability)

เป็นค่าความน่าเชื่อถือที่ประเมินได้จากการเปรียบเทียบข้อมูลว่า เป็นจริงตามความสัมพันธ์และเงื่อนไขของระบบฐานข้อมูลที่ได้กำหนดไว้หรือไม่

สัญลักษณ์ : Rr

ตัวอย่าง : ถ้าในฐานข้อมูลทะเบียนนักศึกษา ได้มีการกำหนดเงื่อนไขว่า จะต้องไม่มีข้อมูลของนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองเทอมติดต่อกัน ฉะนั้นถ้าฐานข้อมูลมีความน่าเชื่อถือลักษณะนี้ ก็แสดงว่ารายการข้อมูลของนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองเทอมติดต่อกันเป็นรายการข้อมูลที่ผิด

อธิบาย : Rr เป็นค่าความน่าเชื่อถือที่ได้จากการเปรียบเทียบข้อมูลกับข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล เพราะฉะนั้นในการประเมินความน่าเชื่อถือขั้นตอนนี้ จะต้องประเมินว่าข้อมูลทั้งหมดที่มีอยู่นั้น เป็นไปตามข้อกำหนดหรือเงื่อนไขที่มีอยู่หรือไม่

2.3 ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความเป็นจริง (Absolute reliability)

เป็นค่าความน่าเชื่อถือของข้อมูลที่ได้จากการเปรียบเทียบข้อมูลที่มีอยู่ในฐานข้อมูลกับข้อมูลตามความเป็นจริง

สัญลักษณ์ : Ra

ตัวอย่าง : รายการข้อมูลรหัสและชื่อนักศึกษา ในฐานข้อมูลนักศึกษา จะต้องตรงกับรหัสและชื่อนักศึกษาที่มีอยู่จริง

อธิบาย : Ra เป็นค่าความน่าเชื่อถือที่ได้จากการสังเกตเห็น (Observation) หรือกล่าวอีกนัย คือค่าความน่าเชื่อถือที่ได้จาก การดูรายการข้อมูลทุก ๆ ตัว และนำไปเปรียบเทียบกับข้อมูลที่มีอยู่จริง ซึ่งจะเห็นได้ว่าค่า Ra หาได้ยากมาก แต่เมื่อสามารถหาได้แล้วจะเป็นเครื่องยืนยันได้อย่างดีว่าระบบฐานข้อมูลหรือระบบสารสนเทศนั้น ๆ มีความน่าเชื่อถือได้มากน้อยเพียงใด

จากที่กล่าวมาทั้งหมดข้างต้น จะสังเกตได้ว่าค่า Rr จะหาได้ยากกว่าค่า Ri และค่า Ra จะหาได้ยากที่สุดในทั้ง 3 ลักษณะ แต่เมื่อหาค่าทั้ง 3 ลักษณะออกมาได้แล้ว ค่า Ra จะเป็นค่าที่เป็นตัวบ่งบอกถึงความน่าเชื่อถือของข้อมูลได้ดีที่สุด

ฐานข้อมูลโดยทั่ว ๆ ไป เมื่อนำมาหาค่าความน่าเชื่อถือ ทั้ง 3 ลักษณะนี้ จะได้ความสัมพันธ์กันดังนี้ คือ

Ra <= Rr <= Ri

ที่เป็นดังนี้ก็เพราะว่า ในรายการข้อมูลใดก็ตามที่มีความถูกต้องในลักษณะของ Ri แต่ไม่ได้หมายความว่ารายการข้อมูลตัวนั้นจะถูกต้องในลักษณะของ Rr และ Ra เช่น ในฐานข้อมูลพนักงานของบริษัทแห่งหนึ่งมีรายการข้อมูลเงินเดือนของ นาย ก. คือ 5,000 บาท โดยที่เงินเดือน 5,000 นั้น ถ้ามองในลักษณะ ของ Ri แล้วถือว่าเป็นข้อมูลที่ถูกต้อง แต่ถ้ามองในลักษณะของ Rr และ Ra แล้วข้อมูลเงินเดือน 5,000 นี้ อาจจะเป็นข้อมูลที่ผิดพลาดก็ได้ เพราะข้อมูลเงินเดือนจริง ๆ ของนาย ก. อาจจะไม่ใช่ 5,000 ก็ได้

ในทางตรงกันข้ามถ้าข้อมูลเงินเดือน เป็นข้อมูลที่ถูกต้องในลักษณะของ Ra ก็หมายความว่าข้อมูลเงินเดือนของนาย ก. จะต้องเท่ากับ 5,000 บาทจริง ๆ หรือสรุปได้ว่าข้อมูลเงินเดือนของนาย ก. จะถูกต้องในลักษณะของ Ri และ Rr ด้วยนั่นเอง

ในความเป็นจริงนั้น จะต้องพยายามหาวิธีการที่จะประเมินหาความน่าเชื่อถือของระบบสารสนเทศ หรือ ระบบฐานข้อมูลให้ถูกต้อง หรือไม่ก็ให้ใกล้เคียงกับความเป็นจริงมากที่สุด ซึ่งสามารถทำได้ โดยอาศัยอ้างอิงกับลักษณะการประเมินค่าความเชื่อถือที่ได้กล่าวมาข้างต้นแล้ว โดยจะไปเน้นหนักในลักษณะการประเมินหาค่าความน่าเชื่อของข้อมูลที่สัมพันธ์กันตามข้อกำหนด หรือเงื่อนไขของฐานข้อมูล (Rr) และลักษณะการประเมินหาค่าความน่าเชื่อของข้อมูลตามความเป็นจริง (Ra) เพราะรูปแบบการประเมินหาค่าความน่าเชื่อถือของข้อมูลแต่ละรายการ (Ri) จะเป็นค่าความน่าเชื่อถือที่มีได้สูงมากอยู่แล้ว เพราะในระบบฐานข้อมูลโดยทั่ว ๆ ไปนั้นจะมีระบบป้องกันความผิดพลาดของข้อมูลประเภทนี้อยู่แล้ว

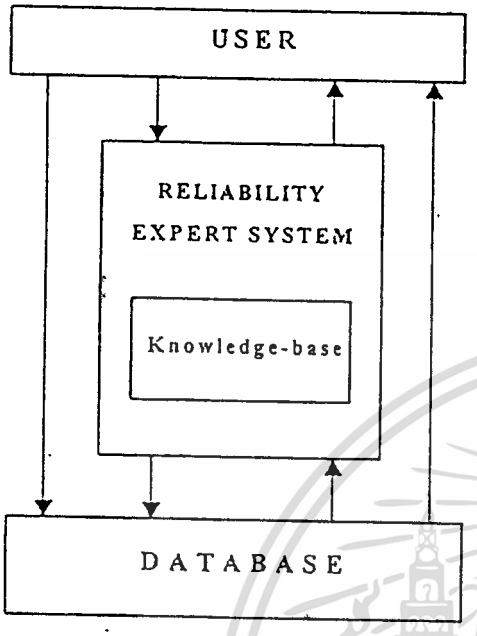
อย่างไรก็ตาม ในการประเมินหาค่าความน่าเชื่อถือของข้อมูลนั้นจะต้องทำทั้ง 3 ลักษณะที่กล่าวมาแล้วข้างต้น จึงจะทำให้การประเมินความน่าเชื่อถือนั้นถูกต้องสมบูรณ์

3. การประเมินความน่าเชื่อถือโดยใช้ระบบผู้เชี่ยวชาญ

ในการประเมินค่าความน่าเชื่อถือของระบบฐานข้อมูลนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เราสามารถนำเอาระบบผู้เชี่ยวชาญเข้ามาประยุกต์ใช้ได้ โดยการสร้างระบบการประเมินค่าความน่าเชื่อถือของข้อมูลที่เป็นระบบผู้เชี่ยวชาญ ซึ่งจะมีรูปแบบการทำงานแสดงในผังงาน ดังรูปที่ 1



รูปที่ 1 แสดงผังการทำงานของระบบการประเมินความน่าเชื่อถือของฐานข้อมูล

จากผังงานในรูปที่ 1 เมื่อผู้ใช้ (User) ต้องการที่จะทราบถึงความน่าเชื่อถือของฐานข้อมูล ก็สามารถที่จะสั่งให้ส่วนที่ทำหน้าที่ประเมินหาความน่าเชื่อถือทำการประเมินหา อาจจะกระทำเมื่อเวลาที่ต้องการจะทราบถึงความน่าเชื่อถือ หรือจะให้กระทำทุก ๆ สัปดาห์ หรือทุก ๆ เดือนก็ได้แล้วแต่ความเหมาะสม แต่กรณีที่ทำการป้อนข้อมูลเข้าฐานข้อมูล ก็จะทำการส่งผ่านข้อมูลจากผู้ใช้ไปสู่ระบบฐานข้อมูลได้โดยตรง โดยไม่ต้องผ่านส่วนของการประเมินความน่าเชื่อถือ ส่วนระบบผู้เชี่ยวชาญนั้นจะมีส่วนประกอบที่เรียกว่า ฐานความรู้ (Knowledge-base) ซึ่งในฐานความรู้นี้ จะเป็นส่วนที่ใช้เก็บความรู้เบื้องต้นในรูปของ ข้อเท็จจริง (Fact) และ ส่วนที่เก็บข้อกำหนดหรือเงื่อนไข (Rule-base) ที่ระบบผู้เชี่ยวชาญจะนำไปใช้ในกลไกการวินิจฉัย

โดยทั่ว ๆ ไปการที่จะสร้างระบบผู้เชี่ยวชาญได้นั้น ส่วนสำคัญก็คือ จะต้องสร้างข้อกำหนดกฎเกณฑ์ของระบบหรือที่เรียกว่า Rule-base ให้ได้เสียก่อน ในการสร้าง Rule-base ขึ้นมานั้น จะต้องมีข้อมูลที่เป็นข้อเท็จจริง (Fact) เพื่อเอาไว้ให้เป็นที่สำหรับอ้างอิงของ Rule-base ที่สร้างขึ้น ในที่นี้ข้อเท็จจริงก็คือ ข้อมูลที่มีอยู่ในฐานข้อมูลนั้นเอง

ในทางปฏิบัติจะทำการสร้าง Rule-base จากวิธีการประเมินความน่าเชื่อถือของข้อมูลทั้ง 3 ลักษณะที่กล่าวมาข้างต้น โดยมีแนวทางในการสร้างดังนี้

3.1 การสร้าง Rule-base สำหรับการประเมินค่าความน่าเชื่อถือของข้อมูลในแต่ละรายการข้อมูล (Rule-base for Ri)

ดังที่ได้อธิบายมาตอนต้นแล้วว่าการประเมินค่าความน่าเชื่อถือลักษณะนี้ ทำได้โดยตรวจสอบว่ารายการข้อมูลทุก ๆ รายการในฐานข้อมูล เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการหรือไม่ เมื่อทราบเช่นนี้แล้วจะต้องสร้างกฎ เพื่อตรวจสอบว่าข้อมูลในแต่ละรายการข้อมูลเป็นไปตามลักษณะพื้นฐานหรือไม่ ถ้าไม่ถูกต้องตามลักษณะพื้นฐาน ก็จะกำหนดให้ค่ากับตัวแปร Nf (จำนวนรายการที่ผิดพลาด) เพิ่มขึ้น 1

ตัวอย่างของ กฎการประเมินค่าความเชื่อถือลักษณะนี้ เช่น

IF 'STUDENT_QUANTITY' < 0 THEN Nf = Nf + 1
IF 'DATE' > 31 THEN Nf = Nf+ 1

จากตัวอย่างของกฎนี้ ถ้าข้อมูลในฐานข้อมูลในแต่ละระเบียนมีรายการข้อมูลคือ STUDENT_QUANTITY หรือ DATE อันใดอันหนึ่งเป็นจริงตามที่กำหนด(ถือว่าเป็นข้อมูลที่ผิดพลาด) จะให้ค่า Nf เพิ่มขึ้น 1 ถ้าเป็นไปตามกฎทั้งสอง ค่า Nf ก็จะเพิ่มขึ้น 2 เป็นต้น

3.2 การสร้าง Rule-base สำหรับการประเมินค่าความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Rule-base for Rr)

การประเมินค่าความน่าเชื่อถือลักษณะนี้จะประเมินจากการทดสอบว่าข้อมูลในฐานข้อมูลนั้น เป็นไปตามความสัมพันธ์และเงื่อนไขที่ระบบฐานข้อมูลกำหนดไว้หรือไม่ เพราะฉะนั้นในการสร้าง Rule-base จะต้องสร้างกฎที่สอดคล้องกับความสัมพันธ์และเงื่อนไขของระบบฐานข้อมูล เช่นตารางในรูปที่ 2

STUDENT

Stu_code	Year	Semeter	GPA	GPA_last
3403247	92	2	2.54	2.80
3404255	92	2	1.89	1.90
3405263	92	2	3.01	2.89
3406456	92	2	2.56	1.98

รูปที่ 2 แสดงตัวอย่างข้อมูล GPA ของนักศึกษาที่ผิดพลาด

ตัวอย่างในรูปที่ 2 ถ้ามีเงื่อนไข 'จะต้องไม่มีข้อมูลของนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองภาคการศึกษาติดต่อกัน' อยู่ในฐานข้อมูล ในทางปฏิบัติจะต้องพยายามสร้างกฎเพื่อตรวจสอบว่ามีนักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.00 สองภาคการศึกษาติดต่อกันหรือไม่ ถ้ามีก็จะให้ Nf เพิ่มขึ้น 1 โดยสามารถเขียนเงื่อนไขได้นี้

IF (GPA < 2.00) AND (GPA_last < 2.00)
THEN Nf = Nf + 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการสร้าง Rule-base นั้นจะต้องสร้างกฎตามความสัมพันธ์ของรายการข้อมูลทุกความสัมพันธ์และทุกเงื่อนไขหรือข้อกำหนดของฐานข้อมูลที่มีอยู่ เมื่อได้กฎแล้วจะนำเอาข้อมูลทุก ๆ ระเบียนในฐานข้อมูลมาเปรียบเทียบกับกฎที่สร้างขึ้นทั้งหมด

3.3 การสร้าง Rule-base สำหรับการประเมินค่าความน่าเชื่อถือของข้อมูลตามความเป็นจริง (Rule-base for Ra)

ในการศึกษาพบว่าวิธีที่จะประเมินค่าความน่าเชื่อถือลักษณะนี้จะทำได้ยากมาก เพราะจะต้องเอารายการข้อมูลทั้งหมดในฐานข้อมูลมาตรวจสอบกับข้อมูลที่มีอยู่จริง โดยในการปฏิบัตินั้นถ้าฐานข้อมูลที่มีขนาดใหญ่มาก ๆ จะทำได้เพียงสุ่มเอาเฉพาะรายการข้อมูลที่มีความสำคัญเท่านั้น หรืออีกทางหนึ่งทำได้โดยการนำผลสรุปหรือผลรวมในรายการข้อมูลที่เป็นชนิดตัวเลข(Numeric) ไปเปรียบเทียบกับข้อมูลจริงที่สามารถหามาได้

ตัวอย่างเช่น ในรูปที่ 3 ผลรวมของค่าหน่วยกิตทั้งหมดที่ได้จาก การเอาผลรวมของหน่วยกิตภาคทฤษฎีของนักศึกษาทั้งหมด (Sum of Credit_Lec) คูณกับ ค่าหน่วยกิตภาคทฤษฎีแต่ละหน่วยและนำมารวมกับผลรวมของหน่วยกิตภาคปฏิบัติของนักศึกษาทั้งหมด

CREDIT

Stu_code	Year	Semeter	Credit_Lec	Credit_Lab
3562358	94	2	12	8
3562812	94	2	10	8
3563584	94	2	15	3
3562545	94	2	12	6
:	:	:	:	:

รูปที่ 3 แสดงข้อมูลจำนวนหน่วยกิตทั้งภาคทฤษฎีและปฏิบัติของนักศึกษาที่ลงทะเบียนเรียน

(Sum of Credit_Lab) คูณกับค่าหน่วยกิตแต่ละหน่วย จะต้องเท่ากับจำนวนเงินค่าหน่วยกิตทั้งหมด ที่ปรากฏอยู่ในใบเสร็จรับเงิน

จากตัวอย่างข้างต้น สามารถเขียนเป็นกฎ โดยให้ Credit_Cost_Fact เป็นค่าหน่วยกิตของนักศึกษาแต่ละคนที่ปรากฏอยู่ในใบเสร็จรับเงิน ได้ดังนี้

```
IF  ((Sum of Credit_Lec  *  Credit_Lec_Cost) +
     (Sum of Credit_Lab  *  Credit_Lab_Cost) <>
      Credit_Cost_Fact)
THEN   Nf = Nf + 1
```

จะเห็นว่าในการสร้าง Rule-base จากลักษณะการประเมินความน่าเชื่อถือทั้ง 3 ลักษณะที่กล่าวมาแล้วข้างต้นนั้น การสร้าง Rule-base ในลักษณะของ Ri และ Rr สามารถเกิดเป็นรูปธรรมได้มากกว่าการประเมินความน่าเชื่อถือในลักษณะ Ra เพราะการสร้าง Rule-base ในลักษณะ Ra ถ้าฐานข้อมูลมีขนาดใหญ่และมีหลายฟิลด์ในแต่ละระเบียนจะไม่สามารถสร้างกฎได้ครอบคลุมรายการข้อมูลทุก ๆ รายการได้ ตัวอย่างเช่น ไม่สามารถสร้างกฎเพื่อตรวจสอบข้อมูลชื่อนักศึกษาทุกคนในฐานข้อมูลว่าถูกต้องหรือไม่ ได้ เพราะฉะนั้น Rule-base จะเป็นกฎที่ได้จากลักษณะการประเมินค่าความน่าเชื่อถือแบบ Ri และ Rr เป็นส่วนใหญ่

อย่างไรก็ตามเมื่อได้ Rule-base ที่ครอบคลุมทุก ๆ เงื่อนไขแล้ว ในการประมวลผลจะนำเอาข้อมูลทุก ๆ ระเบียน ที่ปรากฏอยู่ในตารางทั้งหมดทุกตารางในฐานข้อมูลมาเปรียบเทียบกับ Rule ทั้งหมดใน Rule-base หลังจากขั้นตอนเหล่านี้จะได้ค่าตัวแปร Nf เกิดขึ้น โดยจะนำค่าตัวแปร Nf นี้ ไปคำนวนหาค่าความน่าเชื่อถือได้จากสมการที่ 1

4. บทสรุป

บทความนี้ได้กล่าวถึง การประยุกต์ความน่าเชื่อถือ (Reliability) ที่ใช้อยู่ในด้านวิศวกรรมการผลิตมาใช้กับระบบสารสนเทศ โดยการประยุกต์ใช้นั้น ได้แบ่งลักษณะการประเมินหาค่าความน่าเชื่อถือออกเป็น 3 ลักษณะคือ 1. ค่าความน่าเชื่อถือของรายการข้อมูลในแต่ละรายการ (Internal Reliability : Ri) คือ ค่าความน่าเชื่อถือของข้อมูลที่เป็นไปตามลักษณะพื้นฐานของรายการข้อมูลแต่ละรายการ โดยที่ลักษณะพื้นฐานของรายการข้อมูลนี้ จะเป็นมาตรฐานที่ยอมรับกันโดยทั่ว ๆ ไป 2. ความน่าเชื่อถือของข้อมูลที่สัมพันธ์กันตามข้อกำหนดหรือเงื่อนไขของระบบฐานข้อมูล (Relative Reliability : Rr) คือ ค่าความน่าเชื่อถือที่ประเมินได้จากการเปรียบเทียบข้อมูลว่าเป็นจริง ตามความสัมพันธ์และเงื่อนไขในระบบฐานข้อมูลที่ได้กำหนดไว้หรือไม่ 3. ความน่าเชื่อถือของข้อมูลที่เป็นไปตามความเป็นจริง (Absolute Reliability : Ra) คือ ค่าความน่าเชื่อถือของข้อมูลที่ได้จากการเปรียบเทียบข้อมูลที่มีอยู่ในฐานข้อมูลกับข้อมูลที่เกิดขึ้นตามความเป็นจริง

เมื่อทราบลักษณะในการประเมินค่าความน่าเชื่อถือทั้ง 3 ลักษณะแล้ว ก็จะนำเอาระบบผู้เชี่ยวชาญมาช่วยในการประเมินค่าความน่าเชื่อถือ โดยการที่จะสร้าง Rule-base ตามเงื่อนไขของแต่ละลักษณะของค่าความน่าเชื่อถือที่กล่าวมาข้างต้น

ในการวิจัยนั้นยังพบปัญหาและอุปสรรคในส่วนของลักษณะการประเมินความน่าเชื่อถือของข้อมูลตามความเป็นจริง (Ra) เพราะในการที่จะทำการประเมินความน่าเชื่อถือของข้อมูลลักษณะนี้ จะต้องมีวิธีที่จะสร้าง Rule-base ที่ควบคุมถึงรายการข้อมูลทั้งหมด ซึ่งวิธีที่ใช้อยู่ คือ การนำผลสรุปของรายการข้อมูลที่มีชนิดเป็นตัวเลข (Numeric) มาเปรียบเทียบกับข้อมูลที่คำนวณได้จากความเป็นจริงตามที่กล่าวมาแล้วข้างต้นนั้นไม่สามารถที่จะประเมินหาความน่าเชื่อถือของรายการข้อมูลได้ทั้งหมด เพราะฉะนั้นจะต้องมีการศึกษาค้นคว้าหาวิธีที่เหมาะสมต่อไป

เอกสารอ้างอิง

[1] Nachman Agmon/Niv Ahituv , "Assessing Data Reliability in an Information System", Journal of Management Information System , Vol. 4, No.2, 1987.

[2] G. M. Nijssen/T. A. Halpin , "Conceptual Schema and Ralational Database Design", Prentice Hall, 1989

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[3] Giarratano/Riley , 'Expert Systems Principles and Programming', PWS-KENT publishing Company, 1989

[4] Ralph M. Stair, 'Principle of Information System Amanagerial Approach', Boyd&Fraser , 1992.

[5] J.F. Brule and A. Blount, 'Knowledge Acquisition', McGraw Hill, 1990.

[6] ครรชิต ไมตรี , 'Expert System' , ตำราชุดวิศวกรรมศาสตร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , 1992.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นายพีระศักดิ์ เสรีกุล เกิดเมื่อวันที่ 18 กุมภาพันธ์ พ.ศ. 2514 ที่อำเภอบางกะปิ จังหวัดกรุงเทพมหานคร เป็นบุตรคนที่ 3 ของนายแดง เสรีกุล และนางเชื่อม เสรีกุล สำเร็จการศึกษาระดับมัธยมศึกษาที่โรงเรียนบางกะปิ ในปีการศึกษาา 2530 สำเร็จการศึกษาระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต สาขาคอมพิวเตอร์ ในปีการศึกษา 2534 และเข้าศึกษาในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ที่สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เมื่อ พ.ศ. 2535