ปีการศึกษา 2531

เปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่าง

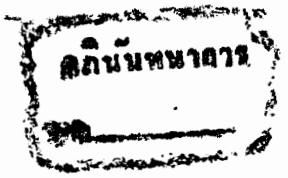
โดย

นายวิศิษฏ์ หิรัญกิตติ

อาจารย์ที่ปรึกษา

ดร. ศุภมิตร จิตตะยโศธร

ศจ.ดร. ไพรัช ธัชยพงษ์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

023241 10.ค.2532

ปริญญานิพนธ์ปีการศึกษา 2531

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง เปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่าง

ผู้จัดทำ

นายวิศิษฏ์ หิรัญกิตติ

.....อาจารย์ที่ปรึกษา

(ดร. ศุภมิตร จิตตะยโศธร)

.....อาจารย์ที่ปรึกษา

(ศจ.ดร.ไพรัช ธัชยพงษ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

023241

สารบัญ

		หน้า
บทคัดย่อ		ก
สารบัญตารางและรูปภาพ		ค
บทที่ 1	บทนำ	1
บทที่ 2	วัตถุประสงค์ ความเป็นมา และ ทฤษฎีที่เกี่ยวข้อง	3
	2.1 วัตถุประสงค์	3
	2.2 ความเป็นมา	4
	2.3 ทฤษฎีที่เกี่ยวข้อง	6
บทที่ 3	เล็กซ์เซลล์	8
	3.1 สถาปัตยกรรมของเล็กซ์เซลล์	8
	3.2 วิธีการเรียนรู้	9
	3.2.1 การเรียนรู้จากตัวอย่าง	10
	3.2.2 ไอดีทรี	13
	3.2.3 การสร้างต้นไม้โดยวิธีของไอดีทรี	14
	3.3 การปรับปรุงไอดีทรี	20
	3.3.1 การปรับปรุงไอดีทรีวิธีที่ 1	20
	3.3.2 การปรับปรุงไอดีทรีวิธีที่ 2	27
	3.3.3 สรุป	32
บทที่ 4	ระบบสารสนเทศยุคที่ห้า	33
บทที่ 5	ตัวประมวลผลสารสนเทศที่เป็นแนวคิด	39
บทที่ 6	แบบจำลองข้อมูลไนแอม	45
บทที่ 7	ไนแอมเบสพีพีจีไอเอส	51
	7.1 เมต้าคอนเซบชวลสกีมม่า	51
	7.1.1 การออกแบบเมต้าคอนเซบชวลสกีมม่า	52
	7.1.2 ส่วนตรวจหอนเซบชวลสกีมม่า	54
	7.2 การสร้างระบบไนแอมเบสพีพีจีไอเอส	55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

7.2.1 สถาปัตยกรรมของไนแอมเบสฟิฟจีไอเอส 55

7.2.2 ซีไอพี 66

7.2.3 ฐานข้อมูล 71

บทที่ 8 การถ่ายความรู้ 73

บทที่ 9 การทดลอง 77

9.1 ขั้นตอนการพัฒนาโครงงาน 77

9.2 ผลการทดลอง 78

9.3 ตัวอย่างการทำงานของระบบ 81

บทที่ 10 สรุปและวิจารณ์ 82

กิตติกรรมประกาศ 84

เอกสารอ้างอิง 85

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่าง

วิศิษฏ์ หิรัญกิตติ

ดร. ศุภมิตร จิตตะยโศธร อาจารย์ที่ปรึกษา

ศจ.ดร. ไพรัช ธัชยพงษ์ อาจารย์ที่ปรึกษา

ปีการศึกษา 2531

บทคัดย่อ

ในปัจจุบันนี้ได้มีการนำระบบผู้เชี่ยวชาญมาใช้งานด้านต่าง ๆ มากขึ้น เนื่องจากได้รับการพัฒนาจนสามารถให้คำปรึกษาและคำแนะนำได้ใกล้เคียงผู้เชี่ยวชาญที่เป็นมนุษย์ แต่ปัญหาที่พบในการพัฒนาระบบผู้เชี่ยวชาญคือ ความยากในการถ่ายทอดความรู้จากผู้เชี่ยวชาญไปสู่ระบบ หนทางหนึ่งในการแก้ไขคือ ใช้ความสามารถในการเรียนรู้จากตัวอย่าง (Learning by examples) เพื่อสร้างกฎขึ้นมา โครงงานนี้ได้สร้าง เล็กซ์เชลล์ (LEX-SHELL) ซึ่งเป็นโปรแกรมต้นแบบของเปลือกระบบผู้เชี่ยวชาญที่พัฒนาโดยใช้ภาษาลิสป์ (LISP) มีความสามารถในการอุปมาน (induce) กฎในรูปของต้นไม้สำหรับการตัดสินใจ (Decision Tree) ซึ่งได้จากตัวอย่างของปัญหา (Training Set) ที่ให้โดยผู้เชี่ยวชาญ นอกจากนี้ยังสามารถแปลงกฎที่อยู่ในรูปของต้นไม้สำหรับการตัดสินใจนี้ให้อยู่ในรูปของ สับเซตของภาษาธรรมชาติ เพื่อใช้กับเปลือกระบบผู้เชี่ยวชาญโดยทั่วไป และอยู่ในรูปของโปรลอกคลอส (PROLOG clauses) เพื่อถ่ายให้กับไนแอมเบสฟิฟจีไอเอส (NIAM-based 5GIS) ซึ่งเป็นเปลือกระบบผู้เชี่ยวชาญอีกระบบหนึ่งที่พัฒนาขึ้นสำหรับโครงงานนี้โดยใช้ภาษาโปรลอก (PROLOG) เพื่อแสดงให้เห็นว่าสามารถนำกฎที่ได้จากเล็กซ์เชลล์ไปใช้กับไนแอมเบสฟิฟจีไอเอสได้

วิทยานิพนธ์ฉบับนี้จะได้กล่าวถึง เล็กซ์เชลล์ การปรับปรุงข้อบกพร่องของไอดีทรี การพัฒนาไนแอมเบสฟิฟจีไอเอส และการถ่ายความรู้ระหว่างเปลือกระบบผู้เชี่ยวชาญทั้งสองระบบดังกล่าว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

AN EXPERT SYSTEM SHELL WITH THE CONCEPTS CLASSIFICATION AND LEARNING CAPABILITIES

Wisit Hirunkitti

Dr. Suphamit Chittayasotorn Advisor

Dr. Pairach Thajchayapong Advisor

1988

Abstract

At present, expert systems gain popularity in real-world applications. Their abilities to give advices and explain reasons behind the advices are recognized and widly accepted of the biggest problems in expert systems development is the "Feigenbaum bottleneck" which is the problem of how to transfer the knowledge from human experts to the systems [1]. So far, inductive learning is found to be one of the most effective solutions to the problem.

This project developes LEX-SHELL: a LISP-based prototype expert system shell that can induce rules which are represented by decision trees from examples in training sets which are given by experts. It can generate rules in subset of natural language form for general expert system shells and in PROLOG clauses form to transfer to NIAM-based 5GIS, another PROLOG-based expert system shell that is also developed by this project. It is important to show that induced rules from LEX-SHELL can be transferr to NIAM-based 5GIS.

All about LEX-SHELL, ID3 improvement, NIAM-based 5GIS developement and knowledge transfer are described in this thesis.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

ในขบวนการสร้างระบบผู้เชี่ยวชาญนั้นวิธีที่นิยมใช้กันมากก็คือ การนำเอาความรู้และความชำนาญเฉพาะด้านจากผู้เชี่ยวชาญใส่ให้กับฐานความรู้ของเปลือกระบบผู้เชี่ยวชาญ ซึ่งเปรียบเสมือนระบบผู้เชี่ยวชาญที่ฐานความรู้ยังปราศจากความรู้ สำหรับการนำเอาความรู้จากผู้เชี่ยวชาญมาใส่ให้กับฐานความรู้นั้น เป็นหน้าที่ของวิศวกรความรู้ที่จะพยายามกลั่นกรองเอาความรู้ในการวิเคราะห์วินิจฉัยปัญหาจากผู้เชี่ยวชาญเฉพาะด้านโดยอาศัยวิธีการสัมภาษณ์ตัวต่อตัว จากนั้นวิศวกรความรู้จะทำการแปรเปลี่ยนความรู้ที่ได้รับมาให้อยู่ในรูปแบบของการแทนความรู้ (Knowledge Representation) ที่เหมาะสมเพื่อใส่ในฐานความรู้ของเปลือกระบบผู้เชี่ยวชาญ หน้าที่ดังกล่าวของวิศกรความรู้นี้เรียกว่า ขบวนการจัดหาความรู้ (Knowledge Acquisition) สำหรับในแง่ของการปฏิบัติแล้วขบวนการจัดหาความรู้เป็นขบวนการที่ยุ่งยากและจะต้องเผชิญอุปสรรคนานัปการ เป็นต้นว่า

1. ผู้เชี่ยวชาญไม่สามารถให้คำอธิบายหรือให้คำอธิบายได้ไม่ชัดเจนเกี่ยวกับหลักเกณฑ์ที่ใช้ในการวินิจฉัย เนื่องจากการสื่อความเข้าใจของตน ให้คนอื่นได้รับทราบนั้นทำได้ลำบาก
2. ในขบวนการวินิจฉัยปัญหาบางอย่างของผู้เชี่ยวชาญ อาจจะใช้วิธีการที่ง่ายมากจนทำให้วิศวกรความรู้ไม่มีความเชื่อมั่นว่าจะสามารถนำไปใช้ในการวินิจฉัยปัญหาได้จริง
3. การได้มาซึ่งกฎเกณฑ์แต่ละข้อของการวินิจฉัยจากผู้เชี่ยวชาญนั้น วิศวกรความรู้จะต้องสูญเสียเวลาเป็นจำนวนมากในการสัมภาษณ์ ดังนั้นการพัฒนาระบบผู้เชี่ยวชาญแต่ละระบบจึงต้องใช้เวลายาวนานมาก ในขณะที่ปัจจุบันความต้องใช้ระบบผู้เชี่ยวชาญเพิ่มขึ้นอย่างรวดเร็ว

ปัญหาต่าง ๆ ที่กล่าวมานี้นั้นเรียกอีกอย่างหนึ่งว่า "Feigenbaum Bottle neck" [1] เป็นปัญหาสำคัญที่ยังเผชิญอยู่ ซึ่งถ้าสามารถเอาชนะได้แล้ว จะทำให้การพัฒนาระบบผู้เชี่ยวชาญทำได้สะดวกรวดเร็วยิ่งขึ้น หนทางหนึ่งในการแก้ไขปัญหานี้ทำได้โดยการเพิ่มความสามารถในการเรียนรู้ เพื่อดึงเอาความรู้จากผู้เชี่ยวชาญใส่ให้กับเปลือกระบบผู้เชี่ยวชาญโดยอัตโนมัติ แทนที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะใช้เฉพาะวิธีการดึงความรู้โดยอาศัยวิศวกรความรู้เพียงอย่างเดียว ส่วนของการเรียนรู้นี้จะสามารถรับเอาความรู้มาจากผู้เชี่ยวชาญแล้วแปรเปลี่ยนให้เป็นแบบของการแทนความรู้ที่กำหนด ซึ่งเปลือกระบบผู้เชี่ยวชาญสามารถนำไปใช้ได้

สำหรับระบบเล็กซ์เชลล์ (LEX-SHELL: Learning-by-Examples Expert System SHELL) นี้เป็นต้นแบบของเปลือกแบบผู้เชี่ยวชาญที่พัฒนาด้วยภาษาลิสป์ (LISP) ซึ่งมีความสามารถในการดึงเอาความรู้ที่มีลักษณะเป็นการจำแนกประเภท (Concept classification) จากผู้เชี่ยวชาญโดยขบวนการเรียนรู้จากตัวอย่างที่ให้โดยผู้เชี่ยวชาญ โดยระบบสามารถที่จะอุปมาน (Induce) กฎที่ใช้วินิจฉัยได้ครอบคลุมตัวอย่างที่ให้โดยผู้เชี่ยวชาญทั้งหมด แล้วแสดงอยู่ในรูปของต้นไม้สำหรับการตัดสินใจ (Decision Tree) ตัวอย่างที่ให้โดยผู้เชี่ยวชาญจะต้องเป็นตัวอย่างที่ผู้เชี่ยวชาญมั่นใจว่าครอบคลุมการจำแนกประเภทได้อย่างสมบูรณ์ และยังสามารถแปลงกฎที่อยู่ในรูปของต้นไม้สำหรับการตัดสินใจไปเป็น กฎในรูปสับเซตของภาษาธรรมชาติ เพื่อสามารถนำไปใช้กับเปลือกระบบผู้เชี่ยวชาญโดยทั่วไป และ กฎที่อยู่ในรูปของโปรลอกคลอส (PROLOG Clauses) เพื่อใช้กับไนแอมฟิฟจีไอเอส (NIAM-based 5GIS)

ไนแอมฟิฟจีไอเอสเป็นระบบที่สร้างขึ้นจากแนวคิดของฟิฟจีไอเอส(5GIS: The 5th Generation Information System) ซึ่งโปรเฟสเซอร์ จี เอ็ม ไนเซน (Prof. G. M. Nijssen) ได้เสนอไว้ โดยใช้ไนแอมเป็นแบบแทนคอนเซบชวลสกีมม่า และสร้างจากภาษาโปรลอก มีความสามารถพิเศษในการเตรียมชุดฝึกหัดเพื่อให้เลกซ์เชลล์อุปมานกฎให้ แล้วนำกฎเหล่านั้นกลับเข้ามาใช้ในระบบ

ในวิทยานิพนธ์ฉบับนี้ จะกล่าวสถาปัตยกรรมเล็กซ์เชลล์ ทฤษฎีไอดีทรี และ วิธีการปรับปรุงวิธีการไอดีทรี ในบทที่ 3 สำหรับบทที่ 4 กล่าวถึงแนวคิดของระบบสารสนเทศยุคที่ 5 หรือ ฟิฟจีไอเอส และจะกล่าวถึงตัวประมวลผลสารสนเทศที่เป็นแนวคิด หรือ ซีไอพี ในบทที่ 5 สำหรับคำอธิบายแบบจำลองข้อมูลไนแอมจะอยู่ในบทที่ 6 ในบทที่ 7 ได้กล่าวถึงการสร้างไนแอมเบสฟิฟจีไอเอส รวมทั้งคอนเซบชวลสกีมม่าที่ใช้อธิบายโครงสร้างของไนแอม ที่เรียกว่า เมต้าคอนเซบชวลสกีมม่าเนื้อหาในบทที่ 8 เป็นการอธิบายการถ่ายความรู้ระหว่างเล็กซ์เชลล์กับไนแอมฟิฟจีไอเอส สำหรับการทดลอง สรุปและวิจารณ์ จะอยู่ในบทที่ 9 และ 10 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

วัตถุประสงค์ ความเป็นมา และ ทฤษฎีที่เกี่ยวข้อง

2.1 วัตถุประสงค์

การจัดทำโครงการวิจัยนี้มีวัตถุประสงค์ที่สำคัญ 6 ประการ คือ

ประการที่หนึ่ง เพื่อกระตุ้นให้มีการค้นคว้าวิจัยระบบผู้เชี่ยวชาญ ตลอดจนถึงการนำเอาความรู้ มาใช้งานในด้านต่าง ๆ ให้แพร่หลายยิ่งขึ้น เป็นการใช้ความรู้ให้เป็นประโยชน์ในรูปแบบของอุตสาหกรรมความรู้ เพื่อช่วยแบ่งเบาภาระของผู้เชี่ยวชาญ เนื่องจากปัจจุบันประเทศยังขาดแคลนผู้เชี่ยวชาญด้านต่าง ๆ เป็นจำนวนมาก และยังสามารถช่วยทดแทนการขาดแคลนบุคคลากรที่จะเป็นกำลังสำคัญในการพัฒนาประเทศต่อไปในอนาคตด้วย

ประการที่สอง เพื่อนำเอาเทคโนโลยีทางด้านความรู้ มาใช้สร้างสรรค์ประโยชน์ทางด้านการศึกษา ด้านเศรษฐกิจ และสังคม ตลอดจนความอยู่ดีกินดี ของประชาชนในชาติ

ประการที่สาม เพื่อค้นคว้าวิธีการใหม่ ๆ ในการสร้างเปลือกระบบผู้เชี่ยวชาญในอีกรูปแบบหนึ่ง โดยอาศัยทฤษฎีระบบผู้เชี่ยวชาญ ทฤษฎีการเรียนรู้ ซึ่งต่างก็เป็นสาขาหนึ่งในปัญญาประดิษฐ์ร่วมกับ เทคโนโลยีทางด้านการออกแบบและการจัดการฐานข้อมูล ประกอบกับ การใช้ความสามารถในการอนุมานของภาษาโปรลอก (PROLOG)

ประการที่สี่ เพื่อเสนอเป็นแนวทางหนึ่งในการพัฒนาเปลือกระบบผู้เชี่ยวชาญ หรือระบบผู้เชี่ยวชาญที่มีการใช้ เทคโนโลยีทางด้านการออกแบบและการจัดการฐานข้อมูล

ประการที่ห้า เพื่อสร้างต้นแบบของเปลือกระบบผู้เชี่ยวชาญที่มีความสามารถในการอุปมานกฎได้จากตัวอย่าง เพื่อใช้ทดสอบแนวความคิดว่ามีความเป็นไปได้ในทางปฏิบัติจริง

ประการสุดท้าย เพื่อแสดงให้เห็นว่าสามารถนำกฎที่สร้างจากส่วนดึงความรู้ของเปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่าง หรือ เล็กซ์เชลล์ (LEX-SHELL) ไปใช้กับเปลือกระบบผู้เชี่ยวชาญอื่น(NIAM-based 5GIS)ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 ความเป็นมา

ความเป็นมาของการเรียนรู้ของเครื่องจักรกล (Machine Learning)

ขบวนการเรียนรู้เป็นขบวนการหนึ่งที่เกี่ยวข้องกับการรับเอาความรู้ใหม่ ๆ เข้ามา การพัฒนาทักษะความนึกคิดจากการสั่งสอนและการฝึกฝน การจัดรูปแบบของความรู้ใหม่ ๆ ให้อยู่ในรูปแบบทั่วไป ตลอดจนการค้นพบข้อเท็จจริงและทฤษฎีใหม่ ๆ จากการสังเกตและการทดลอง

การเรียนรู้ของเครื่องจักรกล (Machine Learning) เป็นสาขาวิชาหนึ่งของปัญญาประดิษฐ์ (Artifical Intelligence) ที่ทำการศึกษาค้นคว้าเพื่อพยายามสร้างแบบจำลองของขบวนการเรียนรู้ให้เกิดมีขึ้นในเครื่องคอมพิวเตอร์

ตลอดเวลาหลายสิบปีที่ผ่านมา การวิจัยค้นคว้าทางด้านการเรียนรู้ของเครื่องจักรกลก็ได้พัฒนาไปด้วยลักษณะและเป้าหมายที่แตกต่างกันไป ซึ่งแบ่งได้เป็น 3 ยุค โดยที่แต่ละยุคจะมีรูปแบบที่แตกต่างกันออกไป ดังนี้

ยุคที่ 1 ยุคของเทคนิคการสร้างแบบจำลองนิวรัล (Neural Modeling Technique) และเทคนิคทางทฤษฎีการตัดสินใจ (Decision Theoretic Technique) เป็นยุคที่มุ่งความสนใจไปที่การสร้างระบบการเรียนรู้ใช้งานอเนกประสงค์ (General Purpose Learning System) ซึ่งเป็นระบบที่ตั้งต้นจากที่ไม่มีโครงสร้างเริ่มต้น หรือ ความรู้เริ่มต้นเกี่ยวกับงานใด ๆ ในระบบเลย หรือถ้ามีก็เพียงเล็กน้อย การวิจัยส่วนใหญ่เชื่อกันว่าโดยอาศัยตาบูล่า ราซ่า (Tabula Rasa) รวมกับการสร้างเครื่องจักรกลที่สร้างจากแบบจำลองนิวรัล (Neural Model-based Machine) แบบต่าง ๆ และโครงสร้างเริ่มต้นที่สุ่มขึ้นมา แล้วระบบนี้จะสามารถเรียนรู้ได้ด้วยตัวเอง ระบบนี้เรียกกันว่า โครงข่ายนิวรัล (Neural Net) การวิจัยในแนวทางนี้ก็มีทั้งทางทฤษฎี และการสร้างระบบฮาร์แวร์พิเศษ ได้แก่ เพอร์เซปตรอน (Perceptron) แพนดิโมเนียม (Pandimonium) และ อะเดลเลี่ยน (Adelaine)

จากประสบการณ์การค้นคว้าวิจัยดังกล่าว ทำให้เกิดหลักการใหม่ในสาขาวิชาการจดจำรูปแบบ (Pattern Recognition) และนำไปสู่การพัฒนาเกี่ยวกับทฤษฎีการตัดสินใจซึ่งการเรียนรู้โดยวิธีนี้ก็คือการสร้างลิเนียร์ฟังก์ชั่น โพลิโนเมียลฟังก์ชั่น หรือ ฟังก์ชั่นรูปแบบอื่น ๆ จากชุดตัวอย่างฝึกหัด (Training Examples) ที่ระบบได้รับ

ผลในทางปฏิบัติของการประยุกต์ใช้การสร้างแบบจำลองนิวรัล และเกี่ยวกับทฤษฎี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การตัดสินใจนั้นประสบผลสำเร็จอยู่เพียงในวงจำกัด สิ่งที่คาดหมายไว้ตั้งที่เคยประกาศออกไปตั้งแต่ต้น ก็ยังห่างไกลจากความเป็นจริง การวิจัยค้นคว้าในแนวทางนี้จึงได้ลดน้อยลงไป จากการศึกษาทางทฤษฎีในสมัยนั้น ได้แสดงให้เห็นถึงความจำกัดของระบบการเรียนรู้ที่ปราศจากความรู้แบบเพอร์เซปตรอน จึงทำให้เกิดการเปลี่ยนแปลงรูปแบบของการวิจัยในยุคต่อมา

ยุคที่ 2 ยุคของการเรียนรู้ที่เน้นแนวความคิดทางด้านสัญญลักษณ์ (Symbolic Concept-Oreinted Learning) ยุคนี้เริ่มในราวปี ค.ศ. 1960 โดยมีนักจิตวิทยาและนักวิจัยปัญญาประดิษฐ์ได้ทำการศึกษาถึง ตัวแบบจำลองของการเรียนรู้ของมนุษย์ เป็นรูปแบบของการใช้การแทนโครงสร้างโดยอาศัยตรรกศาสตร์หรือกราฟ แทนการใช้วิธีการใช้ตัวเลขหรือการคำนวณทางสถิติดังเช่นในยุคแรก ระบบจะทำการเรียนรู้คำอธิบายที่เป็นสัญญลักษณ์ซึ่งใช้แทนความรู้และสร้างข้อสมมุติฐานจากแนวความคิด (Concept) จากคำอธิบายดังกล่าว

ผลงานวิจัยที่เกิดขึ้นในช่วงนี้ได้แก่

- CLS (Concept Learning System) ของ Hunt et al. ในปี ค.ศ. 1966 นับเป็นระบบแรกที่สามารถสร้างกฎจากตัวอย่างได้อย่างอัตโนมัติ โดยแทนกฎอยู่ในรูปของต้นไม้ สำหรับการตัดสินใจ (Decision Trce)

- BOXES ของ Michie และ Chamber เป็นระบบเรียลไทม์ (Real-Time System) ที่สามารถเรียนรู้ในการดุลโพล (Pole) บนรถที่กำลังเคลื่อนที่ ระบบจะทำการดัดแปลงกลุ่มของ โปรดักชั่นรูล (Production Rules) ซึ่งมีทั้งหมด 225 กฎ โดยทดลองรันแสดงการจำลอง (Simulation) บนจอมอเตอร์ ระบบสามารถสร้างความเชี่ยวชาญได้ด้วยตัวเองจากการลองผิดลองถูก หรือการฝึกสอนจากผู้เชี่ยวชาญ

- ผลงานของ Winston และผลงานของ Barrow ร่วมกับ Popplestone เป็นการใช้ โครงข่ายความหมาย (Semantic Nets) ในการบรรยายฉากภาพ (Visual Scenes) ความสามารถในการเรียนรู้ของระบบนี้ก็คือ โปรแกรมสามารถทำการดัดแปลงคำบรรยายภาพจากตัวอย่างฉากที่ได้รับมา

- ผลงานของ Michalski ร่วมกับ Chilausky และ Jacobsen เป็นระบบที่แสดงให้เห็นถึงข้อได้เปรียบของการสร้างกฎโดยวิธีการอุปมาณ (Induction) เมื่อเทียบกับวิธีการสร้างกฎ โดยวิธีการสัมภาษณ์ ดังที่ใช้กันโดยทั่วไป ซึ่งจะช่วยทั้งแรงงานและเวลา

- ผลงานของ Quinlan ซึ่งทำการพัฒนาระบบ CLS ของ Hunt ให้ดีขึ้น คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คัดแปลงให้สามารถรับไฟล์ตัวอย่างขนาดใหญ่มากได้ และมีการตรวจสอบความขัดแย้งในตัวอย่าง แล้วต่อมาก็ได้เพิ่มการคัดเลือกกลุ่มตัวอย่าง เพื่อมาเป็นเวอร์กกิงเซ็ต (working Set) และการอุปมาณใหม่เพื่อปรับปรุงกฎรวมเข้าไป ระบบการเรียนรู้เหล่านี้เป็นชุดของ ID3 ซึ่งก็มี ACLS ASSISTANT Expert-Ease EX-TRAN และ Rule Master

ยุคที่ 3 เป็นยุคที่เริ่มในราวปี 1975 จนถึงปัจจุบัน การค้นคว้าวิจัยได้ขยายขอบเขต จากการเรียนรู้จากชุดตัวอย่างไปสนใจวิธีการเรียนรู้ที่กว้างขวางออกไป ซึ่งงานวิจัยส่วนใหญ่จะอาศัยระบบบรรจุ ความรู้มาก (Knowledge-Rich System) ทำให้แนวโน้มต่อไปของการค้นคว้าวิจัยในยุคนี้ มีดังนี้

1. วิจัยมุ่งสู่ระบบเรียนรู้ที่เน้นความรู้เข้มข้น (Knowledge - Intensive Approcahes)

การวิจัยที่ทำจะเน้นหนักในการใช้ความรู้ที่เกี่ยวกับงานและกฎเกณฑ์ข้อบังคับ (Constraints) เพื่อช่วยชี้นำขบวนการเรียนรู้

2. วิจัยเพื่อสำรวจการเรียนรู้วิธีต่าง ๆ สืบเนื่องมาจากในอดีตการวิจัยได้มุ่งเน้นวิธีการเรียนรู้จากตัวอย่าง ปัจจุบันจึงหันไปพิจารณาการเรียนรู้วิธี อื่น ๆ ดังเช่น การเรียนรู้จากคำสั่ง (Learning from Instruction) การเรียนรู้โดยการเทียบเคียง (Learning by Analogy) การเรียนรู้จากการค้นพบ (Learning From Discovery)

3. วิจัยเพื่อรวมเอาความสามารถในการสร้างและการเลือกงานที่เรียนรู้ (Learning Tasks) ในปัจจุบันมีหลายระดับด้วยกันที่ใช้ฮิวริสติก (Heuristics) ควบคุมความสนใจของระบบ โดยการสร้างงานที่เรียนรู้ การเสนอการทดลองเพื่อรับข้อมูลฝึกหัด (Training Data) และการเลือกรับแนวความคิด

2.3 ทฤษฎีที่เกี่ยวข้อง

แม้ว่า ปัญญาประดิษฐ์ จะได้มีการค้นคว้าวิจัยกันมานานหลายสิบปีแล้วก็ตาม แต่ก็ยังมีลักษณะของศิลปะมากกว่าวิทยาศาสตร์ ดังนั้น นักวิจัยในสาขานี้จึงได้มีการเอาหลักวิชาแขนงอื่น ๆ มาประยุกต์ใช้ร่วมกัน หลักวิชาที่มีการใช้มากในปัญญาประดิษฐ์ ได้แก่ วิชาตรรกศาสตร์ วิชาสถิติ วิชาคณิตศาสตร์ จิตวิทยา ฯลฯ

ในการสร้างส่วนจัดหาความรู้นั้นจะต้องอาศัยขบวนการเรียนรู้ ซึ่งในขบวนการเรียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รู้นั้น วิธีการเรียนรู้แต่ละวิธีก็มีการประยุกต์ใช้ศาสตร์ต่าง ๆ ที่แตกต่างกันไป วิธีการของไอดีทรีได้ประยุกต์ใช้ ทฤษฎีข่าวสาร มาเป็นเกณฑ์ในการเลือกแอทริบิวท์เพื่อสร้างต้นไม้สำหรับการตัดสินใจ

สำหรับระบบสารสนเทศยุคที่ 5 (The 5th Generation Information System) เป็นการพัฒนาการออกแบบและการจัดการฐานข้อมูลในปัจจุบันให้มีประสิทธิภาพมากยิ่งขึ้น และง่ายต่อการติดต่อกับผู้ใช้ แต่อย่างไรก็ตาม ทฤษฎีการออกแบบและการจัดการฐานข้อมูลในปัจจุบัน ก็ถือว่าเป็นพื้นฐานอันสำคัญสำหรับการพัฒนา ระบบสารสนเทศยุคที่ 5 และในการพัฒนาไนแอมเบสฟิฟจีไอเอสนั้นจำเป็นจะต้องอาศัยความรู้เกี่ยวกับแบบจำลองข้อมูลไนแอมเป็นอย่างดี อย่างไรก็ตามในอนาคตอาจจะมีแบบจำลองข้อมูลที่ดีกว่าไนแอม ซึ่งก็จะต้องมีการค้นคว้ากันต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

เล็กซ์เชลล์(LEX-SHELL)

3.1 สถาปัตยกรรมของระบบเล็กซ์เชลล์ (LEX-SHELL ARCHITECTURE)

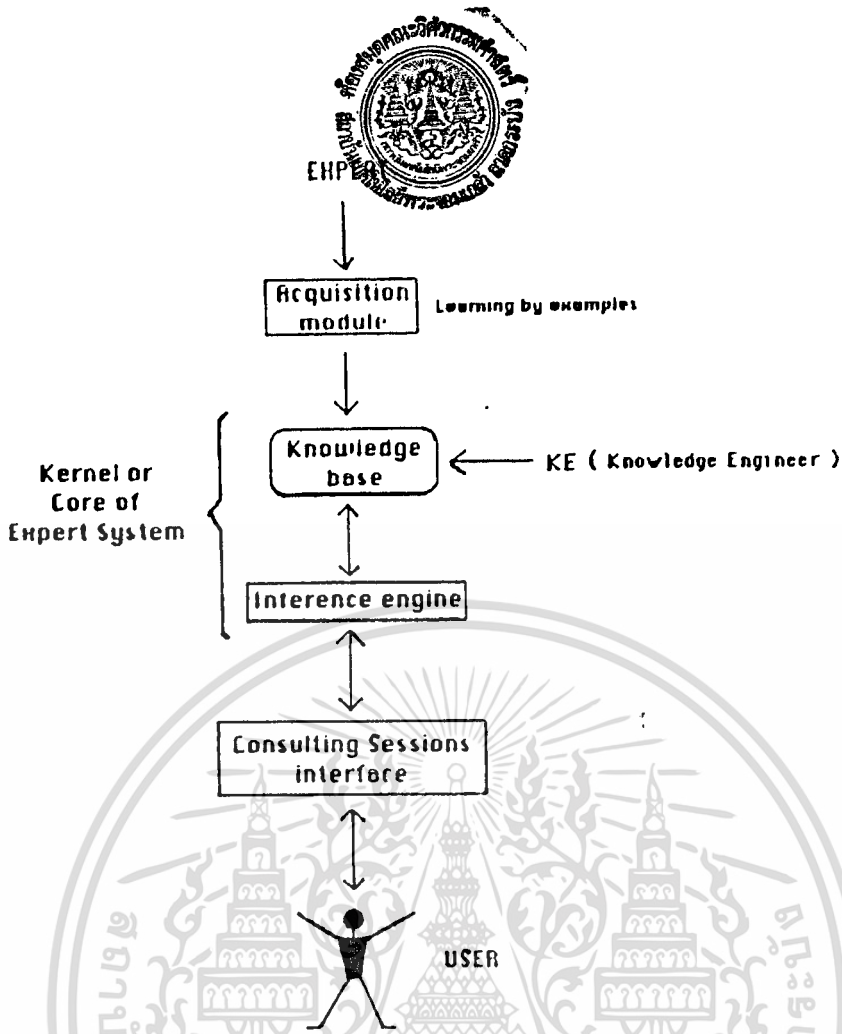
1. ส่วนจัดหาความรู้ (Acquisition Module) เป็นส่วนที่ทำหน้าที่ดึงเอาความรู้จากผู้เชี่ยวชาญโดยอาศัยขบวนการเรียนรู้จากตัวอย่าง (Learning by Examples) โดยรับเอากลุ่มตัวอย่างของวิธีวินิจฉัยปัญหาหรือเรียกว่า ชุดฝึกหัด(Training set) มาจากผู้เชี่ยวชาญแล้วทำการอุปมานกฎซึ่งเป็นการจำแนกประเภทของความรู้ที่ใช้ในการวินิจฉัย โดยแทนกฎให้อยู่ในรูปของต้นไม้สำหรับการตัดสินใจ

2. ฐานความรู้ (Knowledge Base) เป็นส่วนที่ใช้เก็บกฎที่ได้จากการอุปมานจากส่วนจัดหาความรู้ ซึ่งอยู่ในรูปของต้นไม้สำหรับการตัดสินใจ

3. เครื่องจักรอินเฟอเร้นซ์ (Inference Engine) เป็นส่วนที่ใช้ความรู้ซึ่งอยู่ในรูปของต้นไม้สำหรับการตัดสินใจในฐานความรู้ โดยใช้ขบวนการ Data-driven decision tree Search และสามารถจัดการกับความรู้ที่แน่นอนได้เท่านั้น

4. ส่วนจัดการให้คำปรึกษา (Consulting Session) เป็นส่วนที่จะให้คำอธิบายการวินิจฉัยของส่วนเครื่องจักรอินเฟอเร้นซ์ โดยผู้ใช้สามารถที่จะถามหาเหตุผลว่าทำไมระบบจึงถามคำถามเช่นนี้หรือว่าทำไม ถึงได้ข้อสรุปอย่างนี้ และในบางครั้งเมื่อได้ข้อสรุปแล้วระบบอาจจะต้องการแสดงคำอธิบายเพิ่มเติมเกี่ยวกับข้อสรุป ซึ่งก็เป็นหน้าที่ของส่วนนี้เช่นเดียวกัน เมื่อพิจารณาให้ดีแล้วจะเห็นว่า ส่วนจัดการให้คำปรึกษาเป็นส่วนที่สำคัญมากส่วนหนึ่ง เนื่องจากถึงแม้ว่าระบบจะมีความสามารถในการวินิจฉัยดีเพียงใดก็ตาม แต่ถ้าไม่สามารถให้เหตุผลของการวินิจ-ฉัยได้ ก็ถือว่าเป็นระบบที่ยังใช้งานไม่ได้ เนื่องจากการใช้ความรู้โดยขาดซึ่งเหตุผลย่อมไม่สามารถยอมรับได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 แสดงสถาปัตยกรรมของระบบ LEX-SHELL

3.2 วิธีการเรียนรู้ (Learning Method)

จากข้อเท็จจริงที่ว่าผู้เชี่ยวชาญสามารถที่จะให้ตัวอย่างของการแก้ปัญหาได้ง่ายกว่าที่จะแสดงวิธีการแก้ปัญหาออกมาในรูปของกฎโดยตรงเป็นการหลีกเลี่ยง Faigenbaum Bottle neck ดังนั้นขบวนการเรียนรู้จากตัวอย่างจึงเป็นวิธีการที่เหมาะสมในการดึงความรู้จากผู้เชี่ยวชาญ โดยผู้เชี่ยวชาญเพียงแต่ให้ชุดของตัวอย่างที่แสดงการแก้ปัญหา เรียกว่า ชุดฝึกหัด (Training Set) จากนั้นส่วนของการเรียนรู้จะทำการอุปมาณกฎจากตัวอย่างเหล่านั้นโดยตัวมันเองและให้ผลลัพธ์ออกมาในรูปของต้นไม้แห่งการตัดสินใจ ผู้เชี่ยวชาญสามารถที่จะตรวจสอบว่าต้นไม้แห่งการตัดสินใจที่ได้มาสามารถวินิจฉัย ปัญหาตามที่ตนต้องการได้หรือไม่ ถ้ายังไม่ได้ ก็

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

023241

สามารถให้ชุดตัวอย่างใหม่อีกครั้งหรือเพิ่มชุดตัวอย่างในชุดคำสั่งแล้วทดสอบจนกว่าจะได้กฎที่น่าพอใจ ต้นไม้แห่งการตัดสินใจที่ได้สามารถที่จะถูกแปลงให้อยู่ในรูปของโพรดักชั่นรูล เก็บไว้ในฐานความรู้เพื่อใช้งานต่อไป

3.2.1 การเรียนรู้จากตัวอย่าง (Learning by Examples)

วิธีการเรียนรู้จากตัวอย่างนั้น ตัวอย่างแต่ละตัวอย่างที่ให้โดยผู้เชี่ยวชาญจะต้องประกอบด้วยชนิดหรือประเภท (Class) ของตัวอย่างและแอทริบิวท์ (Attributes) ที่ใช้อธิบายตัวอย่างนั้น แอทริบิวท์ที่กล่าวถึงนี้จะต้องเป็นคุณสมบัติที่สำคัญที่มีผลในการแยกประเภท เช่น ลักษณะ อาการ รูปร่าง อายุ ความสามารถต่าง ๆ เป็นต้น ส่วนค่าที่เป็นไปได้ของเรียกว่า ค่า (Value) ซึ่งจะถูกนำมาใช้ในการพิจารณาเพื่อแยกแยะประเภทของตัวอย่างของการแก้ปัญหา (Classification) เช่น ในการวินิจฉัยโรคของแพทย์ แพทย์จะใช้อาการของผู้ป่วยเป็นแอทริบิวท์ ในการแบ่งแยกประเภทของโรค เช่น อาการไข้ อาการไอ น้ำมูก อาการปวดศรีษะ เป็นต้น เช่น ค่าของแอทริบิวท์อาการไข้ก็อาจจะเป็นเล็กน้อย (มีไข้เล็กน้อย) สูง (มีไข้สูง) ไม่มี (ไม่มีไข้) ค่าของ แอทริบิวท์ น้ำมูกก็อาจเป็น ใส (น้ำมูกใส) ข้น (น้ำมูกข้น) ไม่มี (ไม่มีน้ำมูก) เป็นต้น รูปที่ 3.2 เป็นการแสดงกลุ่มตัวอย่างชุดฝึกหัดของการวินิจฉัยโรคอย่างง่ายที่ให้โดยแพทย์ผู้เชี่ยวชาญเพื่อให้ส่วนของการเรียนรู้นำไปอุปมานกฎซึ่งอยู่ในรูปของต้นไม้สำหรับการตัดสินใจ

ตัวอย่าง	อาการไข้	อาการไอ	น้ำมูก	อ่อนเพลีย	เจ็บคอ	ปวดศรีษะ	ท้อง	จุดแดง	ประเภทของโรค
1	มี	เล็กน้อย	ไม่มี	เล็กน้อย	เล็กน้อย	รอบ ๆศรีษะ	ปกติ	ไม่มี	โรค A
2	มี	เล็กน้อย	มี	เล็กน้อย	เล็กน้อย	รอบ ๆศรีษะ	ปกติ	ไม่มี	โรค A
3	มี	เล็กน้อย	ไม่มี	เล็กน้อย	มาก	รอบ ๆศรีษะ	ปกติ	ไม่มี	โรค B
4	มี	ไอแห้ง ๆ	มี	มาก	เล็กน้อย	รอบ ๆศรีษะ	จุกแน่น	ไม่มี	โรค C
5	มี	ไอแห้ง ๆ	ไม่มี	มาก	เล็กน้อย	รอบ ๆศรีษะ	ปกติ	ตามตัว	โรค D
6	ไม่มี	ไม่มี	ไม่มี	เล็กน้อย	เล็กน้อย	ข้างเดียว	ปกติ	ไม่มี	โรค E
7	ไม่มี	ไม่มี	ไม่มี	เล็กน้อย	เล็กน้อย	รอบ ๆศรีษะ	ปกติ	ไม่มี	โรค F

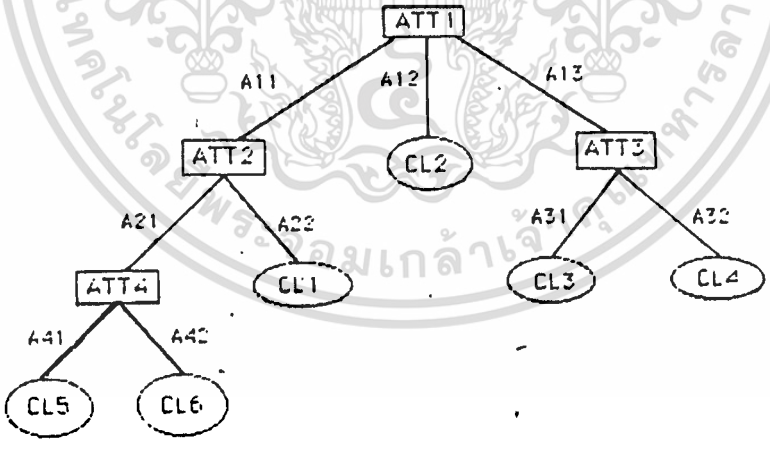
รูปที่ 3.2 แสดงตารางชุดฝึกหัดของตัวอย่างที่ใช้ในการวินิจฉัยโรคอย่างง่าย

ขบวนการที่ใช้ในการเรียนรู้ก็คือ จะนำเอาตัวอย่างจากชุดฝึกหัดไปผ่านขั้นตอนของการอุปมาน (Induction) ซึ่งได้ผลลัพธ์ออกมาเป็นต้นไม้สำหรับการตัดสินใจในกรณีที่ตัวอย่าง 2 ตัวอย่างที่มีประเภท (Class) แตกต่างกัน แต่กลับอธิบายด้วยชุดของค่าของแอทริบิวท์ที่เหมือนกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบจะไม่สามารถนำไปอุปมานกฎได้ เนื่องจากตัวอย่างใด ๆ ก็ตามที่มีประเภทแตกต่างกัน ตัวอย่างเหล่านั้น ย่อมจะอธิบายด้วยชุดของแอทริบิวท์ที่มีค่าที่ต่างกันด้วยเสมอ ถ้าเกิดกรณีดังกล่าวนี้ขึ้นแสดงว่า แอทริบิวท์ที่ผู้เชี่ยวชาญกำหนดไว้ยังไม่พอเพียงจำเป็นจะต้องเพิ่มแอทริบิวท์ใหม่ขึ้นอีกในระบบ เล็กซ์เซลล์ เพื่อเป็นการป้องกันกรณีดังกล่าวระบบจะทำการตรวจชุดของตัวอย่างนั้นเสียก่อน เพื่อไม่ให้เกิดกรณีนี้ขึ้นมา นอกจากนี้การกำหนดแอทริบิวท์ต่าง ๆ เพื่ออธิบายตัวอย่างของผู้เชี่ยวชาญนั้นก็มีความสำคัญ เพราะว่าถ้าหากผู้เชี่ยวชาญกำหนดไม่สมบูรณ์แล้ว ก็อาจจะทำให้การวินิจฉัยผิดพลาดได้ ดังนั้นผู้เชี่ยวชาญจะต้องคำนึงถึงสิ่งนี้ทุกครั้งก่อนที่จะให้ชุดของตัวอย่าง

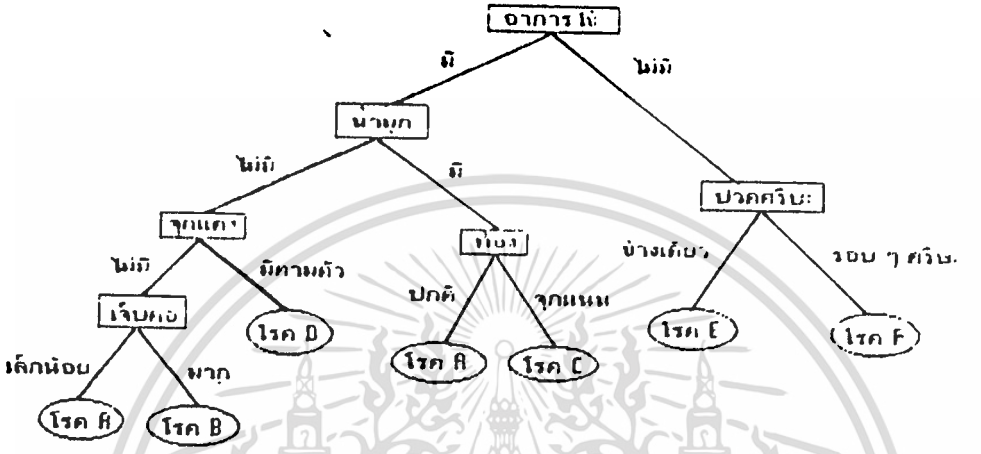
ลักษณะของต้นไม้สำหรับการตัดสินใจที่อุปมานได้จะประกอบด้วยแอทริบิวท์, ค่าของของแอทริบิวท์และประเภทของตัวอย่างในชุดฝึกหัดโดยที่แต่ละแอทริบิวท์จะแทนโนดของต้นไม้ และแต่ละกิ่งที่แยกออกจากโนดของจะถูกกำกับโดยค่าของแอทริบิวท์นั้น แล้วจากกิ่งก็อาจจะเชื่อมต่อไปยังโนดอื่น (แอทริบิวท์อื่น) ตัก จนกว่าจะถึงปลายสุดกิ่ง (Leaf) ของต้นไม้ ซึ่งแทนประเภทของตัวอย่าง ดังรูปที่ 3.3



รูปที่ 3.3 แสดงลักษณะของต้นไม้สำหรับการตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และจากตัวอย่างในชุดฝึกหัด ซึ่งเป็นตัวอย่างของการวินิจฉัยโรคอย่างง่ายสามารถแสดงการแปลประเภทของโรคด้วย ต้นไม้สำหรับการตัดสินใจ โดยอาศัยแอทริบิวท์ค่าของและประเภทของตัวอย่างได้รูปที่ 3.4



รูปที่ 3.4 แสดงต้นไม้สำหรับการตัดสินใจของการวินิจฉัยโรคอย่างง่าย

ต้นไม้สำหรับการตัดสินใจที่ได้จากการอุปมาน สามารถใช้วินิจฉัยปัญหาได้เช่นเดียวกัน การวินิจฉัยจะเริ่มตั้งแต่รูทโนด (Root Node) ของต้นไม้ โดยทำการทดสอบแอทริบิวท์ที่เป็นรูทโนด ด้วยการถามผู้ใช้ เพื่อให้ป้อนค่าของแอทริบิวท์นั้นเข้ามา แล้วค่าที่ได้จะถูกนำมาเลือกกิ่งเพื่อเชื่อมโยงไปทดสอบแอทริบิวท์ที่แทนโนดซึ่งเชื่อมต่อกับกิ่งนั้นต่อไป ทำไปเช่นนี้จนกว่าจะพบประเภท (Class) ที่กิ่งปลายสุดกิ่งของต้นไม้ ซึ่งจะเป็นข้อสรุปที่ได้จากการวินิจฉัยปัญหา

ต้นไม้สำหรับการตัดสินใจที่ได้จากขบวนการเรียนรู้สามารถที่จะวินิจฉัยแยกแยะประเภทของตัวอย่างได้ครอบคลุมทั้งชุดฝึกหัด (Training Set) และยังสามารถใช้เป็นหลักเกณฑ์ในการวินิจฉัยปัญหาประเภทเดียวกันได้ ซึ่งขึ้นอยู่กับว่า ผู้เชี่ยวชาญจะสามารถให้ตัวอย่างได้ครอบคลุมการวินิจฉัยปัญหาประเภทนั้น ๆ ได้ขนาดไหนจะพบว่าสำหรับปัญหามีความซับซ้อน ซึ่งแต่ละตัวอย่างจะต้องอธิบายหรือบรรยายด้วยแอทริบิวท์จำนวนมากแล้ว ต้นไม้สำหรับการตัดสินใจที่อุปมานได้จะมีขนาดใหญ่ เนื่องจากมีจำนวนของแอทริบิวท์ซึ่งเป็นโนดของต้นไม้เป็นจำนวนมากนั่นเอง ดังนั้นเพื่อให้ได้ต้นไม้ที่มีขนาดเล็กที่สุดเท่าที่จะเป็นไปได้ (ถามน้อยคำถามที่สุดเพื่อให้ได้คำตอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยเร็ว) จะต้องพยายามเลือกแอทริบิวท์ ที่เหมาะสมมาใช้ในการสร้างต้นไม้

วิธีการสร้างต้นไม้ โดย ไอดีทรี [2, 3, 4] สามารถให้ต้นไม้ที่มีขนาดเล็กแต่ยังไม่รับรองว่าเป็นต้นไม้ที่มีขนาดเล็กที่สุดอย่างไรก็ตามก็ยังนับว่าเป็นวิธีที่เหมาะสมมาก ในโครงงานนี้ ได้เลือกใช้ ไอดีทรี ในการอุปมานต้นไม้ ในส่วนของการเรียนรู้จากตัวอย่าง

3.2.2 ไอดีทรี (ID3)

วิธีการอุปมานกฎจากตัวอย่างหรือกล่าวอีกอย่างหนึ่งว่าวิธีการสร้าง ต้นไม้สำหรับการตัดสินใจ จาก ชุดฝึกหัด โดยวิธีการ ไอดีทรี นั้น ค้นคิดโดย J.R.Quinlan โดยพัฒนามาจาก CLS (Concept Learning System) ของ E.B. Hunt [5,6] ซึ่งเป็นวิธีการเรียนรู้ในสมัยต้น ๆ ประมาณ ค.ศ. 1963

ขั้นตอนในการสร้างต้นไม้ โดยวิธีของ ไอดีทรี มีดังต่อไปนี้ [4]

1. ทำการเลือกกลุ่มของตัวอย่างบางส่วนจาก ชุดฝึกหัด ขึ้นมา 1 กลุ่ม อย่างสุ่มโดยจะเรียกกลุ่มของตัวอย่างที่เลือกมานี้ว่า วินโดว์ (Window)

2. สร้างต้นไม้สำหรับการตัดสินใจจาก วินโดว์ นั้น ซึ่งต้นไม้ที่ได้ จะสามารถแยกประเภทตัวอย่างได้ครอบคลุมวินโดว์ทั้งหมด สำหรับรายละเอียดของการสร้างต้นไม้ จะกล่าวถึงในหัวข้อต่อไป

3. นำ ต้นไม้ ที่ได้จากข้อ 2 ไปลองทดสอบแยกประเภท ตัวอย่างในชุดฝึกหัดที่ไม่อยู่ในวินโดว์ทีละตัวอย่าง การใช้ต้นไม้ทดสอบแยกประเภทตัวอย่างก็โดยพิจารณาดูว่าตัวอย่างที่กำลังทดสอบนั้นมีประเภทเช่นเดียวกับการใช้ต้นไม้ในการวินิจฉัยประเภทของตัวอย่างหรือไม่ ถ้าได้ประเภทเดียวกันแล้วก็แสดงว่าต้นไม้นั้นสามารถแยกประเภทตัวอย่างนั้นได้ถูกต้อง มิฉะนั้นก็จะถือว่า แยกประเภท ไม่ได้ ถ้าทดสอบแล้วสามารถ แยกประเภท ได้ถูกต้องก็จะทยอยเอาตัวอย่างที่เหลือ ไปลองทดสอบแยกประเภทอีกจนกว่าจะหมด เมื่อลองทดสอบจนไม่เหลือตัวอย่างที่จะให้ทดสอบอีกแล้ว ก็จะได้ ต้นไม้ ที่ใช้ทดสอบนั้นเป็นต้นไม้ที่ได้จากการอุปมานจาก ชุดฝึกหัด ทั้งหมด ซึ่งก็จะสิ้นสุดขบวนการของการสร้าง ต้นไม้ มิฉะนั้นแล้วจะนำเอาตัวอย่างที่แยกประเภทแล้วย้อนไปทำขั้นตอนที่ 2 ใหม่อีก

ขบวนการต่าง ๆ ในการสร้างต้นไม้จากชุดฝึกหัด แสดงได้ดังโฟลว์ชาร์ต รูปที่ 3.5

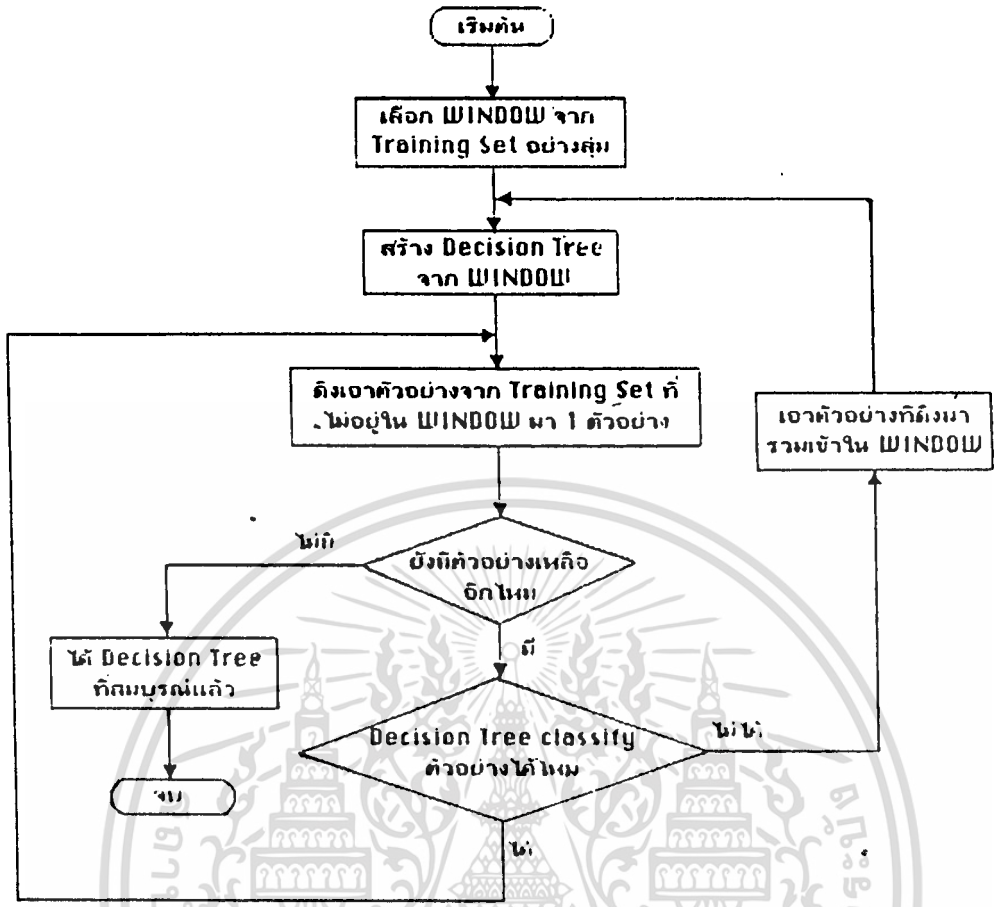
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 การสร้างต้นไม้สำหรับการตัดสินใจโดยวิธีของไอดีทรี

ไอดีทรี ได้ประยุกต์ใช้ทฤษฎีข่าวสาร (Information Theory) [7] ในการเลือกแอทริบิวท์เพื่อสร้างต้นไม้ ซึ่ง ต้นไม้ ที่ได้ จะมีขนาดเล็ก จากที่ได้กล่าวถึงลักษณะของ ต้นไม้ ที่ใช้ในการวินิจฉัยประเภทของตัวอย่างแล้วว่า แต่ละ โนด จะแทนด้วย แอทริบิวท์และกิ่งแทนด้วย ค่าของ แอทริบิวท์ ส่วนปลายกิ่งสุดท้าย (Leaf) ก็จะแทน ประเภท ของตัวอย่าง ถ้าพิจารณาโดยอาศัยทฤษฎีข่าวสารแล้วจะเห็นว่า โครงสร้างของต้นไม้ก็มีลักษณะคล้ายกันกับการเก็บข้อมูลสำหรับถามคำถามเพื่อค้นหา ข่าวสาร นั่นเองโดยที่ แอทริบิวท์ต่าง ๆ ที่แทน โนด ของ ต้นไม้ จะเป็นเสมือนคำถามที่ถามเพื่อรับเอาข่าวสาร ซึ่งคำตอบที่ได้รับมาก็จะเป็นค่าของแอทรบิวิท ์ ทจี่ะนาไปเลอกถามคาถามตอไป ซงกเป็นการถามเพอหาคาของแอทรบิวิทถดไป จนกว่าจะพบ ข่าวสาร ซึ่งก็คือ ประเภท ของตัวอย่างนั่นเอง

การสร้าง ต้นไม้ ให้มีขนาดเล็กนั้นขึ้นอยู่การเลือก แอทริบิวท์ เพื่อมาเป็น โนด ของ ต้นไม้ เป็นสำคัญ การเลือก แอทริบิวท์ ที่ไม่ดีจะทำให้เกิดมี แอทริบิวท์ ที่ซ้ำกันเกิดขึ้นเป็น โนดในต้นไม้หลายแห่งต้นไม้ที่ได้ก็จะมีขนาดใหญ่การเลือก แอทริบิวท์ ที่เหมาะสมจะทำให้จำนวนคำถามน้อยลง ลดการซ้ำซ้อนกันของ แอทริบิวท์ ใน ต้นไม้ ซึ่งทำให้ ต้นไม้มี ขนาดเล็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 แสดงโฟล์วชาร์ตของการสร้างต้นไม้สำหรับการตัดสินใจ

จากทฤษฎีข่าวสาร จะเห็นว่าขนาดของ ต้นไม้ นั้นขึ้นอยู่กับปริมาณของ ข่าวสาร ที่ ต้นไม้ นั้นเก็บอยู่ ถ้าปริมาณของข่าวสารมากย่อมหมายถึงว่าจะต้องใช้จำนวนคำถามมากจึงจะสามารถค้นพบข่าวสารที่ต้องการได้ซึ่งส่งผลให้ต้นไม้มีขนาดใหญ่และในทางตรงกันข้าม ถ้าปริมาณข่าวสาร ที่ ต้นไม้เก็บอยู่มีจำนวนน้อย ต้นไม้ ก็จะมีขนาดเล็ก ซึ่งหมายความว่าถามคำถามจำนวนน้อยคำถามก็จะได้คำตอบที่ต้องการ เพราะฉะนั้นจากชุดฝึกหัด ซึ่งใช้เป็นตัวอย่างชุดเดียวกันเราจึงพยายามที่จะทำให้ได้ต้นไม้ที่เล็กที่สุด

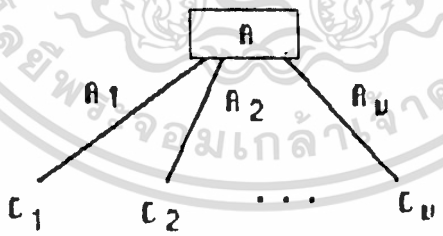
วิธีการสร้าง ต้นไม้ โดยวิธีของ ไอดีทรี ที่จะกล่าวถึงต่อไปนี้ จะสมมุติว่า ประเภท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของตัวอย่างใน ชุดฝึกหัด มีเพียง 2 ประเภท เท่านั้น คือ ประเภท "P" กับ ประเภท "N" ดังนั้นการ แยกประเภท ตัวอย่างใด ๆ ก็ตามความน่าจะเป็นที่ตัวอย่างนั้นจะอยู่ใน ประเภท "P" มีค่าเท่ากับ และความน่าจะเป็นที่ตัวอย่างนั้นจะอยู่ใน ประเภท "N" ก็จะมีค่าเท่ากับ โดยที่ p,n แทนจำนวนของตัวอย่างทั้งหมดใน ชุดฝึกหัด ที่มี ประเภท เป็น "P" และ "N" ตามลำดับ ดังนั้นต้นไม้สำหรับการตัดสินใจที่จะกล่าวถึงนี้สามารถให้ข่าวสารที่เป็น "P" หรือ "N" เท่านั้น ซึ่งค่า ข่าวสาร ที่คาดว่าจะใช้ในการให้ข่าวสาร (Explcted Information) "P" หรือ "N" จากต้นไม้นี้จะมีค่าตามสมการต่อไปนี้

$$I(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$

ถ้าให้ แอทริบิวท์ A ซึ่งมีค่า (Value) ต่าง ๆ คือ $[A_1, A_2, A_3, \ldots\ldots, A_v]$ ขณะเดียวกันก็ให้ แอทริบิวท์ เป็น รูทโนด ของ ต้นไม้ ค่าของ แอทริบิวท์ A จะแยก ต้นไม้ ออกเป็น ต้นไม้ย่อย (Subtree) คือ $[C_1, C_2, C_3, \ldots, C_v]$ โดยที่ C_i หมายถึง ต้นไม้ย่อย ใด ๆ ที่ตัวอย่างที่มีค่าของ แอทริบิวท์ A เป็น Ai ดังรูปที่ 3.6



รูปที่ 3.6 แสดงการแยกของต้นไม้ย่อย โดยค่าของแอทริบิวท์ A

สมมุติว่า ต้นไม้ย่อย C_i นั้นมีตัวอย่างที่มี ประเภทเป็น "P" อยู่ p_i ตัวอย่างและมีตัวอย่างที่มี ประเภท เป็น "N" อยู่ n_i ตัวอย่างดังนั้นค่า ข่าวสาร ที่คาดว่าจะได้รับจาก ต้นไม้ย่อย C_i จะเป็น $I(p_i, n_i)$ จากนั้น การคำนวนค่า ข่าวสาร ที่คาดว่าจะได้รับจาก ต้นไม้ ทั้งหมดโดยที่มี แอทริบิวท์ A เป็น รูทโนด จะสามารถหาได้จากการเฉลี่ยค่า ข่าวสาร ที่คาดว่าจะได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รับจากทุก ต้นไม้ย่อย $[C_1, C_2, C_3, \ldots . C_v]$ ดังสมการต่อไปนี้

$$E(A) = \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I(p_i, n_i)$$

จากความจริงที่ว่า ถ้าต้นไม้สำหรับการตัดสินใจต้นใด มีค่า ข่าวสาร น้อยก็แสดงว่า ต้นไม้นั้น มีขนาดเล็ก ดังนั้นการเลือก แอทริบิวท์ ที่จะมาเป็น รูทโนด ของ ต้นไม้ จึงอาศัยความจริงในข้อนี้ โดยจะทำการเลือก แอทริบิวท์ใด ๆ ก็ตามที่จะมาเป็น รูทโนด ของ ต้นไม้ หรือ ต้นไม้ย่อย ที่ทำให้ค่า ข่าวสาร ที่คาดว่าจะได้รับจาก ต้นไม้ หรือต้นไม้ย่อยคือค่าของ E(A) มีค่าน้อยที่สุดเมื่อเทียบกับค่า E(A) ของ แอทริบิวท์อื่น ๆ ทั้งนี้เพื่อเป็นการรับประกันว่า ต้นไม้ หรือ ต้นไม้ย่อยที่ได้จากการเลือกแอทริบิวท์นั้นเป็นรูทโนดเป็นต้นไม้ที่มีขนาดเล็ก

สำหรับขั้นตอนการสร้าง ต้นไม้ ของ ไอดีทรี จะเริ่มจากการคำนวนหาค่า E(A) ของแต่ละ แอทริบิวท์ ใน ชุดฝึกหัด แล้วเลือกเอาเฉพาะ แอทริบิวท์ ที่ให้ค่า E(A) น้อยที่สุด มาเป็น รูทโนด ของ ต้นไม้ จากนั้นแบ่ง ชุดฝึกหัด ตามกลุ่มของตัวอย่างที่มีค่าตามค่าของแอทริบิวท์ที่เลือกมานั้นออกเป็น ชุดฝึกหัดย่อย ๆ ซึ่งชุดฝึกหัดย่อย ๆ ที่แยกโดยใช้ $A_1, A_2, \ldots . A_v$ ก็จะเป็นของ ต้นไม้ย่อย $C_1, C_2, \ldots . . . . C_v$ ตามลำดับเช่น ชุดฝึกหัดของ C_1 ก็จะเป็น กลุ่มของตัวอย่างทั้งหมดจาก ชุดฝึกหัด เดิมที่มี ค่าของ แอทริบิวท์ เป็น A_1 เป็นต้น (ดูรูปที่ 3.6 ประกอบ)

จากนั้นทำการสร้าง ต้นไม้ย่อย $C_1, C_2, \ldots . C_v$ โดยใช้ ชุดฝึกหัดย่อย ๆ ของแต่ละ ต้นไม้ย่อย ด้วยวิธีการเดียวกัน แต่ตอนนี้การเลือก แอทริบิวท์ ที่จะมาเป็น รูทโนด ของแต่ละ ต้นไม้ย่อย จะไม่พิจารณา แอทริบิวท์ A อีกต่อไป เมื่อเลือก แอทริบิวท์ที่จะมาเป็น รูทโนด ของ ต้นไม้ย่อยได้แล้ว ก็สามารถที่จะแยกออกเป็น ต้นไม้ย่อยลงไปอีก ทำเช่นนี้ไปเรื่อย ๆ จนกว่า ชุดฝึกหัด ของ ต้นไม้ย่อยใด ที่ตัวอย่างทั้งหมดเป็นประเภทเดียวกัน ตอนนี้สามารถที่จะแทนต้นไม้ย่อยนั้น ด้วย ประเภท (Class) นั้นเลย นั่นก็คือเราได้ปลายกิ่งสุดท้าย (Leaf) ของ ต้นไม้แล้วนั่นเอง

ตัวอย่างการสร้างต้นไม้สำหรับการตัดสินใจโดยวิธีการของไอดีทรี จากตารางชุดฝึกหัด ในรูปที่ 3.7 เป็นกลุ่มตัวอย่างอย่างง่ายที่ให้โดยผู้เชี่ยวชาญเพื่อต้องการอุปมาน ต้นไม้สำหรับการตัดสินใจ ที่ใช้ในการวินิจฉัยสภาพอากาศ แต่ละตัวอย่างจะอธิบายด้วย แอทริบิวท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Outlook, Temperature, Humidity และ Windy มีทั้งหมด 14 ตัวอย่าง ส่วน ประเภท (Class) ของตัวอย่างนั้น สมมุติให้มีเพียง 2 ประเภท เท่านั้นคือ ประเภท "P" และ "N"

การสร้างต้นไม้สำหรับการตัดสินใจ ก็เริ่มจากการเลือก แอทริบิวท์ เพื่อมาเป็นรูทโนดของ ต้นไม้ โดยคำนวนหาค่า E(A) ของแต่ละ แอทริบิวท์ จะได้ว่า

สำหรับ แอทริบิวท์ outlook ซึ่งมีค่าเป็น {Sunny,Overcast,Rain}

ตัวอย่างที่มีค่าของ แอทริบิวท์ Outlook เป็น Sunny มี ประเภท เป็น "P" และ "N" เป็นจำนวน 2 และ 3 ตัวอย่างตามลำดับ ดังนั้น

$P_1=2, n_1=3$ $I(P_1,N_1)=0.971$

และขณะเดียวกันสำหรับของ ค่าอื่น ๆ ของ แอทริบิวท์ Outlook

$P_2=4, n=0$ $I(P_2,N_2)=0$

$P_3=3, n=2$ $I(P_3,N_3)=0.971$

จากนั้นสามารถหาค่าของ E(outlook) ได้

โดย $E(\text{outlook}) = 5/14 * I(p_1,n_1)+4/14 * I(p_2,n_2)+ 5/14 * I(p_3,n_3)$

$= 0.694$ bits

สำหรับ E(A) ของ Attribute อื่น ๆ นั่นก็มีค่าดังนี้

E(temperature)= 0.911 bits

E(humidity) = 0.789 bits

E(windy) = 0.892 bits

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

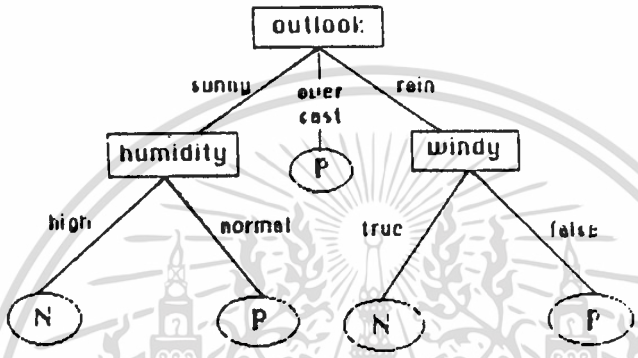
EX No.	ATTRIBUTES				CLASS
	outlook	temperature	humidity	windy	
1	sunny	hot	high	false	N
2	sunny	hot	high	true	N
3	overcast	hot	high	false	P
4	rain	mild	high	false	P
5	rain	cool	normal	false	P
6	rain	cool	normal	true	N
7	overcast	ccol	normal	true	P
8	sunny	mild	high	false	N
9	sunny	cool	normal	false	P
10	rain	mild	normal	false	P
11	sunny	mild	normal	true	P
12	overcast	mild	high	true	P
13	overcast	hot	normal	false	P
14	rain	mild	high	true	N

รูปที่ 3.7 แสดงตารางชุดฝึกหัดของตัวอย่างที่ใช้ในการวินิจฉัยสภาพอากาศ

โดยการพิจารณาค่า E(A) ของแต่ละแอทริบิวท์แล้ว จะเลือก แอทริบิวท์ outlook มาเป็น รูทโนด ของ ต้นไม้ เนื่องจากให้ค่า E(A) น้อยที่สุด

จากนั้นชุดฝึกหัด จะถูกแบ่งเป็น ชุดฝึกหัดย่อย ๆ ตามแต่ละ ค่าของ Outlook แล้ว แต่ละชุดฝึกหัดย่อย ๆ ก็จะถูกใช้สร้าง ต้นไม้ย่อย โดยวิธีการเดียวกันที่ได้แสดงไปแล้ว ต้นไม้สำหรับการตัดสินใจ ที่สมบูรณ์แล้วแสดงดังรูปที่ 3.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 แสดงต้นไม้สำหรับการตัดสินใจที่สร้างเสร็จสมบูรณ์แล้ว

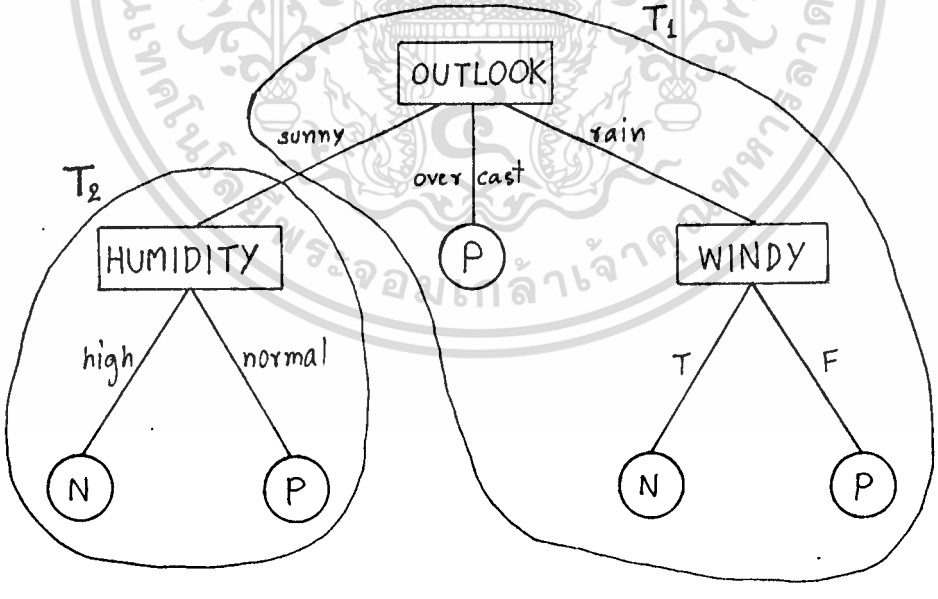
3.3 การปรับปรุง ไอดีทรี

3.3.1 การปรับปรุง ไอดีทรี วิธีที่ 1

จากต้นไม้สำหรับการตัดสินใจในรูป 8 จะเห็นว่าเมื่อระบบใช้ต้นไม้ดังกล่าวในการให้การวินิจฉัยตัวเครื่องจักรอินเฟอเรนซ์จะต้องทำการทดสอบ attribute OUTLOOK ก่อนเสมอ แล้วจึงนำค่าของ attribute OUTLOOK จากผู้ใช้เข้าไปเพื่อเลือกทดสอบ attribute ต่อมา เป็นเช่นนี้จนกว่าตัวเครื่องจักรอินเฟอเรนซ์จะพบ class ซึ่งได้เป็นผลสรุปจากการวินิจฉัย จะเห็นว่าในการให้การวินิจฉัยแต่ละครั้งจะหลีกเหลี่ยงการทดสอบ attribute

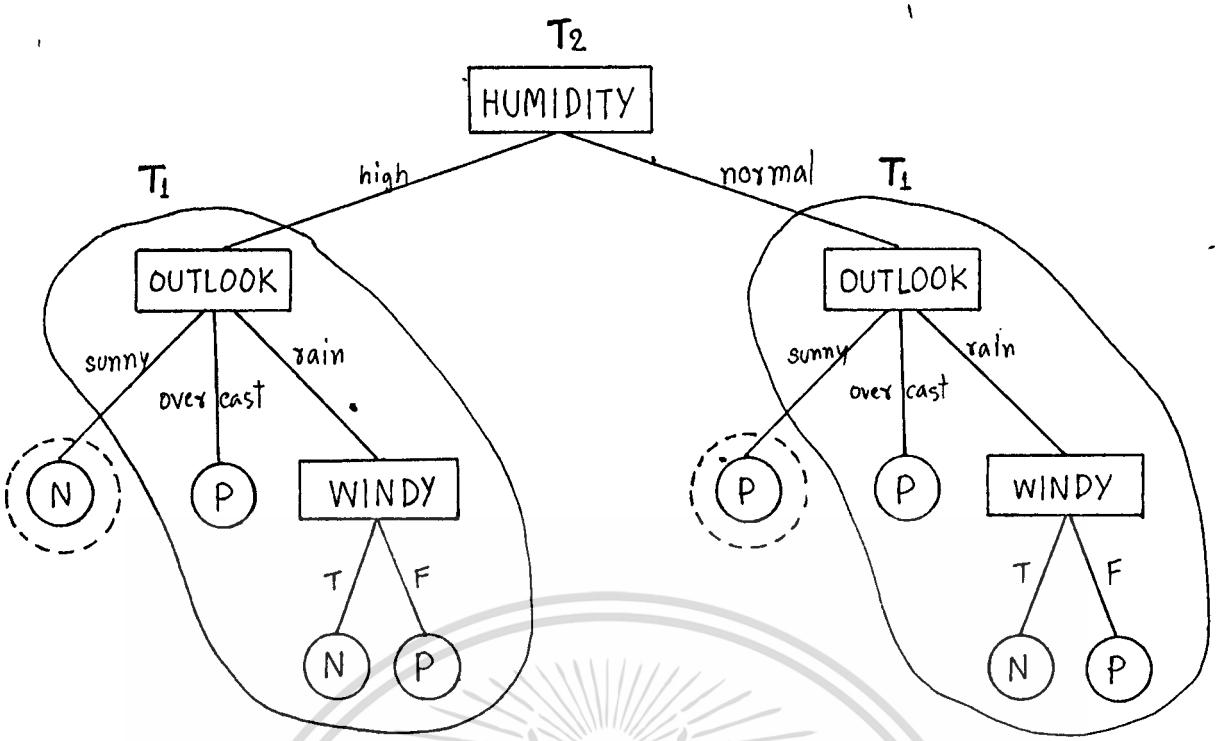
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

OUTLOOK ไม่ได้เลย สมมุติว่าถ้าการให้ค่าของ attribute OUTLOOK จากผู้ใช้แต่ละครั้ง ผู้ใช้จะต้องสูญเสียค่าใช้จ่ายเป็นจำนวนมากเพื่อให้ได้มาซึ่ง information ที่เป็นค่าของ attribute OUTLOOK ดังนั้นควรที่จะมีวิธีการที่หลีกเลี่ยงการทดสอบ attribute OUTLOOK ให้มากที่สุด แต่ระบบก็ยังมีหนทางในการวินิจฉัยโดยไม่จำเป็นต้องทดสอบ attribute OUTLOOK การทำเช่นนี้ได้ก็โดยการย้าย node ที่แทน attrobute OUTLOOK จาก root ของ Tree ลงมาโดยสลับกับ node ซึ่งแทน attribute ที่เป็นลูกในที่นี้คือ attribute HUMIDITY ดูรูปที่ 3.9 การสลับ attribute OUTLOOK กับ attribute HUMIDITY นั้นทำโดยการย้าย Subtree T_1 ซึ่งเป็น Subtree ที่ไม่นับรวม Subtree T_2 ซึ่งเป็น Subtree ที่มี attribute HUMIDITY เป็น root node มาต่อกับ branch ทุก ๆ branch ของ node attribute HUMIDITY สำหรับส่วนที่ต่อกับ branch นั้นอยู่ก่อนก็จะถูกนำเอาไปต่อกับ branch ที่ยังว่างอยู่ของ Subtree T_1 จากรูปที่ 3.9 เมื่อทำการสลับ node atrribute OUTLOOK กับ node attribute HUMIDITY จะได้ Tree ดังรูปที่ 3.10



รูปที่ 3.9 แสดงต้นไม้ย่อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แสดงการย้ายต้นไม้ย่อย

การเปลี่ยนแปลง Tree ในลักษณะดังกล่าวยึดถือหลักการดังต่อไปนี้

1. ถ้าทำการเปลี่ยน path

ในต้นไม้สำหรับการตัดสินใจในรูปที่ 3.9 ให้อยู่ในรูปกฎจะได้ดังนี้

IF THE VALVE OF ATTIBUTE OUTLOOK IS SUNNY

AND THE VALUE OF ATTIBUTE HUMIDITY IS HIGH

THEN THE OBJECT IS IN CLASS N

จะเห็นว่าเราสามารถสลับเงื่อนไขใน IF clause ได้ เนื่องจาก Conective AND เคารพกฎการสลับที่ ดังนั้นกฎสามารถถูกจัดให้อยู่ในรูปแบบใหม่โดยการสลับที่เงื่อนไขใน IF clauseได้เป็น

IF THE VALVE OF ATTIBUTE HUMIDITY IS HIGH

AND THE VALUE OF ATTIBUTE OUTLOOK IS SUNNY

THEN THE OBJECT IS IN CLASS N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

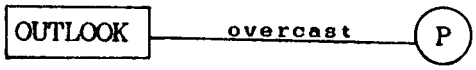
เมื่อพิจารณาเป็น path ของต้นไม้สำหรับการตัดสินใจก็จะได้เป็น

นั่นก็คือเป็นการสลับตำแหน่งของ node ที่แทน attribute OUTLOOK กับ node ที่แทน attribute HUMIDITY นั่นเอง การสลับเช่นนี้จะทำให้ node ของ attribute OUTLOOK ถูกเลื่อนลงมาระดับล่างของ Tree ทำให้ลดความสำคัญที่จะทำการทดสอบลงมา จนอาจที่จะเกิดบางกรณีที่ทำให้สามารถเปลี่ยนรูปของ node ที่แทน attribute OUTLOOK ไปเป็น class ได้

2. การย้าย attribute ที่ต้องการเลื่องลงมาอีกระดับหนึ่งของ Tree ทำได้โดยการสลับตัวมันกับ attribute ที่เป็นลูกของมันตัวใดตัวหนึ่ง จากตัวอย่างที่แล้วเราเลือกสลับ attribute OUTLOOK กับ attribute HUMIDITY เราจะกำหนด Subtree ขึ้นมา 2 Subtree Subtree แรกเรียกว่า Lower Subtree (T_l) ซึ่งเป็น Subtree ที่มี attribute ที่จะสลับกับ attribute ที่ต้องการเลื่อง แทนด้วย N_c (Child Node) เป็น root node อีก Subtree หนึ่งเป็น Subtree ที่มี attribute ที่ต้องการเลื่องซึ่งแทนด้วย N_a (Avoided Node) เป็นรูทโหนดเรียกว่า Upper Subtree (T_u) ซึ่งเป็น Subtree ที่ไม่นับรวม T_l ดังนั้นจึงมี branch หนึ่งว่างอยู่ จากรูปที่ 3.9 T_l คือ T_2 และ T_u คือ T_1 การสลับกันของทั้งสอง attribute ดังกล่าว เกิดจากการนำเอา T_u ไปเชื่อมต่อกับทุก ๆ branch ของ N_c สำหรับสิ่งเดิมที่ต่ออยู่กับแต่ละ branch นั้นจะนำไปต่อกับ branch ที่ว่างอยู่ของ T_u แทน เหมือนกับว่าการเชื่อมต่อของมันกับ branch เดิมถูกแทรกโดย T_u จากตัวอย่างเดิม (N) กับ (P) ซึ่งต่ออยู่กับ N_c คือ attribute HUMIDITY จะถูกแทรกโดย T_l ทั้งคู่ ดังรูปที่ 3.10 การทำเช่นนี้จะทำให้ลักษณะของ Tree เดิมในรูปที่ 3.9 เปลี่ยนแปลงไป ทำให้มีการสลับตำแหน่งของ N_a กับ N_c ตามที่ต้องการ เมื่อพิจารณา Tree รูปที่ 3.10 ในรูปกฎจะพบว่ากฎที่มีเงื่อนไขเป็นการทดสอบค่าของ attribute ที่เป็น N_a กับ N_c จะสลับกันเมื่อเทียบกับกฎเดิม เป็นไปตามหลักการในข้อ 1 สำหรับกฎอื่น ๆ จะเกิดมีเงื่อนไขการทดสอบค่าของ attribute ที่เป็น N_c ปรากฏในทุก ๆ กฎ แต่กฎเดิมนั้นยังมีความหมายเหมือนเดิม เนื่องจากกฎเดิมจะถูกสร้างเป็นกฎใหม่หลาย ๆ กฎ ซึ่งแต่ละกฎจะมีการทดสอบแต่ละค่าของ attribute ที่เป็น N_c ที่เป็นไปได้ทั้งหมด ดังนั้นจึงเหมือนกับว่ากฎใหม่เหล่านั้นไม่ขึ้นกับเงื่อนไขการทดสอบ attribute ที่เป็น N_c เลย เนื่องจากไม่ว่าค่าของ attribute ที่เป็น N_c จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นค่าใด กฎเหล่านั้นก็ยังมีความหมายเหมือนกับเดิมทุกประการ ตัวอย่างเช่น จาก tree เริ่มแรกในรูปที่ 3.9 path

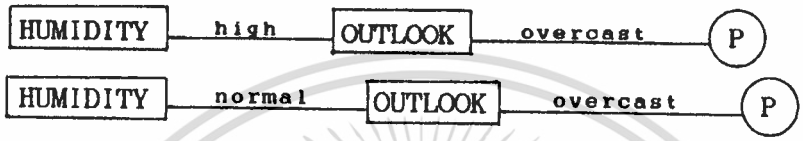


สามารถถูกเขียนให้อยู่ในรูปกฎได้เป็น

IF THE VALUE OF ATTRIBUTE OUTLOOK IS OVERCAST

THEN THE OBJECT IS IN CLASE P

หลังจากสลับ attribute OUTLOOK กับ attribute HUMIDITY แล้วจะได้ Tree ใหม่ในรูปที่ 3.10 ซึ่งมี path เป็น



และสามารถแสดงอยู่ในรูปกฎเป็น

IF THE VALUE OF ATTIBUTE HUMIDITY IS HIGH

AND THE VALUE OF ATTIBUTE OUTLOOK IS OVERCAST

THEN THE OBJECT IS IN CLASS P

และ

IF THE VALUE OF ATTIBUTE HUMIDITY IS NORMAL

AND THE VALUE OF ATTIBUTE OUTLOOK IS OVERCAST

THEN THE OBJECT IS IN CLASS P

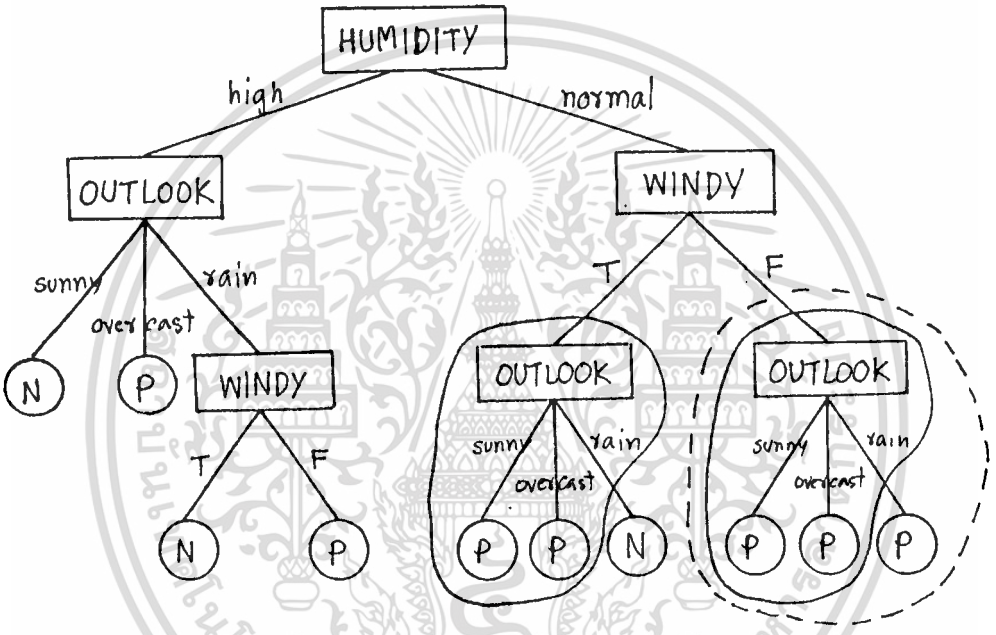
จะเห็นว่าไม่ว่าค่าของ attribute HUMIDITY จะเป็น high หรือ normal กฎก็จะยังสามารถสรุปว่าเป็น class P ถ้าค่าของ attribute OUTLOOK เป็น overcast นั่นคือ กฎจะไม่ขึ้นกับค่าของ attribute HUMIDITY แต่อย่างใด แสดงว่ากฎเดิมไม่มีการเปลี่ยนแปลง

วิธีการปรับปรุง ไอดีทรี จะทำการย้าย attribute ที่ต้องการเลี่ยงลงมาระดับล่าง โดยทำการสลับกับ attribute ลูกของมันตัวใดตัวหนึ่งจนกว่า attribute ที่ต้องการเลี่ยงนั้นมีลูกเป็น class เดียวกันทั้งหมด ก็จะสามารถแทนทั้ง attribute และลูก ๆ ของมันทั้งหมดนั้นด้วย class นั้นทันที จะทำให้ path ที่เคยต่ออยู่กับ attribute ที่ต้องการเลี่ยงที่สามารถแทนได้ด้วย class นั้นไม่ปรากฏ attribute ที่ต้องการเลี่ยงอีกต่อไป เป็นการหลีกเลี่ยงการทดสอบ attribute บางตัวได้ ซึ่งเป็นวิธีการแก้ไขข้อบกพร่องของวิธีการสร้าง Tree ของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไอดีทรี ดังที่ได้กล่าวมาแล้วข้างต้น

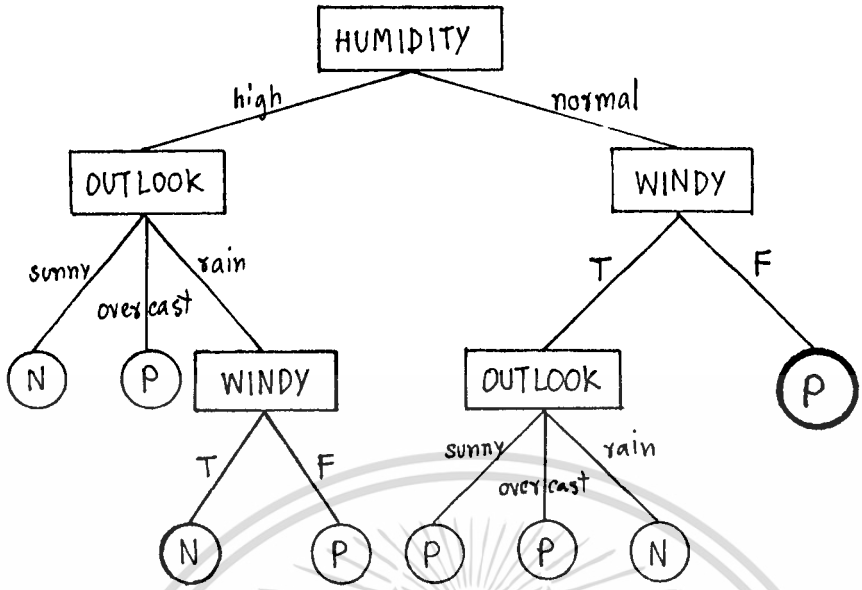
Tree ที่ได้จากรูป 3.10 ยังเป็น Tree ที่ปรับปรุงไม่เสร็จ ขั้นตอนต่อไปจะทำการย้าย attribute OUTLOOK ลงมาอีกโดยทำการสลับกับ attribute WINDY แต่ตอนนี้จะทำเฉพาะกับ attribute OUTLOOK ที่อยู่ทางด้านขวาเท่านั้นเนื่องจากการทำกับ attribute OUTLOOK ที่อยู่ทางด้านซ้ายไม่สามารถยุบ attribute OUTLOOK เป็น class ได้ ดังนั้นจึงไม่มีความจำเป็นในการทำเช่นนั้น จากการสลับกันของ attribute ทั้งสอง Tree ในรูปที่ 3.10 จะเปลี่ยนไปดังรูปที่ 3.11



รูปที่ 3.11 แสดงการย้ายต้นไม้ย่อยอีกครั้งหนึ่ง

Subtree ที่อยู่ในวงล้อมเส้นประในรูปที่ 3.11 สามารถแทนด้วย class P แสดงได้ดังรูปที่ 3.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.12 แสดงการยุบต้นไม้ย่อยเป็นประเภทของตัวอย่าง

จากรูปที่ 3.12 จะเห็นว่า Tree ที่ได้หลังจากการปรับปรุงแล้วจะสามารถหลีกเลี่ยงการทดสอบ attribute OUTLOOK ได้ ถ้าในการวินิจฉัยพบว่า Object นั้นมีค่าของ attribute HUMIDITY เป็น normal และค่าของ attribute WINDY เป็น F แล้วระบบก็จะสามารถให้ข้อสรุปว่าเป็น Object ที่อยู่ใน class P โดยไม่จำเป็นต้องทดสอบ attribute OUTLOOK ก่อนเมื่อเทียบกับการวินิจฉัย Tree ที่ได้จากวิธีของ ไอดีทรี การปรับปรุง Tree จะทำให้เกิดกฎใหม่เพิ่มขึ้นมาเป็นผลเนื่องมาจากการทำให้เกิด การซ้ำซ้อนกันขึ้นในเงื่อนไขของกฎเดิมนั่นเอง

จะเห็นว่าการปรับปรุง Tree จะทำได้ในบางกรณีเท่านั้น สังเกตได้โดยพิจารณา Tree ในรูปที่ 3.8 ที่ leaf ของ Subtree ที่มี attribute ที่ต้องการเลี่ยงเป็น root node (OUTLOOK) จะมี leaf เป็น class P กับทุก ๆ Subtree ที่ต่อมาจาก attribute ที่เราต้องการเลี่ยงการทดสอบ (OUTLOOK) กรณีเช่นนี้เมื่อย้ายตำแหน่งของ OUTLOOK ลงมาล่างสุดของ Tree จะสามารถแทน node ที่แทน attribute OUTLOOK ด้วย class P ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 การปรับปรุงไอดีทรีวิธีที่ 2

เนื่องจากวิธีการของไอดีทรีไม่สามารถอุปมาณชุดฝึกหัดบางชุดได้ ในกรณีที่การแบ่งชุดฝึกหัดเป็นชุดย่อยๆ แล้ว มีบางค่าของแอทริบิวท์ไม่ปรากฏในชุดฝึกหัดย่อยๆ นั้น ดังเช่นตัวอย่างการสร้างต้นไม้จากชุดฝึกหัดต่อไปนี้

HUMAN	EYE_LAYER	SKIN	HAIR	HEIGHT	EYE_COLOR	NOSE	CLASS
1	1_layer	yellow	black	middle	black	sharp	asian
2	1_layer	white	black	short	black	sharp	asian
3	1_layer	white	black	middle	black	sharp	asian
4	2_layer	yellow	black	middle	black	sharp	asian
5	2_layer	yellow	black	middle	brown	sharp	asian
6	2_layer	yellow	black	middle	black	flat	asian
7	2_layer	white	brown	tall	blue	sharp	asian
8	2_layer	black	black	tall	black	sharp	western
9	2_layer	white	black	tall	blue	sharp	western
10	2_layer	yellow	black	tall	blue	sharp	western
11	2_layer	black	black	tall	black	flat	african
12	2_layer	black	black	short	black	flat	african

สามารถแสดงการสร้างต้นไม้โดยวิธีการที่ปรับปรุงจาก ไอดีทรี ได้ดังนี้

จากชุดฝึกหัดนี้ เป็นตัวอย่างของการแบ่งประเภทของชาวเอเชีย ชาวตะวันตก และชาวแอฟฟริกา โดยอาศัยลักษณะทางกายภาพเป็นแอทริบิวท์ในการแบ่งประเภทดังนี้ EYE_LAYER บอกว่ามีตาชั้นเดียว (1_layer) หรือสองชั้น (2_layer), SKIN บอกสีผิว ในที่นี้มีผิวเหลือง ผิวขาว และผิวดำ, HAIR บอกสีผม ซึ่งมีสีดำ สีน้ำตาล และสีบรอนซ์, HEIGHT มี 3 ระดับ คือ เตี้ย ปานกลาง และสูง ส่วนสีตา (EYE_COLOR) ก็มีสีดำ สีน้ำตาล และสีฟ้า และสุดท้าย NOSE มีค่าเป็นจมูกโด่ง และจมูกแบน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสร้างต้นไม้ก็ทำได้เช่นเดียวกับวิธีการของไอดีทรี โดยเริ่มจากการเลือกแอทริบิวท์เพื่อมาเป็นรูทโหนดของต้นไม้เสียก่อน ซึ่งต้องหาค่า E(A) ของแต่ละแอทริบิวท์ ในที่นี้ค่า E(A) หาได้จาก

$$E(A) = \sum_{i=1}^{V} \frac{a_i + w_i + af_i}{a + w + af} * I(a_i, w_i, af_i);$$

โดยที่

$$I(a,w,af) = -\left(\frac{a}{a+w+af} \log_2 \frac{a}{a+w+af} + \frac{w}{a+w+af} \log_2 \frac{w}{a+w+af} + \frac{af}{a+w+af} \log_2 \frac{af}{a+w+af}\right)$$

a,w และ af คือ จำนวนตัวอย่างที่มีประเภทเป็น asian, western และ african ตามลำดับ

พิจารณาที่แอทริบิวท์ EYE_LAYER ซึ่งมีค่าเป็น {1_layer,2_layer}

$a_1 = 3, w_1 = 0, af = 0$ $I(a_1,w_1,af_1) = 0$

$a_2 = 3, w_2 = 4, af = 2$ $I(a_2,w_2,af_2) = 1.53$

$E(EYE_LAYER) = 3/12\ I(a_1,w_1,af_1) + 9/12\ I(a_2,w_2,af_2)$

$= 1.148$ bits

E(SKIN) = 0.864 bits

E(HAIR) = 1.142 bits

E(HEIGHT) = 0.468 bits

E(EYE_COLOR) = 0.866 bits

E(NOSE) = 0.980 bits

เนื่องจาก E(HEIGHT) ต่ำที่สุดจึงเลือกแอทริบิวท์ HEIGHT เป็นรูทโหนดของต้นไม้ จะพบว่าจากขบวนการที่ผ่านมายังใช้วิธีการสร้างต้นไม้แบบไอดีทรีได้ แต่หลังจากนี้ต่อไปจะพบข้อจำกัดของ ไอดีทรี

เมื่อได้รูทโหนดคือแอทริบิวท์ HEIGHT แล้วใช้ค่า short, middle และ tall แบ่งชุดฝึกหัดเป็นชุดย่อยตามค่าเหล่านี้ จะพบว่าตัวอย่างในชุดฝึกหัดย่อยที่แยกโดยใช้ค่า middle นั้นมีประเภทเป็น asian ทั้งหมด ดังนั้นเราแทนส่วนปลายกิ่งสุดท้ายได้ด้วยประเภท asian เชื่อมต่อกับกิ่ง middle ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดฝึกหัดย่อยที่ใช้ค่า short แยกออกมาเป็นดังนี้

HUMAN	EYE_LAYER	SKIN	HAIR	HEIGHT	EYE_COLOR	NOSE	CLASS
2	1_layer	white	black	short	black	sharp	asian
12	2_layer	black	black	short	black	flat	african

จากชุดฝึกหัดย่อยนี้สามารถสร้างเป็นต้นไม้ย้อยที่เชื่อมต่อกันถึง short ได้ดังนี้

เลือกแอทริบิวท์รูทโหนดคำนวณ E(A) ของแอทริบิวท์ที่เหลือ เริ่มพิจารณาที่แอทริบิวท์ EYE_LAYER มีค่าเป็น {1_layer, 2_layer}

$a_1 = 1, w_1 = 0, af = 0$ $I(a_1, w_1, af_1) = 0$

$a_2 = 0, w_2 = 0, af = 1$ $I(a_2, w_2, af_2) = 0$

E(EYE_LAYER) = 0 bits

สำหรับแอทริบิวท์ SKIN มีค่าเป็น {yellow, white, black}

$a_1 = 0, w_1 = 0, af_1 = 0$ $I(a_1, w_1, af_1) = \infty$ (หาค่าไม่ได้)

$a_2 = 1, w_2 = 0, af_2 = 1$ $I(a_2, w_2, af_2) = 0$

$a_3 = 0, w_3 = 0, af_3 = 1$ $I(a_3, w_3, af_3) = 0$

กรณีที่ $I(a_1, w_1, af_1)$ หาค่าไม่ได้ เกิดเนื่องมาจากว่าชุดแบบฝึกหัดย่อยนี้ไม่มีตัวอย่างไหนที่มีค่าของแอทริบิวท์ SKIN เป็น yellow อยู่เลย ไอดีทรี จะไม่สามารถสร้างต้นไม้ได้อีกต่อไป การแก้ไขทำได้โดยการตัดค่า yellow ออกจากค่าที่เป็นไปได้ของแอทริบิวท์ SKIN ก่อนที่จะใช้วิธีการเดิมของไอดีทรี โดยให้ไอดีทรีมองเห็นว่าค่าที่เป็นไปได้ของแอทริบิวท์ SKIN ในขณะนี้มีเพียง white และ black เท่านั้น ซึ่งขบวนการสร้างต้นไม้โดย ไอดีทรี ก็จะสามารถทำงานได้ต่อไป ซึ่งตอนนี้จะมีเพียง $I(a_2, w_2, af_2)$ และ $I(a_3, w_3, af_3)$ ซึ่งเป็น 0 ทั้งคู่

จะได้ E(SKIN) = 0 bits

สำหรับที่แอทริบิวท์ HAIR ก็เช่นกัน ค่อ brown และ bronze จะถูกตัดออกทำให้ค่าของแอทริบิวท์ HAIR เป็น {black} เท่านั้น ซึ่ง

$a_1 = 1, w_1 = 0, af_1 = 0$ $I(a_1, w_1, af_1) = 1$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

E(HAIR) = 1 bits

เช่นเดียวกันถ้ายังใช้วิธีการของไอดีทรีต่อ อาจทำให้เกิดปัญหาตามมาได้ เนื่องจากไอดีทรีจะนำแอทริบิวท์ HAIR ไปใช้พิจารณาในการเลือกรูทโหนดของต้นไม้ย่อยด้วย ซึ่งไม่สมควร เนื่องจากถ้าแอทริบิวท์ HAIR ถูกเลือกเป็นรูทโหนดแล้วมีค่าเหลือเพียงค่าเดียวก็จะแบ่งชุดฝึกหัดได้ชุดฝึกหัดย่อยเหมือนชุดฝึกหัดเดิม แล้วแอทริบิวท์ที่เหลือก็จะถูกเลือกเป็นรูทโหนดของต้นไม้ย่อยต่อไป จะเห็นว่าแอทริบิวท์ HAIR มีบทบาทน้อยมากในการใช้แบ่งประเภท เพราะถึงแม้ว่าจะเลือกมันเป็นโหนดโหนดหนึ่ง แต่มันก็ไม่ได้ช่วยในการใช้แบ่งประเภทให้กับต้นไม้สำหรับการตัดสินใจแต่อย่างใด ดังนั้นจึงไม่สมควรที่จะพิจารณาแอทริบิวท์ใดๆ เป็นรูทโหนดอีกต่อไปถ้าพบว่ามันเหลือค่าเพียงค่าเดียวในขณะที่ตัวอย่างในชุดฝึกหัดย่อย(หรือชุดฝึกหัด) มีจำนวนมากกว่าหนึ่งตัวอย่างขึ้นไป และมีประเภทแตกต่างกัน เราจะตัดแอทริบิวท์ที่เกิดกรณีเช่นนี้ทิ้งไป เพื่อไม่ให้ ไอดีทรี นำมาใช้พิจารณา

และเช่นเดียวกัน ที่แอทริบิวท์ EYE_COLOR จะเหลือค่าเฉพาะ black เท่านั้น จึงไม่มีความจำเป็นต้องพิจารณาแอทริบิวท์นี้

ที่แอทริบิวท์ NOSE ซึ่งมีค่าเป็น {sharp,flat}

$a_1 = 1, w_1 = 0, af = 0$ $I(a_1,w_1,af_1) = 0$

$a_2 = 0, w_2 = 0, af = 1$ $I(a_2,w_2,af_2) = 0$

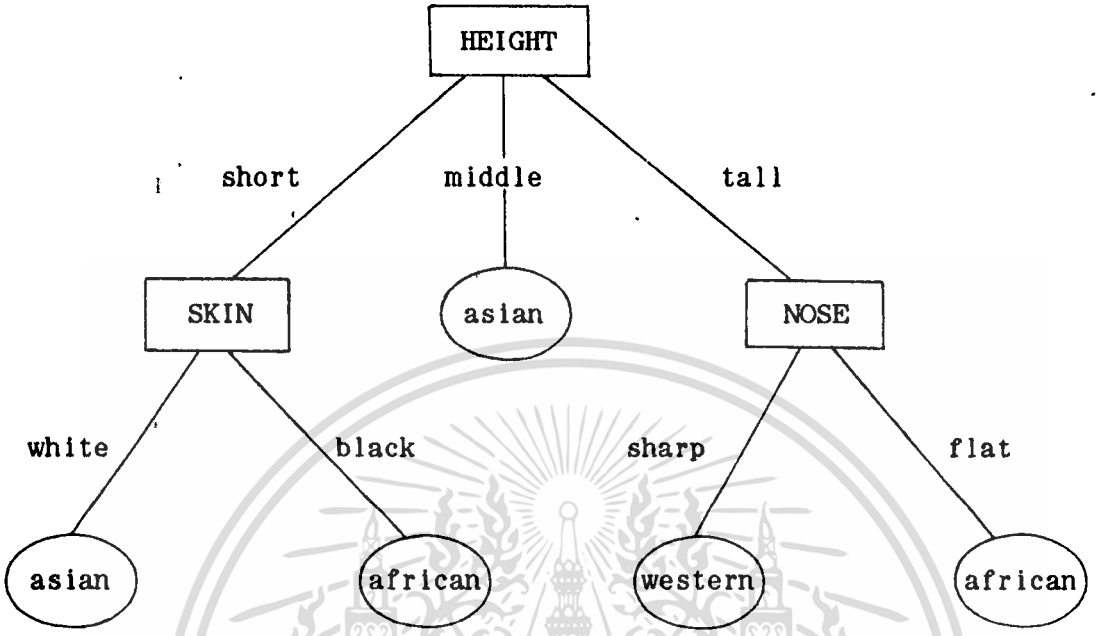
E(NOSE) = 0 bits

ตอนนี้เรามี E(A) ค่าที่น้อยที่สุดมีค่าเท่ากัน คือ E(A) ของแอทริบิวท์ EYE_LAYER SKIN และ NOSE โดยวิธีการของไอดีทรีไม่ได้มีการกำหนดไว้ว่าควรจะเลือกแอทริบิวท์ไหนเป็นรูทโหนด ดังนั้นในโปรแกรมก็ขึ้นอยู่กับโปรแกรมเรียงลำดับ (Sort) ค่าของ E(A) ของแต่ละแอทริบิวท์แล้วก็จะเลือกเอาตัวที่อยู่หน้าสุด (ค่าที่น้อยที่สุด) ซึ่งอาจจะเท่ากับค่าถัดมาก็ได้ มาเป็นรูทโหนด ข้อเสนอแนะก็คือ อาจจะมีการถ่วงน้ำหนักให้กับค่า E(A) ของแอทริบิวท์ที่เท่ากันตามลำดับความสำคัญ (Priority) โดยให้ลำดับความสำคัญของแอทริบิวท์ที่เหมาะสมมากกว่ามีค่ามากกว่า หรืออีกวิธีหนึ่ง (มีในการทำงานของ เล็กซ์เซลล์) คือ อนุญาตให้ผู้ใช้เลือกเองว่าควรจะใช้แอทริบิวท์ใดเป็นโหนดในขณะนั้นจึงจะเหมาะสม

ไม่ว่าจะใช้วิธีเลือกแบบที่ 1 หรือ แบบที่ 2 เมื่อได้รูทโหนดแล้วต่อไปก็ดำเนินการสร้างต้นไม้ตามวิธีของไอดีทรีและวิธีใหม่ที่ผู้จัดทำคิดค้นขึ้นเอง ตอนนี้สมมุติว่าเราเลือกแอทริบิวท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยวิธีการเดียวกับที่ได้กล่าวมาแล้วสามารถสร้างต้นไม้ที่สมบูรณ์ได้ดังรูป



3.3.3 สรุป

โดยอาศัยวิธีการสร้างต้นไม้แบบ ไอดีทรี จะไม่สามารถสร้างต้นไม้จากชุดฝึกหัดบางชุดได้ ถึงแม้ว่าชุดฝึกหัดนั้นจะไม่ได้ขัดแย้งกับกฎเกณฑ์ข้อกำหนดชุดฝึกหัดที่ ไอดีทรี จะยอมรับได้ก็ตาม ดังปัญหาที่ได้กล่าวไปแล้ว แต่โดยอาศัยวิธีการที่ผู้จัดทำโครงงานได้ค้นคิดขึ้น ก็มีผลทำให้ไอดีทรีที่ปรับปรุงแล้วสามารถแก้ไขปัญหาเหล่านี้ได้ และสามารถสร้างต้นไม้สำหรับการตัดสินใจได้อย่างถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ระบบสารสนเทศยุคที่ห้า (The 5th Generation Information Systems)

ระบบสารสนเทศยุคที่ห้าที่จะกล่าวถึงต่อไปนี้เป็นแนวความคิดของโปรเฟสเซอร์ จี เอ็ม ไนเซน และ ด๊อกเตอร์ อี ฟาลเคนเบอร์ก

หลายคนกล่าวว่าระบบสารสนเทศยุคที่ห้า หรือ ฟิฟจีไอเอส (5GIS) เป็นระบบที่ฉลาด ก่อนอื่นมาพิจารณาดูกันว่าระบบที่จะเรียกว่า "ฉลาด" นั้นเป็นอย่างไร

อลัน เทอริง (Alan Turing) ได้เสนอวิธีการทดสอบว่าระบบฉลาดจริงหรือไม่ ซึ่งเรียกง่ายๆว่า "การทดสอบเทอริง" การทดสอบนี้ทำโดยการสมมุติว่ามีห้องสองห้อง ห้องหนึ่งมีคนและอีกห้องหนึ่งมีเครื่องคอมพิวเตอร์ โดยที่เราอยู่ข้างนอกห้องทั้งสอง ทั้งคนและเครื่องคอมพิวเตอร์ต่างก็พยายามบอกว่าตัวเองเป็นคน แล้วให้เราเป็นคนตัดสิน ถ้าเราไม่สามารถบอกได้ว่า คนอยู่ในห้องไหนแน่ ก็แสดงเครื่องคอมพิวเตอร์นั้นฉลาด การทดสอบนี้เราสามารถติดต่อกับคนหรือเครื่องโดยใช้ภาษาธรรมชาติเกี่ยวกับเรื่องที่พูดถึงได้ไม่จำกัด ในปัจจุบันยังไม่มีเครื่องคอมพิวเตอร์ไหนที่สามารถผ่านการทดสอบเทอริงนี้ได้ เนื่องจากสาเหตุสำคัญสองประการคือ

1. เครื่องจะต้องไม่มีความจำกัดทางภาษาธรรมชาติ ซึ่งเป็นปัญหาที่ยังแก้ไม่ตกอยู่ในปัจจุบัน

2. เครื่องจะต้องสามารถสนทนาเรื่องใดๆ ก็ได้ไม่จำกัด

แต่อย่างไรก็ตาม ก็มีตัวอย่างหลายอันที่แสดงให้เห็นว่าเครื่องคอมพิวเตอร์มีความสามารถในระดับขั้นผู้เชี่ยวชาญ ได้แก่ โปรแกรมหมากรุก ระบบผู้เชี่ยวชาญ แต่โปรแกรมเหล่านี้ ถือว่าผ่านการทดสอบเทอริงในขีดจำกัดเท่านั้น คือ จะต้องมีการกำหนดเรื่องที่สนทนาเป็นเฉพาะเรื่องใดเรื่องหนึ่งเท่านั้น

ระบบผู้เชี่ยวชาญ เป็นโปรแกรมหนึ่งที่มีความซับซ้อนเป็นอย่างมาก บางครั้งได้มีการใช้กับฐานข้อมูลเฉพาะเรื่องที่มีขนาดใหญ่มาก ซึ่งถ้าจะเปรียบเทียบกับ ฟิฟจีไอเอส แล้ว จะพบว่าฟิฟจีไอเอสก็คือ ระบบผู้เชี่ยวชาญที่ผู้ใช้สามารถกำหนดโดเมนได้ ซึ่งก็คือ เปลือกระบบผู้เชี่ยวชาญนั่นเอง ระบบจะอนุญาตให้ผู้ใช้ใส่เรื่องราวที่กล่าวถึงในโดเมนเข้าไปได้ แล้วระบบจะสนทนากับผู้ใช้เป็นภาษาธรรมชาติทั้งหมด ความสามารถในการสนทนาของระบบจะขึ้นอยู่กับ การติดต่อกับผู้ใช้ การจัดการฐานข้อมูล และความสามารถในการอนุมาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถึงแม้ว่า ฟิฟจีไอเอสจะต้องการความสามารถในระดับผู้เชี่ยวชาญ แต่ก็ไม่จำเป็นว่าฟิฟจีไอเอสจะมีความสามารถเช่นนั้นทุกครั้งที่ติดต่อกับผู้ใช้ ถึงอย่างไรก็ตามมันจะต้องสามารถติดต่อสนทนาได้เช่นเดียวกับคน ซึ่งกำหนดเป็นกฎข้อที่ 1 ของฟิฟจีไอเอสได้ว่า

" ถ้าคนสองคนพูดคุยกันในเรื่องใดเรื่องหนึ่งอยู่ แล้วแทนคนคนหนึ่งด้วยฟิฟจีไอเอสแล้ว คนอีกคนไม่ควรที่จะเปลี่ยนแปลงภาษาเดิมที่ตนใช้กับคนคนนั้น "

กฎข้อนี้ระบุถึงระดับนอกของฟิฟจีไอเอสว่า ภาษาธรรมชาติที่ใช้ติดต่อกับระบบควรจะมีลักษณะเช่นไร สำหรับภาษาธรรมชาติที่ใช้กันอยู่ ได้แก่ ภาษาอังกฤษ ภาษาไทย ภาษาญี่ปุ่น นั้นมีความซับซ้อนเป็นอย่างมาก จะต้องใช้เวลานานมากกว่าที่จะสามารถแก้ไขข้อจำกัดต่างๆเพื่อใช้กับระบบสารสนเทศ ดังนั้นตอนนี้ก็จะพูดถึงเพียงเฉพาะสับเซตที่มีรูปแบบของภาษาธรรมชาติ ซึ่งเรียกว่า ภาษาธรรมชาติที่มีรูปแบบ(FNL: Formal Natural Language) หรือภาษาโปรแกรมมิงที่เป็นธรรมชาติ(NPL: Natural Programming Language)

แน่นอนเหลือเกินว่าแต่ละภาษาก็จะมีเอฟเอ็นเอล(FNL) ของภาษานั้นๆ เช่น ภาษาอังกฤษก็จะมีเอฟเอ็นเอลหนึ่ง ภาษาไทย และ ญี่ปุ่น ก็เช่นกัน ฟิฟจีไอเอส ควรที่จะสามารถตอบสนองกับเอฟเอ็นเอลอันเดียวกันได้ ตัวอย่างเช่น เราถามคำถามว่า

What is the age of selena?

เราจะได้รับคำตอบว่า

juninen

คำตอบนี้แทบจะไม่มีประโยชน์เลย ถ้าไม่ทราบว่าเป็นภาษาญี่ปุ่นที่แปลว่า " 12 ปี " หรือแม้ว่าเราจะสามารถแปลจากภาษาญี่ปุ่นเป็นภาษาอังกฤษได้ เราก็อาจแปลความหมายคำตอบผิดผลาดอีก เนื่องจากชาวญี่ปุ่นมีหลักเกณฑ์ในการนับอายุผิดแผกแตกต่างไปจากชนชาติทั่วไป คือ ชาวญี่ปุ่นจะนับอายุของเด็กทารกที่อยู่ในครรภ์มารดาว่ามีอายุ 1 ขวบ ขณะที่ชนชาติทั่วไปจะนับเป็นอายุ 0 ขวบ ดังนั้นคำตอบว่า อายุ 12 ปี ในระบบที่เป็นภาษาญี่ปุ่นก็คือ อายุ 11 ปี ในระบบของพวกเรา ดังนั้นนอกจากการติดต่อสนทนาจะต้องใช้ภาษาเดียวกันแล้ว คำที่ใช้จะต้องมีความหมายอันเดียวกันด้วย ซึ่งจะเกิดได้ก็ต่อเมื่อ ประการที่หนึ่ง คู่สนทนาจะต้องพูดในเนื้อหาของเรื่องราว(UoD: Universe of Discourse) หรือ ยูโอดี เดียวกัน ประการที่สอง ประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่พูดเกี่ยวกับเนื่อหาของเรื่องราวนั้นจะต้องไม่กำกวม

จากตัวอย่างที่แล้วสาเหตุของความสับสนเกี่ยวกับอายุของเซเลน่า คือ ไม่ทราบแน่นอนว่าอายุของเธอเป็น 11 หรือ 12 ปี เนื่องมาจากการใช้ยูโอดีที่ต่างกัน ฝ่ายหนึ่งพูดถึงการคิดอายุแบบธรรมดา แต่อีกฝ่ายคิดอายุแบบชาวญี่ปุ่น พิจารณาดูประโยคตัวอย่างต่อไปนี้

Pluto is owned by Mickey Mouse.

ประโยคนี้จะถูกต้อง เมื่อพูดถึงยูโอดีที่เกี่ยวกับตัวละครการ์ตูนของวอล์ท ดิสนีย์ แต่ถ้ามีบางคนที่ไม่คุ้นเคยกับการ์ตูนของวอล์ท ดิสนีย์ อาจจะแปลประโยคนี้ไปทางด้านดาราศาสตร์ที่กล่าวถึงดาวพลูโต ส่งผลให้การสื่อสารล้มเหลวได้

จากที่กล่าวมาแล้ว จึงมีความจำเป็นอย่างยิ่งที่จะต้องหาวิธีการในการอธิบายยูโอดีในระบบสารสนเทศ โครงร่างของยูโอดีที่ระดับแนวคิดของระบบสารสนเทศนั้น เราเรียกว่าคอนเซบชวลสกีมม่า(Conceptual Schema) ซึ่งใช้อธิบายโครงสร้างหรือแบบฟอร์มของยูโอดีนั้น ๆ เกี่ยวกับชนิดของสิ่งของ (Objects) ที่มีอยู่ในยูโอดี บทบาท (Roles) ที่สิ่งของต่างๆ แสดงกัน และ เงื่อนไขข้อบังคับ (Constaints) ที่มีในยูโอดี เป็นต้น หรือกล่าวอีกอย่างหนึ่งได้ว่า คอนเซบชวลสกีมม่าก็คือ แบบแผนทั่วๆไปที่ใช้ในการออกแบบยูโอดี นั่นเอง

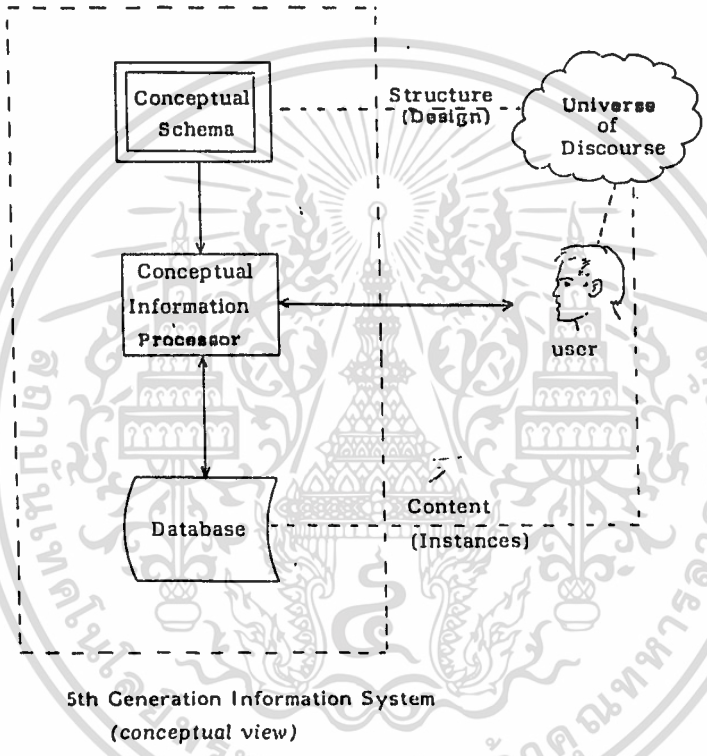
เมื่อคอนเซบชวลสกีมม่าเป็นตัวกำหนดโครงสร้างของยูโอดี ส่วนที่ทำหน้าที่กำหนดรายละเอียด (Content) หรือ ประชากร (Population) ของยูโอดี ก็คือ ฐานข้อมูล (Database) ตามแนวความคิดแล้วอาจจะมองฐานข้อมูลเป็นกลุ่มของประพจน์ (Proposition) ในทางตรรกศาสตร์ซึ่งบ่งบอกถึงความจริงของยูโอดี ประพจน์แต่ละประพจน์จะมีการรวมเอาตัวอย่างของบทบาทเฉพาะที่แสดงโดยสิ่งของหนึ่งสิ่งหรือหลายสิ่งในยูโอดี ประพจน์ต่างๆ อาจจะถูกใส่เข้าหรือลบออกจากฐานข้อมูลก็ได้ ทำให้ฐานข้อมูลเปลี่ยนแปลงไป อย่างไรก็ตามที่เวลาใดเวลาหนึ่งประพจน์ทั้งหมดที่มีอยู่ในฐานข้อมูลจะต้องสอดคล้องกับแบบแผนที่ได้ออกแบบไว้แล้วโดยคอนเซบชวลสกีมม่า นี่เป็นกฎที่สองของฟิฟจีไอเอส ที่กล่าวว่า

" คอนเซบชวลสกีมม่าจะเป็นตัวที่รับผิดชอบทั้งหมดเกี่ยวกับการอนุญาติให้ฐานข้อมูลอยู่ในสภาวะหรือการเปลี่ยนแปลงที่ยอมรับได้เท่านั้น "

เพื่อที่จะให้เป็นตามกฎข้อนี้ จึงทำให้เกิดส่วนประกอบส่วนที่ 3 ขึ้นมาเรียกว่า ตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประมวลผลสารสนเทศที่เป็นแนวคิด(CIP:Conceptual Information Processor) ซึ่งจะเป็นส่วนที่รับผิดชอบเกี่ยวกับการควบคุมดูแลการแก้ไขฐานข้อมูลโดยผู้ใช้และการตอบคำถามของผู้ใช้ ดังนั้น สถาปัตยกรรมทางแนวคิด (Conceptual Architecture) เบื้องต้น ของฟิฟจีไอเอส จึงมีลักษณะดังรูป 4.1 จากรูปนี้สมมุติว่าคอนเซบชวลสกีมม่าได้ถูกใส่ในระบบ และสามารถติดต่อกับระบบได้แล้ว การนำระบบนี้ไปใช้ประยุกต์กับงานอะไรก็ใส่คอนเซบชวลสกีมม่าของงานนั้นๆ เข้าไป ในกรณีที่ยูไอดีมีความซับซ้อนมากๆ อาจจะต้องให้ผู้เชี่ยวชาญเป็นคนออกแบบคอนเซบชวลสกีมม่า



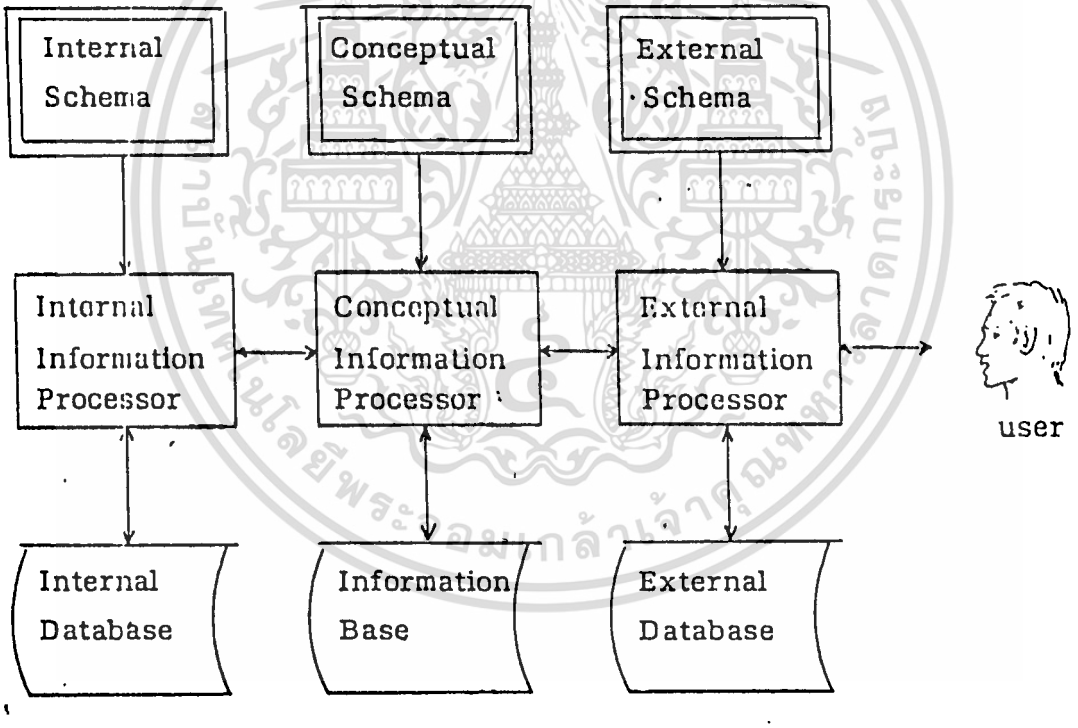
รูปที่ 4.1 สถาปัตยกรรมทางแนวคิดเบื้องต้นของฟิฟจีไอเอส

สำหรับโครงงานในวิทยานิพนธ์เล่มนี้ได้สร้างเปลือกระบบผู้เชี่ยวชาญซึ่งเป็นสถาปัตยกรรมทางแนวคิดเบื้องต้นของฟิฟจีไอเอส จากภาษา PROLOG โดยใช้แบบจำลองข้อมูลไนแอมในการออกแบบคอนเซบชวลสกีมม่า ซึ่งจะกล่าวถึงการรายละเอียดในบทที่ 7

ในการอธิบายสถาปัตยกรรมที่พัฒนาต่อไปจะใช้คำว่าฐานสารสนเทศ (Information base) แทนคำว่าฐานข้อมูล(Database) ดังรูป 4.2 จากรูป ผู้ใช้จะมองเห็นเฉพาะระดับภายนอก(External Level) แทนที่จะเป็นระดับแนวคิด(Conceptual Level) ผู้ใช้สามารถติดต่อกับระดับภายนอกเท่านั้น การออกแบบยูไอดีที่แสดงโดยคอนเซบชวลสกีมม่านั้นเป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนพื้นฐานและเป็นลอจิกคอล (Logical) ที่บรรยายถึงโครงสร้างที่จำเป็นของยูโอดีอย่างสมบูรณ์และถูกต้อง ไม่จำเป็นต้องยุ่งเกี่ยวกับการหาวิธีติดต่อสื่อสารกับผู้ใช้อย่างเป็นมิตรหรืออื่นใดทั้งสิ้น และก็ไม่จำเป็นต้องรู้ถึงรายละเอียดทางด้านกายภาพเกี่ยวกับการจัดเก็บฐานสารสนเทศ ว่าทำอย่างไรถึงจะสามารถทำให้มีประสิทธิภาพมากที่สุดในระดับของฮาร์ดแวร์ของเครื่อง ดังนั้นจึงได้รวมเอาส่วนประกอบที่ไม่ยุ่งเกี่ยวกับระดับแนวคิดเข้าไปในสถาปัตยกรรมด้วย คือ ส่วนประกอบภายนอก และส่วนประกอบภายใน ดังแสดงในรูป 4.2 ส่วนเอ็กเทอนอลสกีมม่า (External Schema) เป็นส่วนที่กำหนดการออกแบบยูโอดีที่รับจากผู้ใช้แต่ละคนโดยตรง สำหรับส่วนอินเทอนอลสกีมม่า (Internal Schema) เป็นส่วนที่บอกว่าการออกแบบทางแนวคิดถูกสร้างในระดับกายภาพได้อย่างไร



3-level architecture of 5th generation information system

รูปที่ 4.2 แสดงสถาปัตยกรรมของฟิฟจีไอเอสที่ขยายต่อไป

การต่อเติมสถาปัตยกรรมเป็นไปตามกฎข้อที่ 3 ของฟิฟจีไอเอส (หลักสำคัญภายใน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนี้

" อินเทอนอลสกีมม่าเป็นตัวที่รับผิดชอบทั้งหมดเกี่ยวกับการระบุการแทนสภาวะของฐานสารสนเทศที่ยอมรับได้ในระดับกายภาพ การแทนฐานสารสนเทศในระดับกายภาพนี้เรียกว่าฐานข้อมูลภายใน (Internal Database) "

นอกจากนี้ยังเป็นไปตามกฎข้อที่ 4 ของ พีพจีไอเอส (หลักสำคัญภายนอก) ดังนี้

"เอ็กเทอนอลสกีมม่าเป็นตัวกำหนดภาพพจน์ (View) ของฐานสารสนเทศสำหรับผู้ใช้เฉพาะคน ฐานข้อมูลที่เกี่ยวข้องกับภาพพจน์ดังกล่าวนี้เรียกว่า ฐานข้อมูลภายนอก (External Database)"

เพื่อที่จะให้สถาปัตยกรรมนี้สมบูรณ์ ส่วนคอนเซบชวลสกีมม่าอาจจะแสดงกลุ่มของประโยคที่อธิบายฐานสารสนเทศของตัวมันเอง (สำหรับในวิทยานิพนธ์เล่มนี้จะเรียกคอนเซบชวลสกีมม่านี้ว่า เมต้าคอนเซบชวลสกีมม่า) จากข้อกำหนดนี้ได้นำไปสู่กฎข้อที่ 5 ของพีพจีไอเอส (หลักสำคัญเมตา (The Meta Principle)) ดังนี้

" คอนเซบชวลสกีมม่า อาจจะถูกมองเป็นฐานสารสนเทศได้ ซึ่งสถานะของฐานสารสนเทศที่ยอมรับได้ถูกกำหนดโดยเมต้าคอนเซบชวลสกีมม่า "

โดยอาศัยตัวช่วยออกแบบคอนเซบชวลสกีมม่า เมต้าคอนเซบชวลสกีมม่าอาจจะถูกใช้ในการสร้างคอนเซบชวลสกีมม่าแบบต่างๆ ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

ตัวประมวลผลสารสนเทศที่เป็นแนวคิด

(CIP : Conceptual Information Processor)

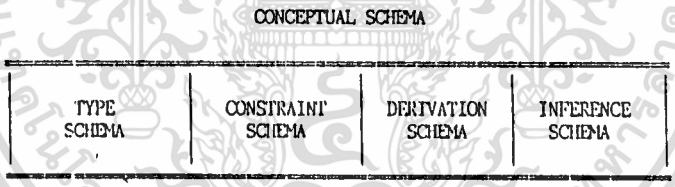
จากที่ได้กล่าวถึงระบบสารสนเทศยุคที่ห้าไปแล้วนั้นจะเห็นว่าส่วนที่มีหน้าที่ติดต่อสื่อสารกับผู้ใช้ในระดับแนวคิดทั้งหมดเป็นหน้าที่ของตัวประมวลผลสารสนเทศที่เป็นแนวคิด หรือเรียกย่อๆ ว่า ซีไอพี (CIP) การติดต่อดังกล่าวแบ่งได้เป็น 2 ขั้นตอนดังนี้

1. ผู้ใช้แก้ไขฐานข้อมูลโดยการเพิ่ม (Insert)หรือ การลบ (Delete) ประพจน์ที่ต้องการซึ่งซีไอพีจะยอมรับการแก้ไขนั้นก็ต่อเมื่อการแก้ไขนั้นยังคงรักษาความถูกต้องตามคอนเซบชวลสกีมม่า

2. ผู้ใช้ป้อนคำถามเกี่ยวกับยูไอดีให้กับระบบ ซึ่งซีไอพีจะเป็นผู้ที่ตอบคำถามนั้น ซีไอพีสามารถให้ข่าวสารเกี่ยวกับคอนเซบชวลสกีมม่า หรือ เกี่ยวกับฐานข้อมูลซึ่งอาจจะได้ข่าวสารมาจากการเก็บโดยตรงหรือจากการอนุมาน (Deduction)

จากการทำงานทั้ง 2 ขั้นตอนดังกล่าว ซีไอพี จึงเปรียบเสมือนเป็น ตัวกรองข้อมูล (Data Filter) และตัวให้ข่าวสาร (Information Supplier) ตามลำดับ

ในที่นี้จะไม่ขอกล่าวถึงว่าจะใส่คอนเซบชวลสกีมม่าเข้าไปได้อย่างไร แต่จะขอกล่าวถึงส่วนประกอบของคอนเซบชวลสกีมม่า ซึ่งประกอบด้วยสกีมม่าย่อย 4 สกีมม่า ดังรูป 5.1



รูปที่ 5.1 แสดงส่วนประกอบของคอนเซบชวลสกีมม่า

- สกีมม่าชนิด (Type Schema) บ่งบอกถึงชนิดของประพจน์ที่อาจจะถูกเก็บในฐานข้อมูล โดยทั่วไปแล้วจะบ่งบอกถึงชนิดของเอนติตี้ที่ยอมรับได้ เช่น นักศึกษา วิชาเรียน ภาควิชา ชนิดของเลเบ้ลที่ใช้อ้างอิง เอนติตี้ เช่น ชื่อนักศึกษา ระหัสวิชา ชื่อภาควิชา และบรรดาบทบาท (Roles) ทั้งหลายที่แสดงโดยเอนติตี้หรือเลเบ้ล เช่น เรียน สังกัด รวมทั้งความสัมพันธ์ที่เป็นชนิดย่อย (Subtyne) ที่อาจจะมีการระบุ เช่น นักศึกษาชายเป็นชนิดย่อยของชนิดเอนติตี้นักศึกษา

- คอนสเตนท์สกีมม่า (Constraint Schema) เป็นตัวแสดงคอนสเตนท์หรือข้อจำกัดต่างๆ ที่ใช้กับชนิดที่มีอยู่ในยูไอดี ซึ่งอาจจะเป็นแบบสแตติค (Static) หรือแบบไดนามิค (Dynamic) คอนสเตนท์แบบสแตติคสามารถใช้ได้กับทุกๆ สภาวะของฐานข้อมูล ตัวอย่างเช่น เราต้องการว่า ห้ามบุคคลคนหนึ่งแต่งงานกับบุคคลอื่นมากกว่าหนึ่งคนขึ้นไปในเวลาขณะใดขณะหนึ่ง นี้เป็นตัวอย่างของยูนิกเนสคอนสเตน์ (Uniquenss Constraint) นอกจากนี้ยังมีคอนสเตนท์แบบสแตติคอีกหลายชนิด คือ คอนสเตนท์ความจำเป็น (Necessity Constraint) คอนสเตนท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความถี่ (Frequency Constraint) คอนสเตนท์เท่ากัน (Equality Constraint) คอนสเตนท์สับเซต (Subset Constraint) และคอนสเตนท์ยกเว้น (Exclusion Constraint)

สำหรับคอนสเตนท์แบบไดนามิคนั้น จะห้ามการเปลี่ยนแปลงสภาวะบางอย่างของฐานข้อมูล ตัวอย่างเช่น แม้ว่านักศึกษาอาจจะลงทะเบียนในปี 9 ในเวลาหนึ่ง และลงทะเบียนในปี 10 ในอีกเวลาหนึ่ง แต่โรงเรียนจะต้องห้ามนักเรียนกระโดดจากการลงทะเบียนปี 9 ไปเป็นการลงทะเบียนปี 11

คอนสเตนท์นั้นเรียกได้ว่าเป็นกฎที่ทำให้เกิดความถูกต้อง (Validation Rules) เนื่องจากมันจะยินยอมสำหรับปอปปูเลชั่น (Population) ของฐานข้อมูลที่เป็นปอปปูเลชั่นที่ถูกต้องเท่านั้น ปอปปูเลชั่นที่ถูกต้องหมายถึงว่า เป็นปอปปูเลชั่นที่สอดคล้องกับคอนเซบชวลสกีมม่า จะเห็นได้ว่าคอนเดนท์ชนิดต่างๆ ส่วนใหญ่สามารถแสดงในคอนเซบชวลสกีมม่าได้เลย แต่ก็มีบางคอนสเตท์ โดยเฉพาะอย่างยิ่งคอนสเตนท์การเปลี่ยนแปลง (Transition Constraints) ซึ่งจำเป็นต้องแสดงโดยวิธีการอื่น คือ โดยประโยค ตาราง หรือ กราฟ

- เดริเวชั่นสกีมม่า (Derivation Schema) เป็นตัวให้กฎการคำนวณทางคณิตศาสตร์เพื่อที่จะสามารถให้ข่าวสารบางอย่างที่ไม่มีอยู่ในฐานข้อมูลออกมาได้ โดยการคำนวณตามกฎเกณฑ์ดังกล่าว ตัวอย่างเช่น ค่าผลรวมและค่าเฉลี่ยอาจจะคำนวณได้โดยการกระทำทางคณิตศาสตร์กับข้อมูลที่เหมาะสม แม้ว่าเราสามารถเก็บค่าที่คำนวณได้นี้ไว้ในฐานข้อมูล แต่ก็เป็นการไม่สมควร ไม่ใช่เหตุผลเพราะว่า เพื่อประหยัดเนื้อที่ของหน่วยความจำ แต่เพื่อทำให้การแก้ไขฐานข้อมูลเป็นไปได้ง่าย

ตัวอย่างเช่น สมมุติว่าเราต้องการได้ อายุของนักเรียนในห้องแต่ละคนและอายุเฉลี่ยของนักเรียนในห้อง ถ้าเราเก็บอายุเฉลี่ยไว้ในฐานข้อมูล เราก็จะต้องทำการจัดเตรียมข้อมูลนี้โดยการคำนวณค่านี้ใหม่ทุกครั้งที่มีการเพิ่มนักเรียนเข้าหรือตัดออกจากห้อง หรือทุกๆ ครั้งที่อายุของนักเรียนแต่ละคนเพิ่มขึ้น เพื่อแก้ไขปัญหานี้ เราเลือกเก็บเพียงจำนวนของนักเรียนและอายุของนักเรียนแต่ละคน ส่วนอายุเฉลี่ยนั้นเราสามารถหามาได้โดยใช้กฎการคำนวณทางคณิตศาสตร์เพื่อหาค่าเฉลี่ย โดยการหาผลรวมของอายุแล้วหารด้วยจำนวนของนักเรียน แต่การจัดการโดยวิธีนี้ก็ยังไม่ดี เราไม่จำเป็นต้องเก็บจำนวนของนักเรียนในห้องเนื่องจากสามารถหาได้จากการนับจำนวนของประพจน์ที่อยู่ในรูปแบบ "student x is in class y " เพื่อเป็นการหลีกเลี่ยงที่ต้องแก้ไขทุกครั้งที่มีนักเรียนเข้ามาเพิ่มในห้องหรือออกจากห้อง และจริง ๆ แล้วก็ไม่มีความจำเป็นจะต้องเก็บอายุของนักเรียนแต่ละคน แต่ควรที่จะเก็บวันเกิดของแต่ละคนรวมทั้งวันที่ปัจจุบันไว้ ซึ่งระบบจะสามารถหาอายุของนักเรียนแต่ละคนได้จากการลบวันที่ปัจจุบันกับวันเกิดของเขา สำหรับการแก้ไขวันที่ปัจจุบันทุกๆวันนั้น สามารถทำได้อันโนมัติโดยใช้นาฬิกาของระบบ โดยการจัดการเช่นนี้ทำให้ไม่ต้องกังวลกับการแก้ไขอายุของนักเรียน เมื่อ 1 ปีผ่านไป

- อินเฟอเร้นซ์สกีมม่า (Inference Schema) เป็นตัวที่เก็บกฎทางตรรก (Logical Rules) ซึ่งสามารถใช้ในการอนุมานข่าวสารที่ไม่มีอยู่ในฐานข้อมูลโดยตรง กฎการวิ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นิจฉัย (Inference Rules) จะเป็นตัวเพิ่มพูนความสามารถในการให้ข่าวสารคำตอบกับผู้ใช้ได้มากขึ้น เมื่อซีไอพีทำงานในลักษณะเช่นนี้เราเรียกว่า ตัวขยายข่าวสาร (Information Amplifier) นอกจากนี้กฎการวินิจฉัยยังสามารถช่วยซีไอพีป้องกันไม่ให้ผู้ใช้ทำการแก้ไขอะไรที่ไม่สอดคล้องกับคอนเซพชวลสกีมม่า ดังนั้นกฎการวินิจฉัยก็ถือว่ามีส่วนร่วมในหน้าที่ของซีไอพีในฐานะตัวกรองข้อมูลเช่นกัน ขอให้พิจารณาตัวอย่างง่าย ๆ ข่าวสารส่วนใหญ่จะอยู่ในรูปของประพจน์ในฐานข้อมูล ก่อนอื่นจะขอกล่าวถึงข้อแตกต่างระหว่างประพจน์และประโยคเสียก่อน ประพจน์นั้นมีค่าเป็นไปได้เพียง 2 ค่าเท่านั้น คือ จริงกับเท็จ ส่วนประโยคนั้นใช้แทนประพจน์โดยประกอบด้วยคำที่เรียงกันอย่างถูกต้องตามหลักไวยกรณ์ ดังนั้นในระดับแนวคิดประพจน์ใด ๆ ในฐานข้อมูลจะถูกแทนด้วยประโยค ทำให้ซีไอพีจัดการกับประโยคที่เป็นการบอกกล่าว (Declarative Sentences) โดยตรงแทนที่จะจัดการกับความหมายที่อยู่ในประโยค สมมุติว่าเราต้องการเก็บประโยคประเภท

person X is married to person Y

ไว้ในฐานข้อมูล จากความรู้ในเรื่องการแต่งงานเราสามารถรวมคอนสเตนท์สกีมม่าต่อไปนี้เข้าไปด้วย

C1 Each person can be married to at most one person

C2 Human marriage is irreflexive i.e.
no person may mary himself/herseft

สมมุติว่าเราปอปปูเลขยูไอดีของเราตามนี้

add: person Alan is married to person Betty ------> acepted

ซีไอพีพิจารณาประโยคนี้พบว่าเป็นประเภทของประโยคที่ถูกต้องและไม่กระทบกระเทือนกับคอนสเตนท์ไหน ดังนั้นมันก็จะเพิ่มประโยคนี้เข้าไปในฐานข้อมูลและส่งข้อความ"accepted" ออกมาว่ารับแล้ว กรณีที่การแก้ไขขัดแย้งกับคอนเซบชวลสกีมม่า ซีไอพีจะไม่ยอมรับ ดังตัวอย่างต่อไปนี้

add: cat Doris is married to cat Felix ---> rejected.wrong type

add: person Alan is married to person Sue ---> rejected. C1 violated

add: person Alan is married to person Alan ---> rejected. C2 violated

สมมุติว่าตอนนี้จะเพิ่มประโยค

add: person Betty is married to person Colin

เข้าไป ถึงแม้ว่าระบบจะเก็บประโยค "Person is married to person Betty" เข้าไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้ว แต่ยังไม่มีประโยค "Person Betty is married to person Alan" ก็ตาม แต่ประโยคนี้ขัดแย้งกับ C1 ที่บอกว่า บุคคล X แต่งงานกับบุคค Y ได้อย่างมาก 1 คนเท่านั้น แต่อย่างไรก็ ชีไอพีก็ยังรับประโยคนี้ วิธีที่หลีกเลี่ยงข้อผิดพลาดนี้ ก็โดยการเพิ่มกฎเข้าไปในอินเฟอเรนสกีม ม่าดังนี้

I1 Human marriage is symmetric

i.e. given any people X and Y

if X is married to Y then Y is married to X

เมื่อไรที่มีกฎการวินิจฉัยอยู่ในระบบ จะถือว่าชีไอพีมีเครื่องจักรกลวินิจฉัย (Inference engine) ซึ่งเป็นตัววินิจฉัยกฎเหล่านี้ ในกรณีเช่นนี้เครื่องจักรกลวินิจฉัยจะถูกสั่งให้เริ่มทำงาน โดยมันจะมีกฎการแทนค่าตัวแปรทางตรรก (Logical Rules of Universal Instantiation (UI)) และกฎการยืนยันความจริงของเงื่อนไข (Rules of Affirming the Antecedent (AA)) ในการแก้ไขแต่ละครั้ง

ยูไอ(UI) จะทำให้เราสามารถยืนยันตัวอย่างของประพจน์ทั่วไป(Generel Proposition) ได้ซึ่งก็คือ การแทนค่า X ด้วย Alan และ Y ด้วย Betty ในกฎ I1 นั่นเอง ซึ่งเราจะได้

if Alan is married to Betty then Betty is married to Alan

ส่วนเอเอ (AA) จะทำให้สรุปส่วนที่เป็นผลสรุปของเงื่อนไขไปได้ ถ้าเงื่อนไขถูกยืนยันว่าเป็นจริงกำหนดประพจน์ P และQ ที่อยู่ในรูป

if P then Q

ถ้า P จริง

เราสามารถสรุปได้ว่า Q จริง

ดังนั้นไม่จำเป็นจะต้องเก็บว่า "Betty is married to Alan" เอเอก็สามารถสรุปให้เราได้ จากประโยค "Alan is married to Betty" ที่มีอยู่ในฐานข้อมูล โดยใช้คอนสเตนท์ C1 ร่วมกับประโยคที่สรุปได้ ทำให้ไม่สามารถใส่ประโยค "Betty is married to Colin" เข้ามาได้

นอกจากกฎการวินิจฉัยจะสามารถช่วยชีไอพีในลักษณะของตัวกรองข้อมูลแล้ว มันยังช่วยในลักษณะของตัวขยายข่าวสารด้วย จากลักษณะข้างบน ชีไอพีใช้วิธีการอนุมานไปข้างหน้า สำหรับในตัวอย่างนี้สมมุติว่ามันใช้การอนุมานย้อนหลังร่วมกับการซ้อนทับรูปแบบ (Backward Chaining with Pattern Matching) ซึ่งเป็นลักษณะการทำงานพื้นฐานของภาษาโปรลอก (PROLOG) โดยกำหนดเป้าหมายเป็น "Betty is married to Alan" ชีไอพีจะค้นหากฎที่มีเป้าหมายนี้ในส่วนข้อสรุปของกฎ ซึ่งจากกฎที่ว่า " Y is married to X if X is married to Y " จากการซ้อนทับเป้าหมายในส่วนข้อสรุป Y ก็จะมีค่าเป็น Betty และ X ก็จะมีค่าเป็น Alan

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกฎเป้าหมายจะเป็นจริงได้ก็ต่อเมื่อ "Alan is married to Bettty" ในส่วนเงื่อนไขเป็นจริง ซึ่งก็เป็นเช่นนั้น ทำให้ยืนยันได้ว่าเป้าหมายเป็นจริง จากตัวอย่างที่ผ่านมา ซีไอพีได้ใช้ส่วนประกอบ 4 ส่วนนี้ในการควบคุมดูแลการแก้ไขฐานข้อมูล ตลอดจนกว่าให้ข่าวสารจากการโต้ตอบคำถามของผู้ใช้ จากลักษณะเช่นนี้อาจจะเปรียบคอนเซบชวลสกีมม่าได้ดัง "ผู้บัญญัติกฎ" เนื่องจากเป็นผู้กำหนดกฎและข้อบังคับอื่นๆ สำหรับยูไอดี ส่วนซีไอพีนั้นเปรียบได้ดัง "ผู้รักษากฎ" เนื่องจากมันจะเป็นมันจะดูแลการแก้ไขฐานข้อมูลของผู้ใช้ให้เป็นไปตามกฎ นอกจากนี้ยังเป็นตัวให้ข่าวสารกับผู้ใช้อีกด้วย

สำหรับฟิฟจีไอเอส (5GIS) จะอนุญาติให้มีการแก้ไขฐานข้อมูลได้ 2 วิธีคือการเพิ่มประโยคเข้าไปในฐานข้อมูล และ การลบประโยคออกจากฐานข้อมูล ซึ่งอาจจะทำทีละ 1 ประโยค เรียกว่า ทรานเซคชั่นแบบง่าย (Simple Transaction) หรือการทำทีละหลายๆ ประโยคพร้อมกัน เรียกว่า ทรานเซคชั่นแบบคอมเพานด์(Compound Transaction) ซึ่งใช้คำว่า "begin" และ "end" บ่งบอกจุดเริ่มและจุดสิ้นสุดของทรานเซคชั่น สิ่งที่สำคัญก็คือ ซีไอพีจะพิจารณาเป็นผลรวมๆ (Collective effect) ของทรานเซคชั่น แทนที่จะพิจารณาทีละส่วนย่อยๆ ดังนั้น ซีไอพีอาจจะรับทรานเซคชั่นแบบคอมเพาด์ในขณะที่ถ้าพิจารณาแต่ละส่วนที่เป็นทรานเซคชั่นแบบง่ายๆ แล้วจะไม่ยอมรับ

ตัวอย่างเช่น กลับไปใช้ ยูไอดี การแต่งงานอีกครั้งหนึ่งและพิจารณาทรานเซคชั่นต่อไปนี้โดยเริ่มต้นฐานข้อมูลยังว่างอยู่

User	CIP
add: Alan is married to Betty	accepted
begin	
add: Colin is married to Betty	
delete: Alan is married to Betty	
end	accepted

ผู้ใช้ใส่ทรานเซคชั่น 2 ทรานเซคชั่น ทรานเซคชั่นแรกเป็นแบบง่ายและทรานเซคชั่นหลังเป็นแบบคอมเพานด์ ซีไอพียอมรับทั้งคู่ ถ้าสังเกตจะพบว่าส่วนแรกของคอมเพานด์ทรานเซคชั่น (add: Colin is married to Betty) นั้นขัดแย้งกับคอนสเตนท์ C1 จากตัวอย่างก่อนๆ แต่ถ้าพิจารณาอย่างรวมๆ ทั้งทรานเซคชั่นแล้วจะไม่ขัดแย้งกับคอนเซบชวลสกีมม่า เพื่อให้เข้าใจการรับคอมเพาด์ทรานเซคชั่นของซีไอพีได้ดียิ่งขึ้น ขอให้พิจารณาว่า ซีไอพีจะพิจารณาดูผลกระทบรวมของทรานเซคชั่นทั้งหมดก่อนที่จะทำการเปลี่ยนแปลงจริงกับฐานข้อมูล ซีไอพีจะยอมรับทรานเซคชั่นใดๆ ก็ต่อเมื่อผลกระทบรวมนั้น สอดคล้องกับคอนเซบชวลสกีมม่า

เมื่อซีไอพียอมรับคอมเพานด์ทรานเซคชั่นแล้วมันจะพยายามแก้ไขทีละส่วนของทรานเซคชั่นในฐานข้อมูลเพื่อไม่ให้ขัดแย้งกับคอนเซบชวลสกีมม่า จากตัวอย่างนี้ทำได้โดยการกระทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทรานเซคชั่นย้อนลำดับจากที่ผู้ใช้ให้เข้ามาคือ

delete: Alan is married to Betty

add: Colin is married to Betty

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

แบบจำลองข้อมูลไนแอม (The NIAM Conceptual Schema Model)

แบบจำลองข้อมูลไนแอม เป็นแบบจำลองข้อมูลที่ค้นคิดโดย ศาสตราจารย์ จี เอ็ม ไนเซน (Professor G.M. Nijssen) ในราวปลายปี ค.ศ. 1977 คำว่า "NIAM" ย่อมาจาก Nijssen Information Analysis Methodology

ไนแอมเป็นแบบจำลองข้อมูลแบบหนึ่งที่ทรงประสิทธิภาพเป็นอย่างมาก โดยมีพื้นฐานมาจากแนวความคิดทางภาษาธรรมชาติแบบโครงสร้างลึก (Deep Structure Natural Language Concepts) ซึ่งมีรูปประโยคเป็น <ประธาน, กิริยา, กรรม> เท่านั้น โครงสร้างภาษาธรรมชาติแบบนี้ได้ถูกใช้แทนความหมายที่สมบูรณ์ของ ยูโอดี (UoD : Universe of Discourse) ในคอนเซบชวลสกีมมา (Conceptual Sehcma)

ไนแอมมีส่วนประกอบพื้นฐานดังต่อไปนี้

1. ชนิดเอนติตี้ (Entity Type)
2. ชนิดเลเบ้ล (Label Type)
3. ชนิดอ้างอิง (Reference Type)
4. ชนิดความจริง (Fact Type)
5. คอนสเตนท์ (Constraint)

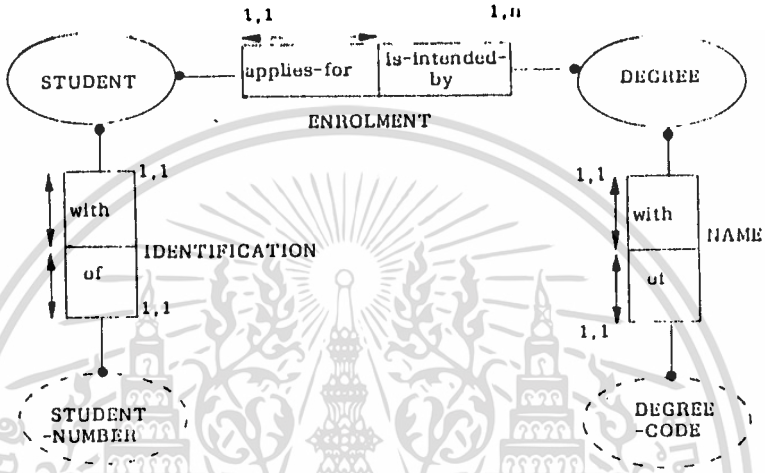
ส่วนประกอบพื้นฐานทั้ง 5 แสดงได้ดังตัวอย่างในรูปที่ 6.1 เอนติตี้เป็นหน่วยพื้นฐานที่สุดของระบบฐานข้อมูล เช่น ในมหาวิทยาลัย อาจารย์นักศึกษา วิชาเรียน ภาควิชา ฯลฯ ถือว่าเป็นเอนติตี้ สำหรับชนิดเอนติตี้นั้น เป็นเซ็ทที่มีสมาชิกเป็นตัวอย่างเอนติตี้ (Entity Instance) เช่น ภาควิชาวิศวกรรมคอมพิวเตอร์ ถือว่าเป็นสมาชิกของชนิดเอนติตี้ ภาควิชา นาย ก นาย ข เป็นนักศึกษาถือว่าเป็นสมาชิกในชนิดเอนติตี้นักศึกษาเป็นต้น

เนื่องจากเอนติตี้ไม่สามารถใช้อ้างถึงตัวมันเองได้ แต่จะอาศัยสิ่งหนึ่งคือ เลเบ้ล (Label) ในการระบุถึงตัวมัน ซึ่งเลเบ้ลเป็นตัวอย่าง (Instances) ในชนิดเลเบ้ล(Label Type) อีกทีหนึ่ง ดังนั้นเลเบ้ลจึงเป็นสิ่งที่ใช้ระบุถึงเอนติตี้ เช่น ชนิดเอนติตี้นักศึกษาอาจจะถูกระบุโดยใช้ชนิดเลเบ้ล ชื่อ หรือ รหัสประจำตัว อย่างใดอย่างหนึ่ง

สำหรับความสัมพันธ์ระหว่างชนิดเอนติตี้กับชนิดเลเบ้ล เรียกว่า ชนิดอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Reference Type) ดังแสดงในรูปที่ 6.1 'STUDENT' และ 'DEGREE' เป็นชนิดเอนติตี้ ส่วน 'STUDENT-NUMBER' และ'DEGREE-CODE' เป็นชนิดเลเบิ้ล 'IDENTIFICATION' และ 'NAME' นั้นเป็นชนิดอ้างอิง ซึ่งใช้อ้างอิงชื่อของชนิดเอนติตี้ 'STUDENT' และ 'DEGREE' ตามลำดับ



รูปที่ 6.1

ตัวอย่างง่าย ๆ ของชนิดอ้างอิง และชนิดเลเบิ้ลมีดังนี้

เลเบิ้ล คือ 102297816, 437863252 (STUDENT-NUMBER)

และ BSc, BE, MBBS, (DEGREE-CODE)

อ้างอิง STUDENT กับ STUDENT-NUMBER 102297816

และ DEGREE กับ DEGREE-CODE BE

สิ่งที่กำหนดความสัมพันธ์ระหว่างชนิดเอนติตี้กับชนิดเอนติตี้ด้วยกันเรียกว่า " ชนิดความจริง " (Fact Type) เงื่อนไขที่จำเป็นประการหนึ่งก็คือ การที่จะแทนชนิดเอนติตี้อันใดอันหนึ่งในคอนเซบชวลสกีมา ชนิดเอนติตี้นั้นจะต้องเกี่ยวข้องกับชนิดความจริงอย่างน้อยหนึ่งอัน ความเกี่ยวข้องของชนิดเอนติตี้ในชนิดความจริงแทนได้ด้วยบทบาท (Role) หรือ กิริยา (Verb) หนึ่งในชนิดความจริง ชนิดความจริงที่มีบทบาทเข้าเกี่ยวข้องเพียงหนึ่งบทบาท เรียกว่า ความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกี่ยวข้องแบบ อันนารี่ (Unary) กรณีที่มีสองบทบาท เรียกว่า ความเกี่ยวข้องแบบไบนารี่ (Binary) หรือกรณีที่มีบทบาทหลายบทบาทเข้าเกี่ยวข้องในชนิดความจริงเรียกว่า ความเกี่ยวข้องแบบ เอ็นนารี่ (N-ary) ทุก ๆ บทบาทของชนิดความจริงจะเกี่ยวข้องกับชนิดเอนติตี้เพียงชนิดเดียวเท่านั้น รูปที่ 6.1 แสดงให้เห็นว่าชนิดความจริง 'ENROLMENT' เป็นชนิดความจริงแบบไบนารี่ สำหรับแต่ละชนิดอ้างอิงก็จะเป็นความเกี่ยวข้องแบบไบนารี่ระหว่างชนิดเอนติตี้กับชนิดเลเบ้ล

จำนวนครั้งที่เอนติตี้หรือเลเบ้ลเกี่ยวข้องกับบทบาทหนึ่งในชนิดความจริงหรือชนิดอ้างอิงสามารถกำหนดได้ด้วยคอนสเตนท์ความถี่ที่ปรากฏบนบทบาทอันนั้น ความถี่ที่ปรากฏของบทบาทใดๆ สามารถแทนได้ด้วยเลขจำนวนธรรมชาติ (Natural Number) n ใด ๆ (เมื่อ n > = 0) โดยที่เลขตัวแรก จะแสดงถึงจำนวนครั้งที่น้อยที่สุดที่แต่ละตัวอย่างของชนิดเอนติตี้ หรือชนิดเลเบ้ลเกี่ยวข้องกับบทบาทนั้น สำหรับตัวเลขตัวที่สองจะบอกถึงจำนวนครั้งที่ตัวอย่างของชนิดเอนติตี้หรือชนิดเลเบ้ลเกี่ยวข้องกับบทบาทนั้นมากที่สุด เช่นเดียวกัน ค่าความถี่ที่มากที่สุดจะต้องมากกว่า 0 เสมอ แต่ค่าที่น้อยที่สุดอาจจะเป็น 0 หรือ 1 ก็ได้ ดูรูปที่ 6.1 ตัวเลข 2 ตัว ที่คั่นด้วย "," นั้น เป็นความถี่ที่ปรากฏของบทบาทของแต่ละชนิดความจริง

ความจริง (Fact) แต่ละความจริง แสดงได้เป็น พรีดิเคต (Predicate) บนเอนติตี้ ซึ่งมีรูปแบบดังนี้

```
<fact> : <predicate) (<entity-type 1, label-type 1, label 1>
                      [[,<entity-type 2, label-type 2, label 2>]...])
```

ดังนั้นความจริงต่าง ๆ ในคอนเซบชวลสกีมมา ในรูปที่ 6.1 เขียนตามรูปแบบของความจริงได้เป็น

```
' The student with student-nember 102297816 applies-for
  the degree with degree-code BSc '
```

หรือ อาจจะเขียนย้อนกลับได้เป็น

```
' The degree with degree-code BSc is-intended-by the
  student with student-number 102297816 '
```

พรีดิเคตในความจริงเหล่านี้มี 'applies-for' และ 'is-inteded-by' ซึ่งเป็นชื่อของบทบาทของชนิดความจริง 'ENROLMENT'

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากข้อดีที่ไนแอม สามารถแสดงความแตกต่างระหว่างเอนติตี้กับเลเบิ้ลได้อย่างชัดเจนส่งผลให้ ไนแอม สามารถแทนความจริงในลักษณะที่ใกล้เคียงกับภาษาธรรมชาติ (Natural Language) ได้โดยอาศัยพรีดิเคต จากลักษณะดังกล่าวทำให้ผู้ใช้สามารถสร้างคอนเซบชวลสกีมาทั้งหมดจากภาษาธรรมชาติได้

สำหรับชนิดเอนติตี้แบบเนสต์ (Nested Entity Types) สร้างมาจากการมองชนิดความจริงเป็นชนิดเอนติตี้ของชนิดความจริงอื่น จากรูปที่ 6.2 ชนิดเอนติตี้แบบเนสต์ 'SUPPLY' มาจากการมองชนิดความจริงแบบเทอนารี่เป็นชนิดเอนติตี้ตัวหนึ่ง ตัวอย่างของชนิดเอนติตี้ประเภทนี้ก็คือ ตัวอย่างของชนิดความจริงนี้ เช่น ตัวอย่างหนึ่งของชนิดเอ็นติตี้ 'SUPPLY' เป็น

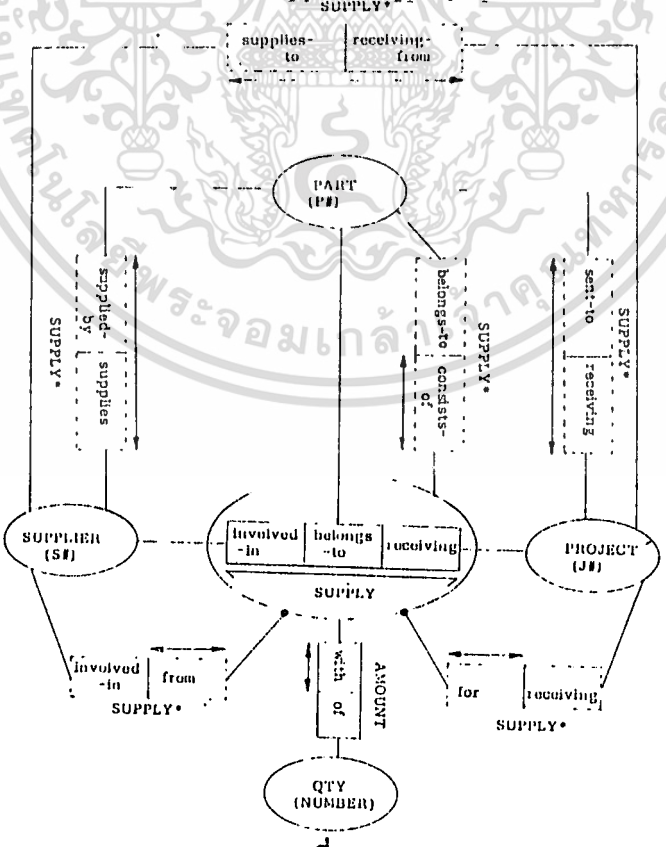
'The supply from John for project J1 consists of part P2 '

ถึงแม้ว่าความจริงนี้จะไม่สามารถแตกย่อยออกเป็นความจริงย่อยๆ ได้แต่ผู้ใช้รับเอา (Retrieve) ความจริงย่อยเหล่านี้ไปได้ คือ

'The part P2 is supplied by the supplier John '

และ

'The part P2 is sent to the project J1 '



รูปที่ 6.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะเด่นอย่างหนึ่งของไนแอม คือ สามารถแสดงกฎเกณฑ์ข้อบังคับต่างๆ หรือคอนสเตนท์ความจริงที่มีอยู่ในคอนเซบชวลสกีมม่า ซึ่งคอนสเตนท์นี้จะเป็นตัวกำหนดความถูกต้อง (Consistency) ของฐานข้อมูล สามารถอธิบายคอนสเตนท์ต่างๆ ที่มีในไนแอมได้ดังนี้

- ถ้าตัวอย่างของชนิดความจริงถูกกำหนดโดยสับเซตที่น้อยที่สุดของบทบาทบนชนิดความจริงนั้น โดยที่กลุ่มของบทบาทนั้นเป็นไอเดนติไฟเออน์ (Identifier) สำหรับชนิดความจริงนั้นหมายถึงชุดเอนติตี้ที่เกี่ยวข้องกับกลุ่มบทบาทนั้นทั้งหมดนั้น เกิดขึ้นได้ครั้งเดียวเท่านั้น คอนสเตนท์ประเภทนี้เรียกว่า ยูนิกเนสคอนสเตนท์ (Uniqueness Constraint) หรือ ไอเดนตไฟเออร์คอนสเตนท์ (Identifier Constraint) แสดงได้โดยเส้นตรงที่มีลูกศรทั้งสองข้างบนกลุ่มบทบาทที่เกิดขึ้นได้ครั้งเดียวนั้น สำหรับชนิดความจริงใดๆ ที่มี N บทบาทแล้ว จะต้องมียูนิกเนสคอดสเตนท์อย่างน้อย 1 อันที่เกี่ยวข้องกับบทบาทอย่างน้อย N-1 บทบาท ตัวยูนิคเนสคอนสเตนท์นี่เอง ที่ใช้ในการกำหนดคีย์ในระดับสกีมม่า ของฐานข้อมูลแบบรีเลชั่นนอล (Relational Database) จากรูป 6.3 บทบาท R1 เป็นยูนิกเนสคอนสเตนท์ของชนิดความจริง F1 นอกจากนี้ยูนิกเนสคอนสเตนท์ก็สามารถกำหนดบนชนิดอ้างอิงได้ด้วย

- ถ้ามีโททอลโรลคอนสเตนท์ (Total Role Constraint) กำกับที่บทบาท R ที่มีชนิดเอนติตี้หรือชนิดเลเบ้ล X เกี่ยวข้องอยู่แล้ว ทุกๆตัวอย่าง X ที่เป็นไปได้จะต้องเข้ามาเกี่ยวข้องกับบทบาท R อย่างน้อย 1 ครั้ง โททอลโรลคอนสเตนท์แสดงได้โดยจุดที่ตำแหน่งที่บทบาทเชื่อมต่อกับชนิดเอ็นติตี้หรือชนิดเลเบ้ล สำหรับในระดับของสกีมม่าของฐานข้อมูลแบบรีเลชันนอล โททอลโรลคอนสเตนท์จะเป็นตัวระบุว่าอนุญาตให้มีค่านัล (Null Values) ที่ส่วนที่เกี่ยวข้องกับบทบาทนั้นหรือไม่ ดูรูปที่ 6.3

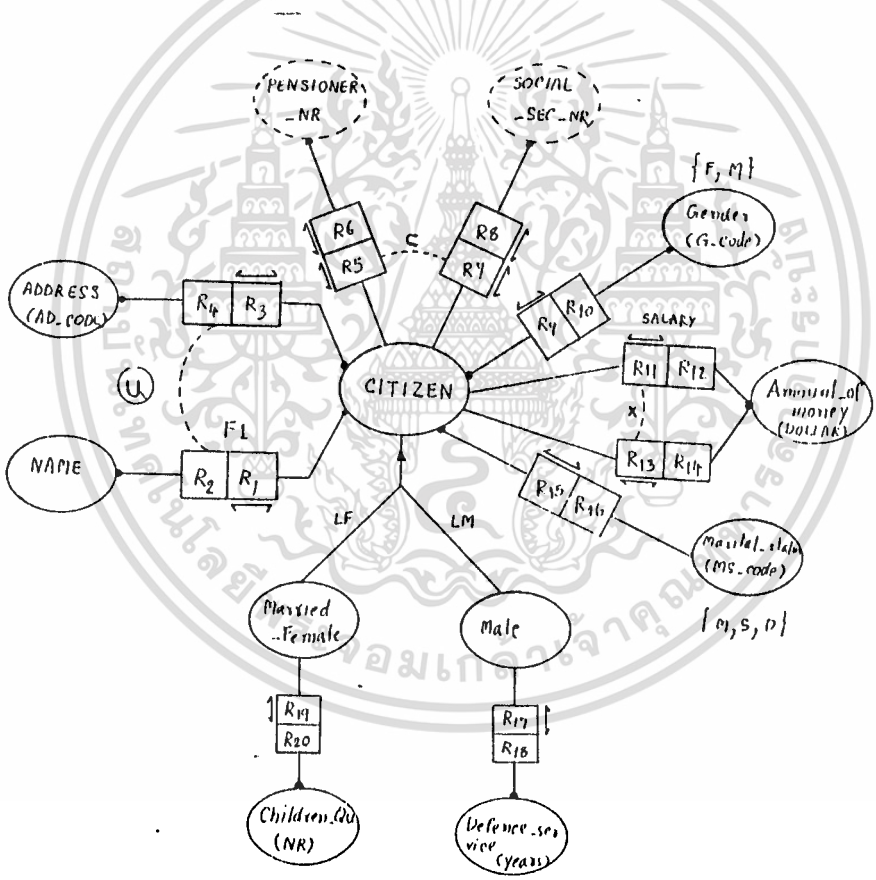
- สำหรับชนิดความจริงหรือชนิดอ้างอิง แต่ละอย่างของมันจะกำหนดโดยยูนิกเนสคอนสเตนท์ เช่นเดียวกันสำหรับชนิดเอนติตี้ ถ้าต้องการการอ้างอิงที่เป็นเอกลักษณ์เฉพาะของแต่ละตัวอย่างเอนติตี้ การอ้างอิงแบบนี้เปรียบเสมือนกลุ่มของชนิดเลเบ้ลที่การผสมกันของเลเบ้ลใช้อ้างถึงเอนติตี้แต่ละตัวเท่านั้น เราเรียกการอ้างอิงเช่นนี้ว่า การอ้างอิงชื่อ (Naming Reference)

- ถ้ามีสับเซตคอนสเตนท์ (Subset Constraint) ระหว่างบทบาท R_{A1} และ R_{A2} ของชนิดความจริง (หรือชนิดอ้างอิง) S1 และ S2 ตามลำดับ และบทบาททั้งสองเกี่ยวข้องกับชนิดเอนติตี้ A แล้ว หมายความว่า ตัวอย่างเอนติตี้ทุกเอนติตี้ของ A ที่เกี่ยวข้องกับ R_{A1} จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต้องเกี่ยวข้องกับ R_{A2} (RA1 C RA2) สังเกตว่าสับเซตคอนสเตนท์ไม่สนใจจำนวนครั้งเอนติตี้ปรากฏในชนิดความจริงทั้งสอง จากรูป 6.3 สับเซตคอนสเต่นท์บอกว่าถ้าเลเบ้ลใดเป็นชื่อเริ่มแรก(Initials) แล้วจะต้องเป็นชื่อสกุล (Surname) ด้วย

- ถ้ามีเอ็กคลูชั่นคอนสเตนท์ระหว่างบทบาท R_{B1} และ R_{B2} ของชนิดความจริง (หรือชนิดอ้างอิง) S1 และ S2 ตามลำดับ และบทบาททั้งคู่เกี่ยวข้องกับชนิดเอนติตี้ B แล้ว หมายความว่า ถ้าเอนติตี้ตัวอย่างใดปรากฏที่ R_{B1} แล้ว จะไม่ปรากฏที่ R_{B2} (R_{B1} X R_{B2}) หรือตรงกันข้าม กล่าวอีกความหมายหนึ่งก็คือ การอินเตอร์เซคชั่นกันระหว่างเซตของเอนติตี้ที่เกี่ยวข้องกับบทบาททั้งสองจะได้ผลลัพธ์เป็นเซตว่าง ดังนั้นบทบาทไหนที่เกี่ยวข้องกับเอ็กคลูชั่นคอนสเตนท์แล้วจะต้องไม่เกี่ยวข้องกับโททอลโรลคอนสเตนท์



รูปที่ 6.3 แสดงคอนเซบชวลสกีมม่าที่มีคอนสเตนท์ต่าง ๆ

สำหรับบทนี้เป็นแต่เพียงบทแนะนำแบบจำลองข้อมูลไนแอมเท่านั้น ผู้สนใจสามารถที่จะหารายละเอียดเพิ่มเติมได้จากเอกสารอ้างอิง ที่มีในตอนท้ายของวิทยานิพนธ์ฉบับนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7

ไนแอมเบสฟิฟจิไอเอส

(The NIAM-based 5th Generation Information System)

7.1 เมต้าคอนเซบชวลสกีมม่า(META-CONCEPTUAL SCHEMA)

จากกฎข้อที่ 5 ของฟิฟจิไอเอส (5GIS) ในบทที่ 4 กล่าวว่า

"คอนเซบชวลสกีมม่าอาจจะถูกมองเป็นฐานสารสนเทศได้ ซึ่งสถานะของฐานสารสนเทศที่ยอมรับได้ถูกกำหนดโดยเมต้าคอนเซบชวลสกีมม่า"

เนื่องจากฟิฟจิไอเอสจะไม่ยอมให้มีคอนเซบชวลสกีมม่าที่ผิดผลาดอยู่ในระบบได้ เนื่องจากคอนเซบชวลสกีมม่าเป็นตัวที่ใช้อธิบายยูโอดี ถ้ามันเกิดผิดพลาดแล้วก็ย่อมไม่สามารถใช้อธิบายยูโอดีที่ต้องการ ทำให้เราไม่สามารถใช้งานฟิฟจิไอเอสได้ ดังนั้นระบบที่จะเรียกว่า ฟิฟจิไอเอส จำเป็นจะต้องเป็นไปตามกฎที่ 6 คือ ระบบจะมีประโยคกลุ่มหนึ่ง ซึ่งทำหน้าที่อธิบายลักษณะของโครงสร้างของคอนเซบชวลสกีมม่า เราเรียกกลุ่มประโยคนี้ว่า เมต้าคอนเซบชวลสกีมม่า ตามข้อกำหนดของฟิฟจิไอเอสจะพบว่าเมต้าคอนเซบชวลสกีมม่าจะเป็นส่วนหนึ่งในคอนเซบชวลสกีมม่า ทำให้ผู้ใช้สามารถถามข้อมูลเกี่ยวกับโครงสร้างของคอนเซบชวลสกีมม่าจากซีไอพีได้ หน้าที่ของเมต้าคอนเซบชวลสกีมม่าตามกฎข้อที่ 5 นั้น มันจะถูกใช้ในหน้าที่ของตัวกรองการออกแบบ (Design Filter) เพื่อกลั่นกรองหาข้อผิดพลาดของโครงสร้างของคอนเซบชวลสกีมม่าที่ออกแบบไว้แล้ว ซึ่งจะนำมาใส่ให้กับระบบ ดังนั้น ระบบจะต้องมีเมต้าคอนเซบชวลสกีมม่าอยู่ในตอนเริ่มแรกในขณะที่ยังไม่มีคอนเซบชวลสกีมม่า ซึ่งจะถูกนำมาใส่ให้กับระบบในภายหลัง ตัวตรวจสอบคอนเซบชวลสกีมม่า (Conceptual Schema Checker) จะใช้มันในการตรวจสอบคอนเซบชวลสกีมม่าเสียก่อน ที่จะทำการแทนคอนเซบชวลสกีมม่าเข้าสู่ระบบ แล้วจึงอนุญาตให้ผู้ใช้ทำการแก้ไขฐานข้อมูล และ ใช้ข้อมูลในฐานข้อมูลต่อไป ซึ่งเป็นหน้าที่ของซีไอพี

ในระบบฟิฟจิไอเอสนั้นจะใช้กลุ่มประโยค (Sentences) แทนคอนเซบชวลสกีมม่า ซึ่งไนแอมก็ถูกค้นคิดขึ้นมาใช้แทนคอนเซบชวลสกีมม่าเช่นเดียวกัน เนื่องจากมีพื้นฐานมาจากภาษาธรรมชาติ แบบโครงสร้างลึก (Deep Structured Natural Language) ดังนั้นโครงสร้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างไรก็ตามเมต้าคอนเซบชวลสกีมม่านี้ยังรับประกันไม่ได้ว่าสามารถอธิบายครอบคลุมกฎไวยากรณ์ไนแอมได้ทุกกรณี เนื่องจากถ้าจะทำให้ได้เช่นนั้นจะต้องใช้วิธีคิดค้นและทดสอบแบบเป็นเวลานาน เพื่อให้สามารถครอบคลุมได้ทุกกรณี แต่ถึงอย่างไรก็ตามเมต้าคอนเซบชวลสกีมม่าต้นแบบนี้ก็สามารถนำไปตรวจสอบความถูกต้องของคอนเซบชวลสกีมม่าที่นำมาทดสอบได้ผลเป็นที่น่าพอใจ เมื่อไรที่มีการพัฒนาให้ดียิ่งขึ้น ก็สามารถเปลี่ยนเข้าไปใช้ในไนแอมเบสฟิฟจีไอเอสของเราใหม่ได้โดยเปลี่ยนเฉพาะส่วนฐานข้อมูลของโปรลอกที่แทนเมต้าคอนเซบชวลสกีมม่า สำหรับส่วนโปรแกรมไม่จำเป็นต้องแก้ไขใดๆ ทั้งสิ้น

7.1.2 ส่วนตรวจสอบคอนเซบชวลสกีมม่า (Conceptual Schema Checker)

อย่างไรก็ตามลำพังจะอาศัยเฉพาะเมต้าคอนเซบชวลสกีมม่าในการตรวจสอบความผิดพลาดของคอนเซบชวลสกีมม่านั้นยังทำไม่ได้ เนื่องจากเมต้าคอนเซบชวลสกีมม่าเป็นเพียงส่วนที่เป็นความรู้ (คำบรรยาย) ที่เกี่ยวกับกฎไวยากรณ์จะต้องมีส่วนที่ทำหน้าที่ใช้ความรู้ในการตรวจสอบ เราจะเรียกส่วนนี้ว่าส่วนตรวจสอบแบบจำลองข้อมูลไนแอม

ส่วนนี้ทำหน้าที่ตรวจสอบการปอปปูเลทเมต้าคอนเซบชวลสกีมม่า โดยตัวอย่างความจริงทั้งหมดในที่นี้คือคอนเซบชวลสกีมม่าว่าสอดคล้องตามเมตาคอนเซบชวลสกีมม่าหรือไม่ โดยการปอปปูเลทตัวอย่างความจริงทั้งหมด (คอนเซบชวลสกีมม่า) ลงในเมต้าคอนเซลชวลสกีมม่า แล้วทำการตรวจสอบว่าเป็นไปตามโดเมนยูนิกเนสคอนสเตนท์และโรลคอนสเตนท์หรือไม่ ถ้าเกิดการขัดแย้งขึ้นส่วนนี้จะทำการรายงานว่าการขัดแย้งโดเมนหรือคอนสเตนท์ที่ชนิดความจริงไหนบ้าง เพื่อให้ข้อมูลแก่ผู้ใช้นำไปแก้ไขคอนเซบชวลต่อไป ซึ่งหน้าที่ดังกล่าวนี้เรียกว่า ตัวกรองการออกแบบ (Design Filter) การรายงานอาจจะมีการรายงานที่เป็นข้อผิดผลาดที่จะต้องแก้ไขให้ถูกต้องก่อนหรือการรายงานที่บอกเตือนว่า คอนเซบชวลสกีมม่านี้ไม่มีส่วนประกอบบางส่วน เช่น รายงานว่าคอนเซบชวลสกีมม่านี้ไม่มีชนิดอ้างอิง เป็นการเตือนผู้ใช้เท่านั้น ถ้าไม่มีการรายงานข้อผิดพลาดที่ผู้ใช้จะต้องแก้ไขแล้ว ก็ถือว่าคอนเซบชวลสกีมม่านี้ผ่านตัวกรองการออกแบบ ซึ่งถือว่าคอนเซบชวลสกีมม่านี้แทนโดยไนแอมถูกต้องแล้ว จะถูกนำไปแทนเป็นส่วนคอนเซบชวลสกีมม่าในไนแอมเบสฟิฟจีไอเอสต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดย<domain>ในที่นี้จะเป็นเซตของตัวอย่างชนิดที่เป็นไปได้ของชนิด <type> สำหรับคอนเซบชวลสกีมม่านี้สามารถประกาศชนิดเอนติตี้ได้ว่า

```
type_domain(entity,[adherent,nr_session,religion,speaker,text,
        main_teacher]).
```

จะสังเกตเห็นว่า ส่วนประกาศโดเมนชนิดมีลักษณะคล้ายกับการประกาศชนิด แบบอินิวเมอร์เรท (Enumerate Type) ในภาษาปาสคาล (PASCAL)

2.ส่วนแทนคอนเซบชวลสกีมม่า ประกอบด้วย

2.1 ส่วนประกาศโดเมนของชนิดเลเบ้ลซึ่งมีรูปแบบดังนี้

```
fact_domain(<domain name>,<domain type>,<domain>).
```

<domain name> เป็นส่วนที่บอกชื่อของโดเมน

<domain type> เป็นส่วนที่บอกประเภทของโดเมน ซึ่งมี 3 ประเภท คือ

1. ประเภทอะตอม (atom) จะมีส่วน <domain> เป็นอินิวเมอร์เรทเซต

2. ประเภทเลขจำนวนเต็ม (integer) จะมีส่วน <domain> เป็นช่วงขอบเขตของตัวเลขจำนวนเต็ม มีค่าตั้งแต่ <minimum integer> ไปถึง <maximum integer> ซึ่งมีรูปแบบเป็น [<minimum intrger>,<maximum integer>]

3. ประเภทเลขจำนวนจริง (real) จะมีส่วน <domain> เป็นช่วงขอบเขตของตัวเลขจำนวนจริง มีค่าตั้งแต่ <minimum real> ไปถึง <maximum real> ซึ่งมีรูปแบบเป็น [<minimum real>,<maximum real>]

4. ประเภทใดๆ คือ เป็นประเภทอะไรก็ได้ที่มีขนาดของโดเมนใหญ่มากจนไม่สามารถกำหนดไว้ได้ ผู้ออกแบบคอนเซบชวลสามารถตั้งเป็นชื่อประเภทอะไรก็ได้ ซึ่งต้องเป็นชื่อของอะตอมใด ๆ ในโปรลอก

ตัวอย่างเช่น กำหนดว่าจำนวนครั้งของการพูดของคนหนึ่งคนใดในหัวข้อศาสนาเดียวมีได้ตั้ง แต่ 1 ครั้งไปจนถึง 10 ครั้ง ดังนั้น โดเมนของจำนวนครั้งในการพูดซึ่งให้ใช้ชื่อว่า 's_o_session_num' สามารถประกาศได้ดังนี้

```
fact_domain(s_o_session_num,integer,[0,10]).
```

สำหรับ

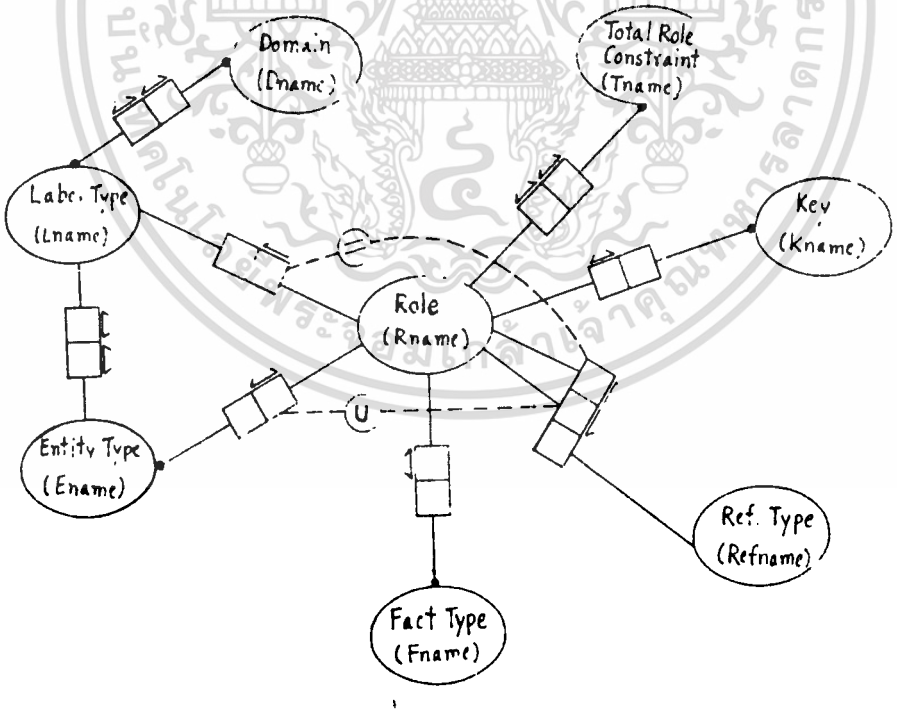
```
fact_domain(mteacher-name,atom,[christ,mohammed,krishna,gautama,
        confucius]).
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นการประกาศโดเมนชื่อศาสดาของศาสนา

2.2 ส่วนปอปปูเลชั่นของเมตาคอนเซบชวลสกีมม่า(Meta-conceptual Schema Population) จากที่ได้กล่าวถึงเมตาคอนเซบชวลสกีมม่าไปแล้วในหัวข้อที่ 7.1 เห็นว่าเมตาคอนเซบชวลสกีมม่านั้นเป็นโครงร่างของคอนเซบชวลสกีมม่า ดังนั้น คอนเซบชวลสกีมม่าก็คือ ปอปปูเลชั่นของเมต้าคอนเซบชวลสกีมม่านั่นเอง ในการแทนชนิดความจริงของไนแอม เราใช้ความจริงของโปรลอก โดยให้ฟังก์เตอร์ (Functor) แทนชื่อของชนิดความจริง และ อาร์กูเมนต์ (Arguments) แทนเอนติตี้ที่เกี่ยวข้องกับบทบาทที่ประกอบกันเป็นชนิดความจริงนี้ โดยเรากำหนดตำแหน่งของบทบาทที่เกี่ยวข้องกับชนิดความจริงเป็นลำดับที่แน่นอน แต่ตามแนวคิดของไนแอมแล้วบทบาทเหล่านี้สามารถเปลี่ยนตำแหน่งเพื่อประกอบกันเป็นชนิดความจริงได้อย่างอิสระ

NIAM ของเมต้าคอนเซบชวลสกีมม่าแสดงได้ดังรูปที่ 7.4



รูปที่ 7.4 แสดงเมต้าคอนเซบชวลสกีมม่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับการแทนคอนเซบชวลสกีมาทั้งหมด ในรูปของความจริงของโปรลอก แสดงได้ดังนี้

```
/*************************************************************/
/* Represent CS of a conference on the five major religions  */
/*************************************************************/
/*****************************/
/* part 1 declare type domain */
/*****************************/

/* s_o = set_of , c = code , r = range , nr = number */
type_domain(domain,[s_o_text_name,s_o_session_num,s_o_religion_c,
            million_r,mteacher_name,person_name,s_o_text_c,
            s_o_religion_name]).
type_domain(label,[text_name,session_num,religion_c,million,
            person_name,text_c,mteacher_name,religion_name]).
type_domain(entity,[adherent,nr_session,religion,speaker,text,
            main_teacher]).
type_domain(role,[r1,r2,r3,r4,r5,r6,r7,r8,r9,r10,r11,r12,r13,r14,r15]).
type_domain(facttype,[f1,f2,f3,f4,f5]).
type_domain(reftype,[rf1,rf2]).
type_domain(mand,[t1,t2,t3,t4,t5,t6,t7,t8,t9,t10]).
type_domain(key,[k1,k2,k3,k4,k5,k6,k7,k8,k9,k10]).
/*****************************************************/
/*      part 2 conceptual schema representation      */
/* part 2.1 declare the domains of the entitys that involve facte */

fact_domain(s_o_text_name,atom,[new_testament,old_testament,koran,
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            bhagavad_gita,upanishads,sutras,analects]).
fact_domain(s_o_session_num,integer,[1,10]).
fact_domain(s_o_religion_name,atom,[buddhism,christianity,
            confucianism,hinduism,islam]).
fact_domain(s_o_religion_c,atom,[bud,chr,con,hin,isl]).
fact_domain(million_r,integer,[300,2000]).
fact_domain(person_name,atom,[chandra,davis,green,jones,smith]).
fact_domain(mteacher_name,atom,[christ,mohammed,krishna,gautama,
            confucius]).
fact_domain(s_o_text_c,atom,[a,bg,k,nt,ot,s,u]).

/* part 2.2 Meta-Conceptual Schema population */
label_domain(text_name,s_o_text_name).
label_domain(session_num,s_o_session_num).
label_domain(religion_c,s_o_religion_c).
label_domain(million,million_r).
label_domain(person_name,person_name).
label_domain(text_c,s_o_text_c).
label_domain(mteacher_name,mteacher_name).
label_domain(religion_name,s_o_religion_name).

entity_label(adherent,million).
entity_label(nr_session,session_num).
entity_label(religion,religion_c).
entity_label(speaker,person_name).
entity_label(main_teacher,mteacher_name).
entity_label(text,text_c).
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
role_entity(r1,adherent).

role_entity(r2,religion).

role_entity(r4,religion).

role_entity(r6,religion).

role_entity(r10,religion).

role_entity(r12,religion).

role_entity(r5,nr_session).

role_entity(r7,speaker).

role_entity(r9,speaker).

role_entity(r3,main_teacher).

role_entity(r11,text).

role_entity(r14,text).

role_entity(r8,text).


role_label(r13,religion_name).

role_label(r15,text_name).


facttype_role(f1,r1).

facttype_role(f1,r2).

facttype_role(f2,r3).

facttype_role(f2,r4).

facttype_role(f3,r5).

facttype_role(f3,r6).

facttype_role(f3,r7).

facttype_role(f4,r10).

facttype_role(f4,r11).
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
facttype_role(f5,r8).
facttype_role(f5,r9).

reftype_erole_lrole(rf1,r12,r13).
reftype_erole_lrole(rf2,r14,r15).

tconstr_role(t1,r2).
tconstr_role(t2,r4).
tconstr_role(t3,r10).
tconstr_role(t4,r12).
tconstr_role(t5,r3).
tconstr_role(t6,r7).
tconstr_role(t7,r11).
tconstr_role(t8,r14).
tconstr_role(t9,r13).
tconstr_role(t10,r15).

key_role(k1,r2).
key_role(k2,r3).
key_role(k3,r4).
key_role(k4,r6).
key_role(k4,r7).
key_role(k5,r8).
key_role(k5,r9).
key_role(k6,r11).
key_role(k7,r12).
key_role(k8,r13).
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
key_role(k9,r14).
key_role(k10,r15).
/*****************************************/
/*  part 3 assign query answer for facttype  */
/*****************************************/
fact_nl(rf1,['is the code of the religion names ','']).
fact_nl(rf2,['is the code of the text names ','']).
fact_nl(f1,[' millions have religion code ','']).
fact_nl(f2,['is the main teacher of religion ','']).
fact_nl(f3,[' times that',' religion is represented by ','']).
fact_nl(f4,[' religion has text ','']).
fact_nl(f5,[' book is read by ','']).
```

คอนเซบชวลสกีมม่าใดที่จะถูกแทนในระบบ จะต้องเป็นคอนเซบชวลสกีมม่าที่ผ่านตัวกรองการออกแบบ (Design Filter) เสียก่อน โดยเป็นการทำหน้าที่ของ ตัวตรวจสอบคอนเซบชวลสกีมม่า (Conceptual Schema Checker) ถ้าเกิดข้อผิดผลาดที่ต้องแก้ไข เนื่องจากแสดงคอนเซบชวลสกีมม่าผิด ผู้ออกแบบต้องแก้ไขให้ถูกต้องเสียก่อนโดยอาศัยข้อความของการรายงานข้อผิดพลาดประกอบ แต่ถ้าคอนเซบชวลที่ออกแบบมาถูกต้อง (สอดคล้องตามกฎไวยกรณ์ไนแอม) แล้วคอนเซบชวลสกีมม่านี้ถึงจะถูกเก็บไว้ในระบบ

ตัวกรองการออกแบบมีการทำงาน 3 ขั้นตอน คือ

1. การตรวจสอบโดเมนชนิด
2. การตรวจสอบคีย์ (ยูนิกเนสคอนเตนท์) บนชนิดความจริง
3. การตรวจสอบแมนดาโทรี่ (โททอลโรลคอนสเตนท์) ของบทบาทบนชนิดความจริง

1. การตรวจสอบโดเมนชนิด จะตรวจสอบดูว่า ปอปปูเลชั่นเป็นไปตามโดเมนชนิดที่ประกาศไว้หรือไม่ถ้าไม่จะมีการรายงานออกมาว่า เลเบ้ลของเอนติตี้ไหนขัดแย้งกับโดเมนชนิดอะไร โดยพรีดิเขต(Predicate) ที่ทำการตราจสอบคือ 'm_ck_type_domaim' ซึ่งจะรับเอาอาร์กูเมนต์เป็นชนิดความจริงทั้งหมดแล้วไปค้นหาเอนติตี้ที่เกี่ยวข้องกับชนิดความจริงนั้น ทำการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตรวจสอบโดเมน ถ้าเกิดข้อผิดพลาดจะสะสมไว้ใช้ให้พรีดิเขต 'err_ck_domain' เป็นตัวรายงานข้อผิดพลาดทั้งหมดให้ผู้ออกแบบได้รับทราบ โดยพ้องว่าเอนติตี้ไหนขัดแย้งกับโดเมนชนิดไหน

2. การตรวจสอบคีย์ (ยูนิกเนสคอนสเตนท์) บนชนิดความจริง จะทำการตรวจสอบคีย์บนทุกชนิดความจริงที่มีบทบาทที่ถูกกำกับโดยยูนิกเนสคอนสเตนท์ โดยใช้พรีดิเขต 'm_ck_key_all_f' ซึ่งต้องการอาร์กูเมนต์ชนิดความจริงทั้งหมดที่มีบทบาทที่เป็นคีย์ โดยกำหนดให้กลุ่มของบางบทบาทเป็นคีย์แล้วทดสอบดูว่าเป็นจริงหรือไม่ ถ้าไม่จะมีการบันทึกไว้ว่าการขัดแย้งคีย์ที่ชนิดความจริงอะไรบ้าง ซึ่งพรีดิเขต 'err_ck_key' จะทำหน้าที่รายงานข้อผิดพลาดดังกล่าว

3. การตรวจสอบเมนดาโทรี่(โททอลโรลคอนสเตนท์)ของบทบาทบนชนิดความจริง จะตรวจสอบกับทุกชนิดความจริงที่มีบทบาทใดถูกกำกับด้วยโททอลโรลคอนสเตนท์ โดยจะทำทีละบทบาทในแต่ละชนิดความจริง โดยพิจารณาดูว่า ทุกเอนติตี้ที่เกี่ยวข้องกับบทบาทนั้นทั้งหมดครบถ้วนตามโดเมนของเอนติตี้นั้นหรือไม่ ถ้าไม่แสดงว่าเกิดข้อผิดพลาดขึ้น จะมีการบันทึกไว้ว่าบทบาทไหนในชนิดความจริงใดขัดแย้งกับโดเมนชื่อว่าอะไร พรีดิเขตที่ทำหน้าที่ดังกล่าวคือ 'm_ck_mand_all_f' ซึ่งต้องการอาร์กูเมนต์เป็นชนิดความจริงทั้งหมดที่มีบทบาทถูกกำกับโดยโททอลโรลคอนสเตนท์ พรีดิเขต 'err_ck_mand' จะทำหน้าที่รายงานข้อผิดพลาดดังกล่าว

ถึงแม้ว่าเราจะสามารถแทนคอนเซบชวลสกีมม่าอยู่ในรูปของปอปปูเลชั่นของเมต้าคอนเซบชวลสกีมม่าก็ตาม แต่การแทนในลักษณะนี้ยังไม่เหมาะกับการนำไปใช้งานโดยซีไอพี (CIP) เนื่องจากจะต้องมีการค้นหาข่าวสารเป็นจำนวนมาก ทุกครั้งที่ต้องการใช้มัน เช่น เมื่อผู้ใช้ต้องการปอปปูเลทคอนเซบชวลสกีมม่าโดยการเพิ่มความจริงเข้ามา 1 ความจริง ซีไอพีจะต้องตรวจสอบว่าความจริงนี้ เป็นตัวอย่างของชนิดความจริงอะไรแล้วมีในคอนเซบชวลสกีมม่าหรือไม่ ถ้ามีจะต้องค้นหาต่อไปอีกว่าประกอบจากบทบาทอะไรบ้าง แล้วดูว่าบทบาทถูกกำกับโดยคีย์ หรือโททอลโรลคอนสเตนท์ไหม เพื่อจะใช้ตรวจสอบความจริงที่รับเข้ามาให้สอดคล้องกับคอนเซบชวลสกีมม่า จะเห็นว่าเป็นการสิ้นเปลืองเวลาเป็นอันมาก ดังนั้นผู้จัดทำโครงงานจึงได้เสนอวิธีการแปลงคอนเซบชวลสกีมม่าให้อยู่ในรูปแบบที่ซีไอพีสามารถนำไปใช้ได้ง่ายและรวดเร็ว และพยายามรักษาการแทนความหมายของคอนเซบชวลสกีมม่าให้มากที่สุด รูปแบบดังกล่าวแสดงได้ดังนี้

```
[<fact type>,<set of roles>,<set of some roles that indicated by key>,
 <set of some roles that indicated by total role constraint>,
 <set of domain names>]
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากว่าผู้ใช้จะติดต่อกับซีไอพีโดยใช้ความจริงเท่านั้น และจากชนิดความจริงก็จะสามารถชักจูงไปค้นหาข่าวสารต่างๆ ในคอนเซบชวลสกีมม่าได้ ดังนั้นระบบของเราจึงถึงว่าการรวบรวมข่าวสารเกี่ยวกับชนิดความจริงจะสามารถแทนความหมายของคอนเซบชวลสกีมม่าได้มากพอสมควร แม้ว่าในบางครั้งอาจจะต้องไปใช้ข่าวสารในเมตาคอนเซบชวลสกีมม่าก็ตาม แต่ก็ทำให้ทุ่นเวลาและสะดวกต่อการสร้าง (Implementation) มาก

รูปแบบดังกล่าวอธิบายได้ว่า

"ชนิดความจริง <fact type> ประกอบด้วยกลุ่มบทบาทคือ <set of roles> ซึ่งกลุ่มบทบาทที่ถูกกำกับโดยคีย์มี <set of some roles that indicated by key> และกลุ่มบทบาทที่ถูกกำกับโดยโททอลโรลคอนสเตนท์มี <set of some roles that indicated by total role constraint> โดยที่แต่ละเอนติตี้ที่ยุ่งเกี่ยวกับบทบาททั้งหมดมีโดเมน <set of domain names> ตามลำดับของบทบาทใน <set of roles> "

เหตุผลที่เราแทนคอนเซบชวลสกีมม่าในลักษณะเช่นนี้เนื่องจาก เราได้กำหนดไว้ว่า

"ชนิดความจริงสร้างจากบทบาทต่างๆ ที่มีตำแหน่งแน่นอน"

และการแทนชนิดความจริงเช่นนี้ทำให้ซีไอพีสามารถทำการตราจสอบการแก้ไขฐานข้อมูลของผู้ใช้ได้รวดเร็วขึ้น

พรีดิเขตที่ทำหน้าที่ในการสร้างแบบแทนนี้ คือ "record_fact_def" ซึ่งต้องการชื่อชนิดความจริงที่ต้องการแทน รูปแบบการแทนชนิดความจริงที่ได้จะมี 2 ประเภท คือ

1. ประเภทที่ใช้บอกข่าวสารเกี่ยวกับชนิดความจริง จากรูปแบบการแทนความจริงที่กล่าวถึงก่อนหน้านี้ ซึ่งบทบาทที่กล่าวถึงในแบบจะแทนด้วยชื่อของบทบาทนั้น เช่น แบบที่แทนชนิดความจริง f2 แสดงได้ดังนี้

[f2,[r3,r4],[[r3],[r4]],[r3,r4],[mteacher_name,s_o_religion_c]]

2. ประเภทที่ซีไอพีจะใช้ในการดูแลการแก้ไขฐานข้อมูลของผู้ใช้ ประเภทนี้แตกต่างกับประเภทแรก คือบทบาทในประเภทนี้ จะเป็นชื่อตัวแปรที่มีตำแหน่งสมนัยกับบทบาทนั้น ดังนั้นจากตัวอย่างการแทนในประเภทที่ 1 แสดงเป็นประเภทที่ 2 ได้ดังนี้

[f2,[R1,R2],[[R1],[R2]],[R1,R2],[mteacher_name,s_o_religion_c]]

ตัวแปรเหล่านี้โปรแกรมจะใช้เพื่อนำไปเก็บเอนติตี้ทั้งหมดที่เกี่ยวข้องกับบทบาทที่ตำแหน่งที่ตัวแปรนั้นอยู่ ซึ่งจะกล่าวถึงการใช้ตัวแปรเหล่านี้ในส่วนของซีไอพี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของพรีดิเขต 'record_fact_def' เริ่มโดยพรีดิเขต 'pack_fact' ทำการรวบรวมบทบาททั้งหมดที่ประกอบกันเป็นขนิดความจริงนั้นโดยถือตามลำดับของการจัดเก็บในฐานข้อมูลของโปรลอกจากตัวอย่างจะได้กลุ่มบทบาทเป็น [r3,r4] จากกลุ่มบทบาทนี้ก็ถูกนำไปหา บทบาทที่ถูกกำกับโดยคีย์หรือโททอลโรลคอนสเตนท์ โดยใช้พรีดิเขต 'pack_key' และ 'mand_o_f' จะได้ผลออกมาเป็น[[r3],[r4]] และ[r3,r4] ตามลำดับ สำหรับโดเมนของเอนติตี้ที่เกี่ยวข้องกับบทบาทหามาได้โดยพรีดิเขต 'domain_o_f' ก็จะได้กลุ่มโดเมนเป็น[mteacher_name,s_o_religion_c] หลังจากนั้นแต่ละส่วนจะถูกเก็บเป็นอาร์กูเมนต์ของฟังก์เตอร์ 'cs_fact_info' ไว้ในฐานข้อมูล (Database) ของโปรลอกเป็น

```
cs_fact_info(f2,[r3,r4],[[r3,[r4]],[r3,r4],[mteacher_name,
                s_o_religion_c]).
```

สำหรับประเภทที่ 2 นั้นได้มาจากการนำ [r3,r4] ไปเปลี่ยนเป็นตัวแปร [R1, R2] โดยพรีดิเขต'gen_var' แล้วเปลี่ยน[[r3],[r4]] และ [r3,r4] เป็น[[R1],[R2]] [R1,R2] โดยพรีดิเขต 'indexll' และ 'indexl' ตามลำดับ ทั้งหมดจะถูกเก็บเป็นอาร์กูเมนต์ของฟังก์เตอร์ 'cs_fact' ไว้ในฐานข้อมูลของโปรลอก ดังนี้

```
cs_fact(f2,[R1,R2],[[R1],[R2]],[R1,R2],[mteacher_name,
          s_o_religion_c]).
```

7.2.2 ซีไอพี (CIP: Conceptual Information Processor)

เป็นส่วนของระบบที่ทำหน้าที่ติดต่อกับผู้ใช้ ซึ่งได้กล่าวไปแล้วในบทที่ 5 ซีไอพีของไนแอมเบสฟิฟจีไอเอส มีหน้าที่สำคัญ 3 ประการคือ

1. ตอบคำถามของผู้ใช้เกี่ยวกับฐานข้อมูล

2. อนุญาติให้มีการแก้ไขฐานข้อมูลเมื่อการแก้ไขนั้นยังทำให้สถานของฐานข้อมูลสอดคล้องกับคอนเซบชวลสกีมม่า

3. สร้างชุดฝึกหัด (Training Set) เพื่อนำไปอุปมานกฎโดยไอดีทรี แล้วนำกฎที่ได้มาเป็นกฎการวินิจฉัย(Inference Rules)

ข้อที่ 1 และ 2 เป็นหน้าที่ของซีไอพีที่ฟิฟจีไอเอสได้กำหนดไว้ แต่สำหรับข้อที่ 3 เป็นหน้าที่พิเศษที่เพิ่มเข้ามา ซึ่งจะกล่าวถึงรายละเอียดในบทต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต่อไปนี้จะเป็นคำอธิบายการทำงานตามหน้าที่ทั้งสอง

1. หน้าที่ในการตอบคำถามของผู้ใช้เกี่ยวกับฐานข้อมูล เป็นความรับผิดชอบของพรีดิเขต `query' ซึ่งจะเป็นตัวที่อ่านคำถาม (Query) ซึ่งมีรูปแบบเดียวกันกับ ความจริงของโปรลอก จากผู้ใช้คือมีฟังก์เตอร์และอาร์กูเมนต์ แต่ฟังก์เตอร์ในที่นี้จะแทนชนิดความจริงและอาร์กูเมนต์แทนเอนติตี้ที่เกี่ยวข้องกับบทบาทบนชนิดความจริงนั้น สำหรับความจริงของคอนเซบชวลสกีมม่าก็มีรูปแบบเดียวกันนี้ จึงถูกเก็บในฐานข้อมูลของโปรลอก ดังนั้นการตอบคำถามจึงเป็นเพียงการใช้โปรลอกพิสูจน์ (Prove) ความจริงนั้น แต่ก่อนที่จะทำเช่นนั้นจะต้องตรวจสอบเสียก่อนว่า ฟังก์เตอร์ของคำถามเป็นชนิดที่มีในคอนเซบชวลสกีมม่า เมื่อโปรลอกทำการพิสูจน์ให้แล้วอาร์กูเมนต์ใดที่เป็นตัวแปรในคำถาม ก็จะมีค่าตามที่ค้นหาความจริงที่มีในฐานข้อมูล คำตอบที่ได้จะแสดงออกมาในลักษณะของภาษาธรรมชาติที่มีรูปแบบซึ่งจัดการโดยพรีดิเขต 'write_in_nl' สำหรับประโยคที่ต้องการให้แสดงออกมาเป็นภาษาธรรมชาติที่มีรูปแบบนี้ ผู้ออกแบบคอนเซบชวลสกีมม่าสามารถกำหนดได้ตามชนิดความจริง หรือชนิดอ้างอิงโดยกำหนดได้ด้วยความจริงของโปรลอก ในส่วนที่ 3 ของการแทนคอนเซบชวลสกีมม่า ได้ดังนี้

```
fact_nl(<fact name>,[<words saperate by comma>]).
```

เช่น `fact_nl(rf1,['is the code of the religion names',''])`.

สมมุติระบบมีความจริง `rf1(bud, buddhism)`.

```
Query: rf1 (bud, X).
CIP: bud it the code of the religion names buddhism.
```

2. หน้าที่อนุญาติให้มีการแก้ไขฐานข้อมูล เมื่อการแก้ไขนั้นยังทำให้สถานะของฐานข้อมูลสอดคล้องกับคอนเซบชวลสกีมม่า การแก้ไขฐานข้อมูลมี 2 ประเภทคือ การเพิ่ม (Insert) ความจริงและการลบ (Delete) ความจริง

2.1 การเพิ่มความจริง กระทำได้ทีละความจริงเท่านั้น การเพิ่มใดที่ทำให้สถานะของฐานข้อมูลไม่สอดคล้องกับคอนเซบชวลสกีมม่า แล้วซีไอพีจะไม่ยอมรับการเพิ่มนั้น (ความจริงนั้นถูกปฏิเสธ(Reject)) เมื่อผู้ใช้ใส่ความจริงเข้ามา ซีไอพีมีขบวนตรวจสอบ 4 ด้วยกันขั้นตอนคือ

1. ขั้นตอนตรวจสอบรูปแบบของความจริง โดยพิจารณาดูว่าฟังก์เตอร์ของความจริงเป็นชนิดความจริงในคอนเซบชวลสกีมม่าหรือไม่ และมีจำนวนอาร์กูเมนซึ่งเป็นเอนติตี้ที่เกี่ยวข้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กับบทบาททั้งหมดบนชนิดความจริงนั้นถูกต้องหรือไม่ เมื่อตรวจสอบพบว่าถูกต้องแล้วก็จะดำเนินขั้นตอนต่อไป

2. ขั้นตอนตรวจสอบโดเมน จะทำการตรวจสอบโดเมนของทุกๆ เอนติตี้ที่เกี่ยวข้องกับบทบาทบนชนิดความจริงนั้น โดยที่ซีไอพีอนุญาตให้มีการระบุได้ว่าเอนติตี้เหล่านั้นถูกต้องตามโดเมนเสมอโดยใช้พรีฟิกซ์ (Prefix) 'valid' นำหน้าเอนติตี้นั้นในความจริงที่ใส่เข้ามา ถ้าเอนติตี้นี้ไม่อยู่ในขอบเขตของโดเมนที่สมนัยกันแล้ว โดเมนนั้นจะถูกขยายให้ครอบคลุมเอนติตี้นี้ด้วย มิฉะนั้นโดเมนนั้นจะคงเหมือนเดิม ในกรณีที่ไม่ได้ใช้พรีฟิกซ์ 'valid' แล้ว เอนติตี้นั้น จะถูกตรวจสอบว่าอยู่ในโดเมนของมันหรือไม่ ถ้าไม่ ซีไอพีจะฟ้องข้อผิดพลาดออกมาว่า เอนติตี้นี้ขัดข้องกับโดเมนใด แล้วซีไอพีจะไม่ยอมรับความจริงที่ผู้ใช้ต้องการเพิ่ม พริดิเขตที่ทำหน้าที่ดังกล่าวคือ 'ck_doms_err' ต้องการอาร์กูเมนต์เป็นเอนติตี้ทุกเอนติตี้ที่เกี่ยวข้องกับชนิดความจริง และชื่อของโดเมนที่สมนัยทั้งหมด ซึ่งรับมาจากแบบชนิดความจริงประเภทที่ 2

3. ขั้นตอนตรวจสอบคีย์ เมื่อโดเมนของแต่ละเอนติตี้ถูกต้องแล้ว ซีไอพีจะทดลองเพิ่มความจริงเข้าไปในฐานข้อมูล แล้วตรวจสอบว่าเกิดการขัดแย้งคีย์ (ยูนิกเนสคอนเสนท์) ที่อยู่บนชนิดความจริงนี้หรือไม่ เราสามารถทราบข้อมูลที่ว่า กลุ่มบทบาทไหนเป็นคีย์บนชนิดความจริงได้จากแบบแทนชนิด ความจริงประเภทที่ 2 เช่นกัน เราจะใช้ประเภทที่ 1 ก็ต่อเมื่อฟ้องข้อผิดพลาด กลุ่มตัวแปรที่แทนคีย์จากแบบแทนชนิดความจริงประเภทที่ 2 จะถูกนำไปใช้แมทชิง (Matching) เฉพาะกับกลุ่มเอนติตี้ที่เกี่ยวข้องกับกลุ่มบทบาทที่เป็นคีย์ที่ทุก ๆความจริงที่อยู่ในฐานข้อมูล และเก็บผลไว้ในลิสต์ (List) แล้วตรวจสอบดูว่า กลุ่มของเอนติตี้เหล่านั้นเป็นหนึ่งเดียว (Unique) หรือไม่ ถ้าไม่แสดงว่าเกิดการขัดแย้งคีย์เกิดขึ้น ระบบจะฟ้องข้อผิดพลาดออกมาว่า เกิดการขัดแย้งคีย์ที่คีย์ไหนบนชนิดความจริงนั้น แล้วซีไอพีจะไม่ยอมรับความจริงนั้น แต่ถ้าเป็นหนึ่งเดียวแล้วแสดงว่าสอดคล้องจะดำเนิดขั้นตอนต่อไป สำหรับพรีดิเขตที่ทำหน้าที่ตรวจสอบดังกล่าวคือ 'ck_key_ err' ต้องการอาร์กูเมนต์เป็นคีย์ที่อยู่ในรูปตัวแปรและโครงสร้าง (Structurc) ของโปรลอกที่มีฟังก์เตอร์เป็นชื่อของชนิดความจริง และมีอาร์กูเมนต์เป็นตัวแปรที่แทนบทบาททั้งหมดที่ประกอบเป็นชนิดความจริงนั้น ซึ่งอาร์กูเมนต์เหล่านี้จะสมนัยกับคีย์ที่กล่าวถึง เช่น ck_key_err([[R1],[R2]],f(R1,R2)). เป็นการตรวจสอบคีย์บนชนิดความจริง f2

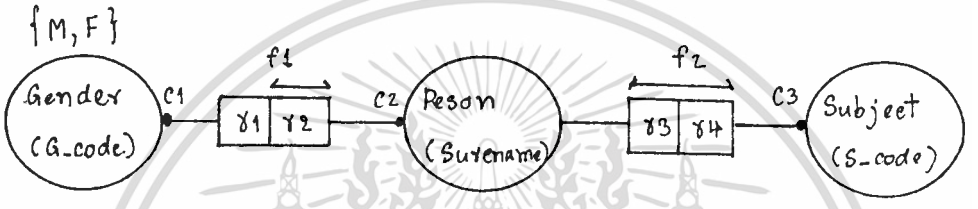
4. ขั้นตอนตรวจสอบแมนดาทอรี่ หรือตรวจสอบโททอลโรลคอนเสนท์ มีด้วยกัน 2 โหมด คือ โหมดเช็คแมนดาทอรี่ออน (ck_mand_on) และโหมดเช็คแมนอาทอรี่ออฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(ck_mand_off) ซึ่งผู้ใช้สามารถกำหนดเลือกโหมดได้ก่อนทำการเพิ่ม (Insert)

1. โหมดเช็ดแมนดาทอรี่ออน การเพิ่มในโหมดนี้ถ้าขัดแย้งกับไททอลโรลคอนสเตนท์ไหนแล้ว ซีไอพีจะไม่ยอมรับความจริงที่ใส่เข้ามา เนื่องจากเอนติตี้หนึ่งสามารถเกี่ยวกับกับบทบาทบนชนิดความจริงได้มากกกว่า 1 ชนิด ดังนั้นการเพิ่มความจริง 1 ความจริงอาจทำให้เกิดการขัดแย้งไททอลโรลคอนสเตนท์ได้หลายตัว

ตัวอย่างการเพิ่มความจริงในคอนเซบชวลสกีมม่ารูปที่ 7.5



รูปที่ 7.5 แสดงคอนเซบชวลสกีมม่าตัวอย่าง

User CIP

(ck_insert_on) insert=>f2(brown,'CS112') rejected.

CIP: Because f1(brown,X) has not recorded yet.

ซีไอพีไม่ยอมรับความจริง f2(brown,'CS112') เนื่องจาก ในฐานข้อมูลจะต้องระบุว่า brown มีเพศอะไรเสียก่อน f2(brown,CS112) ถึงจะถูกอนุญาติให้เพิ่มได้ พรีดิเขตที่ทำหน้าที่ตรวจสอบดังกล่าวคือ 'ck_insert_namd' การทำงานของมันมีขั้นตอนต่อไปนี้ (สมมุติให้มีการเพิ่มความจริง f2(brown,CS112) ตามตัวอย่างข้างบน)

1. จากความจริงที่รับมาพบว่าเป็นตัวอย่างความจริงของชนิดความจริง f2 ซึ่ง f2 สร้างมาจากกลุ่มบทบาท [r3,r4] (ทราบจากแบบแทนชนิดความจริงประเภทที่ 1)

2. หาชนิดเอนติตี้ที่เกี่ยวข้องกับ r3 และ r4 โดยใช้ข้อมูลของเมต้าคอนเซบชวลสกีมม่า จะได้กลุ่มชนิดเอนติตี้เป็น [person,subject]

3. จากกลุ่มชนิดเอนติตี้พิจารณาดูบทบาทที่เชื่อมต่อกับแต่ละชนิดเอนติตี้ทุกตัวในกลุ่มโดยใช้ข้อมูลของเมต้าคอนเซบชวลสกีมม่าจะได้กลุ่มบทบาทเป็น [r2,r3,r4]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ตัดบทบาทที่อยู่บนชนิดความจริง f2 ออกจะได้ [r2] จากนั้นตัดบทบาทที่ไม่ถูกกำกับโดยโททอลโรลคอนสเตนท์ทิ้งไป ผลก็ยังคงเป็น [r2]

5. ตอนนี้เราจะตรวจสอบว่าเอนติตี้ brown ที่เกี่ยวข้องกับ r3 ได้มีความจริงในฐานข้อมูลที่ brown เกี่ยวข้องกับ r2 หรือไม่ ถ้ามีแสดงว่าความจริง f2(brown,'CS112') ไม่ได้ขัดแย้งกับโททอลโรลคอนสเตนท์ C2 จะยินยอมให้เพิ่มเข้าไปในฐานข้อมูลได้ มิฉะนั้นจะฟ้องข้อผิดพลาดออกมาว่า f2(brown,X) ยังไม่ถูกบันทึกในฐานข้อมูล

เมื่อซีไอพีพบว่าเกิดข้อขัดแย้งกับโททอลโรลคอนสเตนท์ขึ้น ก็จะไม่ยอมเพิ่มความจริงนั้นเข้าไปในฐานข้อมูล

2. โหมดเช็คแมนดาทอรี่ออฟ โหมดนี้คล้ายกับโหมดเช็คแมนดาทอรี่ออนคือ ใช้การทำงานของพรีดิเขต 'ck_insert_mand' แต่ในโหมดนี้ เมื่อในโหมดนี้เมื่อซีไอพีพบว่าการเพิ่มความจริงเกิดข้อขัดแย้งกับโททอลโรลคอนสเตนท์ขึ้นแล้ว ก็จะยังอนุญาตให้มีการเพิ่มความจริงนั้นได้แต่จะเตือนให้ผู้ใช้เพิ่มความจริงที่ทำให้เกิดการขัดแย้งนั้นขึ้นตามมาทีหลัง เช่นจากตัวอย่างที่แล้ว ซีไอพีจะยอมรับความจริง f2 (brown,CS112) แต่จะเตือนให้ผู้ใช้เพิ่มความจริง f1 (brown,'M') ตามมาโดยอาศัยโหมดนี้จะทำให้ผู้ใช้สามารถเพิ่มความจริงเสมือนในลักษณะของแบทช์ (Batch) ได้ มิฉะนั้นแล้วการเพิ่มความจริงที่ละ 1 ความจริง และเมื่อความจริงนั้นทำให้เกิดการขัดแย้งโททอลโรลคอนสเตนท์ในโหมดที่ 1 แล้ว ความจริงนั้นจะไม่มีโอกาสถูกเพิ่มเข้าไปในฐานข้อมูลได้เลย

2. การลบ (Detete) ความจริงออกจากฐานข้อมูล ทำโดยพรีดิเขต 'delete_fact' เราจะสมมุติว่าคอนเซบชวลสกีมม่าจากตัวอย่างที่แล้ว มีปอปปูเลชั่นครบถ้วนสอดคล้องกับคอนเซบชวลสกีมม่าแล้ว หมายถึงทุกตัวอย่างของชนิดเอนติตี้ Gender,Person และ Subject ถูกบันทึกทั้งหมดและมีการกำหนดโดเมนเรียบร้อยแล้ว สมมุติให้มีความจริง f1(brown,'M') และ f2(brown,'CS112') อยู่ในฐานข้อมูล และผู้ใช้ต้องการลบ f1(brown,'M') ออกจากฐานข้อมูล พรีดิเขต'delete_fact' จะดำเนินการตามขั้นตอนต่อไปนี้

1. ตรวจสอบดูว่า f2 เป็นชนิดความจริงที่มีอยู่ในคอนเซบชวลสกีมม่าหรือไม่ ถ้ามีจะทำขั้นตอนต่อไป มิฉะนั้นจะฟ้องผิดพลาดออกมา

2. ตรวจสอบจำนวนเอนติตี้ที่เกี่ยวข้องกับบทบาทบนชนิดความจริง f2 ถ้าถูกต้องทำขั้นตอนต่อไป มิฉะนั้นจะฟ้องข้อผิดพลาดออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ตรวจสอบดูว่าความจริงนี้มีอยู่ในฐานข้อมูลหรือไม่ ถ้ามีทำขั้นตอนต่อไป มิฉะนั้นจะฟ้องข้อผิดพลาดออกมา

4. ทดสอบว่าความจริงที่ต้องการลบนี้เป็นตัวอย่างของชนิดความจริงที่มีคีย์เป็นวันเซลล์คีย์ (One Cell Key) และบทบาทที่เป็นวันเซลล์คีย์นี้ ถูกกำกับโดยโททอลโรลคอนสเตนท์ด้วยหรือไม่ ถ้าใช่ ระบบจะฟ้องออกมาว่าไม่สามารถลบตัวอย่างใด ๆ ของชนิดความจริงนี้ได้ เนื่องจากการลบตัวอย่างของชนิดความจริงที่มีบทบาทใดเป็นวันเซลล์คีย์ และบทบาทนั้นถูกกำกับโดยโททอลโรลคอนสเตนท์แล้ว จะทำให้เกิดข้อขัดแย้งกับโททอลโรลคอนสเตนท์ตัวนั้น จากตัวอย่างนี้ การลบความจริง f1(brown,'M') จะขัดแย้งกับคอนสเตนท์ C2 ดังนั้น ซีไอพีจะฟ้องว่าจะลบความจริงใด ๆ ของ f1 ไม่ได้เลย

5. ถ้าตัวอย่างที่ต้องการลบไม่เป็นดังขั้นตอนที่ 4 ซีไอพีจะดำเนินการในขั้นนี้ต่อ ตอนนี้สมมุติว่าคอนเซปชวลสกีมม่าในรูปที่ 7.5 เปลี่ยนไป คือ ให้ C2 ไปกำกับที่ r3 แทนที่จะเป็น r2 ดราวนี้ต้องการลบความจริง f2(brown,'CS112') ซึ่งจะพบว่า การลบความจริงนี้ จะถูกดำเนินขั้นตอนจาก 1 ไปถึง 4 และมาถึงขั้นตอนนี้ ขั้นตอนนี้จะทำการตรวจสอบโดยอาศัยข้อกำหนดของ C2 ที่ว่า เอนติตี้ไหนที่ปรากฏในตัวอย่างของชนิดความจริง f1 แล้วจะต้องปรากฏในตัวอย่างของชนิดความจริง f2 ด้วย จากที่สมมุติไว้ว่า มีความจริง f1(brown,'M') อยู่ในฐานข้อมูลแล้ว นั่นคือ ถ้าลบ f2(brown,'CS112') ออกจากฐานข้อมูลแล้ว จะต้องเหลือความจริงที่มีรูปแบบเป็น f2(brown,X) อยู่ในฐานข้อมูลอย่างน้อย 1 ความจริง ซีไอพีจึงจะอนุญาตให้ลบความจริง f2(brown,'CS112') ได้ มิฉะนั้นจะขัดแย้งกับ C2 ซึ่งซีไอพีจะฟ้องข้อผิดพลาดออกมาว่า ขัดแย้งกับโททอลโรลคอนสเตนท์ แล้วปฏิเสธการลบนี้

ความสามารถพิเศษของพรีดิเขต 'delete_fact' คือสามารถลบความจริงหลายๆ ความจริงได้โดยการแทนตัวแปรในตำแหน่งของเอนติตี้ของความจริงที่ใส่เข้ามา เช่น f2(x,'CS112') หมายถึงการลบ X ทุกคนที่เรียนวิธี CS112 ความสามารถพิเศษนี้เกิดจากการใช้คุณสมบัติการย้อนรอย (Backtracking) ของโปรลอกนั่นเอง

7.2.3 ฐานข้อมูล (Database)

ฐานข้อมูลของระบบไนแอมเบสฟิฟจีไอเอสนี้เป็นส่วนหนึ่งของฐานข้อมูลของโปรลอก ดังนั้น เพื่อให้ระบบมีความปลอดภัย (Security) เราจะต้องมีการควบคุมการใช้บิวด์อินพรีดิเขต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

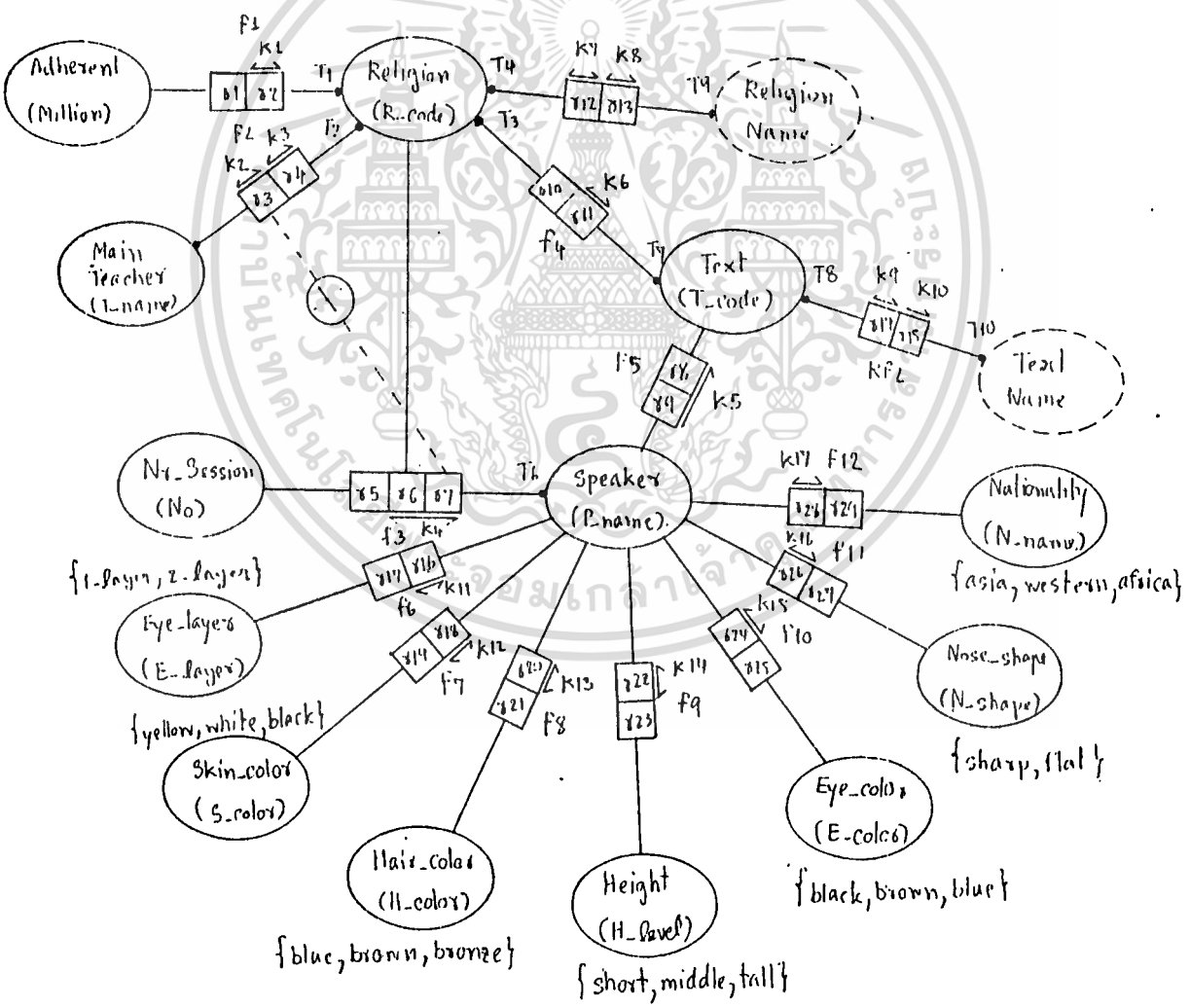
(Build-in Predicates) ของโปรลอกที่เกี่ยวกับประมวลผู้ฐานข้อมูล ได้แก่ clause, listing (สำหรับการแสดงความจริง) asserta,assertz,assert (สำหรับการเพิ่มความจริง) retract,retractall (สำหรับการลบความจริง) โดยกำหนดลำดับความสำคัญให้กับผู้ใช้แต่ละระดับ แต่อย่างไรก็ตามระบบเรายินยอมให้มีการใช้งานระบบในระดับล่าง (Low Level) ได้ โดยผ่านเปลือกของโปรลอกอินเตอร์พรีตเตอร์ (PROLOG Interpreter Shell) ซึ่งเหมาะสำหรับผู้ที่ทำหน้าที่ดูแลฐานข้อมูลเป็นผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 8

การถ่ายความรู้(Knowledge Transfer)

จากสถาปัตยกรรมของไนแอมฟิฟจีไอเอสในบทที่แล้ว นอกจากชีไอพีจะมีหน้าที่ในการหาคำตอบคำถามและการควบคุมการแก้ไขฐานข้อมูลของผู้ใช้แล้วยังมีหน้าที่พิเศษอีกประการหนึ่ง คือ การให้สิทธิ์กับผู้ใช้(ผู้เชี่ยวชาญ)ในการเลือกสร้างความจริงอนุมาน (Deduced Fact) จากตัวอย่างความจริงของชนิดความจริงอื่นๆ ก่อนที่ผู้ใช้จะทำการเลือกชนิดความจริงไหนเป็นความจริงอนุมาน ผู้ใช้จะต้องรู้ว่าชนิดความจริงไหนสามารถอนุมานได้จากชนิดความจริงอื่นอะไรบ้าง และจะต้องมีตัวอย่างความจริงของชนิดความจริงเหล่านี้เพียงพอที่บอกกล่าวได้เช่นนั้น จากคอนเซบชวลสกีมม่าต่อไปนี้



รูปที่ 8.1 แสดงคอนเซบชวลสกีมม่าที่ใช้แสดงการถ่ายความรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซีไอพีจะมีโหมดพิเศษชื่อว่าโหมดสร้างความจริงอนุมาน (Create_deduce_fact) เมื่อผู้ใช้เลือกโหมดนี้ ซีไอพีจะแสดงข้อมูลเกี่ยวกับชนิดความจริงทั้งหมดในคอนเซบชวลสกีมม่าให้ผู้ใช้เลือกชนิดความจริงที่ต้องการสร้างเป็นความจริงอนุมาน และชนิดความจริงอื่นที่จะใช้บอกกล่าวในการอนุมานให้ได้ความจริงนั้น จากตัวอย่างนี้ผู้ใช้เลือก f12 เป็นชนิดความจริงที่ใช้สร้างความจริงอนุมานและเลือก f6,f7,f8,f9,f10 และ f11 เป็นชนิดความจริงที่ใช้บอกกล่าวในการอนุมานให้ได้ความจริงอนุมานของชนิดความจริง f12 จากนั้นซีไอพีจะรอรับการให้ตัวอย่างของความจริงที่ได้กล่าวไปตอนต้นจากผู้ใช้ในลักษณะของโอเอ็นเอฟ (ONF: Optimal Normal Form) ซึ่งโอเอ็นเอฟนี้สร้างโดยซีไอพีจากชนิดความจริง f6,f7,f8,f9,f10,f11 และ f12 จากตัวอย่างผู้ใช้ก็จะใส่ตัวอย่างของความจริงชุดได้ดังนี้

```
ONF:P_name  E_layer  S_color  H_color  H_level  E_color  N_shape  N_name
 =>   'A'   1_layer  yellow   black    middle   black    sharp    asian
```

สำหรับชุดตัวอย่างอื่นๆ ผู้ใช้ก็ใส่ได้ในลักษณะนี้เช่นเดียวกัน

เมื่อซีไอพีรับชุดความจริงในลักษณะโอเอ็นเอฟแล้วมันจะทำหน้าในการเพิ่มความจริงย่อยแต่ละความจริงจากความจริงในลักษณะของโอเอ็นเอฟ เข้าไปในฐานข้อมูลโดยจะต้องสอดคล้องกับคอนเซบชวลสกีมม่า ซึ่งมีลักษณะเช่นเดียวกับที่กล่าวไปในบทที่แล้ว ถ้าความจริงไหนเพิ่มแล้ว เกิดข้อแย้งขึ้นซีไอพีจะปฏิเสธทั้งชุดความจริงโอเอ็นเอฟนั้นพร้อมกับฟ้องข้อผิดพลาด ผู้ใช้จะต้แกไข้ใหถกตูองในการปอนชดตุวอยาง่ใหม ถ่าบ้รากฏวาส่ามารถเพมคีวามจรงโหลาน่ได้ทงชั้ด ซีไอพีจะเก็บชุดความจริงในลักษณะของโอเอ็นเอฟนี้เป็น 1 บรรทัดของชุดฝึกหัด (Training) ชุดความจริงจะถูกผู้ใช้ใส่เข้าไปเรื่อย ๆ จนกระทั่งผู้ใช้พอใจ

ตอนนี้ซีไอพีก็ได้เก็บทุกบรรทัดของชุดฝึกหัดไว้แล้ว ซึ่งจะเป็นส่วนตัวของตารางชุดฝึกหัด (Training Set Table) สำหรับส่วนหัวของตารางชุดฝึกหักซีไอพีสามารถสร้างได้จากชนิดความจริงที่ถูกเลือกเป็นความจริงอนุมาน (f12) และชนิดความจริงที่ใช้บอกกล่าวการอนุมาน (f6,f7,f8,f9,f10,f11) โดยพิจารณาจากชนิดเอนติตี้ที่เกี่ยวข้องกับบทบาทบนชนิดความจริงเหล่านี้ ซึ่งถ้าจะเปรียบเทียบกับแนวคิดของไอดีทรี (ID3) แล้วชนิดเอนติตี้ที่ร่วมของทุกชนิดความจริง (Speaker) จะหมายถึง สิ่งของ (Object) ที่จะถูกแบ่งประเภท บางตัวอย่างเอนติตี้ของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชนิดเอนติตี้ซึ่งไม่ใช่ชนิดเอนติตี้ร่วมที่อยู่บนชนิดความจริงอนุมาน คือ ประเภทที่เป็นไปได้ของสิ่งของ และสำหรับชนิดเอนติตี้อื่นๆ ที่เกี่ยวข้องกับชนิดความจริงที่ใช้บอกกล่าวการอนุมาน (f6,f7, f8,f9,f10,f11) คือ แอทริบิวท์ที่ใช้อธิบายสิ่งของตามตัวอย่างนี้ คือ Eye_layer, Skin_color,Hair_color,Height,Eye_color และ Nose_shape สำหรับตัวอย่างเอนติตี้ของชนิดเอนติตี้เหล่านี้ก็คือ ค่าของแอทริบิวท์นั้นๆ

ซีไอพีจะใช้ความสัมพันธ์ดังกล่าวนี้ในการสร้างหัวของตารางชุดฝึกหัด จากตัวอย่างนี้ได้หัวตารางชุดฝึกหัดเป็น

SPEAKER EYE_LAYER SKIN HAIR HEIGHT EYE_COLOR NOSE_SHAPE NATIONALITY

เมื่อซีไอพีได้ส่วนหัวของตารางและส่วนตัวของตารางแล้ว เพื่อให้ตัวแปลงกฎจากรูปต้นไม้สำหรับการตัดสินใจไปเป็นรูปโปรลอกคลอสได้ทราบว่าจะตั้งชื่อพรีดิเขตของคลอสว่าอย่างไร ซีไอพีจะทำการบันทึกความสัมพันธ์ของเอนติตี้ต่าง ๆ กับชนิดความจริง f6,f7,f8,f9,f10, f11 และ f12 เช่น ชนิดเอนติตี้ Speaker สัมพันธ์กับเอนติตี้ Nationality บนชนิดความจริง f12 ลงไปพร้อม ๆ กับตารางชุดฝึกหัดบนไฟล์ เพื่อส่งให้เล็กซ์เชลล์นำไปสร้างโปรลอกคลอสให้แล้วนำโปรลอกคลอสกลับเข้ามาแทนเป็นความจริงอนุมานในระบบ

ข้อจำกัดของขบวนการถ่ายความรู้

1. บทบาททั้งหมดบนชนิดความจริงถูกกำหนดลำดับแน่นอน โดยบทบาทแรกจะเป็นบทบาทที่เกี่ยวข้องกับเอนติตี้ร่วมเสมอ

2. ชนิดความจริงที่เป็นชนิดความจริงอนุมาน และ ชนิดความจริงที่เป็นเงื่อนไขของการอนุมาน จะต้องเป็นชนิดความจริงแบบไบนารี่โดยมีบทบาทแรกเป็นคีย์

3. จะมีเอนติตี้หนึ่งเป็นเอนติตี้ที่ร่วมกันของชนิดความจริงอนุมาน และ ชนิดความจริงที่เป็นเงื่อนไขของการอนุมาน

4. โดเมนของทุกเอนติตี้ที่เกี่ยวข้องกับชนิดความจริงเหล่านี้ ต้องเป็นประเภท อะตอมเท่านั้น

สังเกตได้ดังตัวอย่างคอนเซบชวลสกีมม่าดังรูปที่ 8.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 9

ผลการทดลอง

9.1 ขั้นตอนการพัฒนาโครงงาน

ลักษณะโครงงานเปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่างแบ่งออกเป็น 2 ส่วน คือ ส่วนการพัฒนาต้นแบบระบบผู้เชี่ยวชาญที่สามารถสรุปจากตัวอย่าง (LEX-SHELL) และส่วนการพัฒนาต้นแบบระบบสารสนเทศยุคที่ห้าไนแอม (NIAM-based 5GIS) ดังแสดงรายละเอียดของขั้นตอนในการพัฒนาอย่างคร่าวๆ ได้ดังนี้

ขั้นตอนที่ 1 ขั้นตอนของการพัฒนาเลกซ์เชลล์ แบ่งออกเป็น

- ศึกษาทฤษฎีที่เกี่ยวข้องกับระบบผู้เชี่ยวชาญ ทฤษฎีไอดีทรี ทฤษฎีข่าวสาร
- ศึกษาข้อจำกัดและข้อบกพร่องของไอดีทรี ค้นคิดหาวิธีแก้ไขข้อบกพร่องเหล่านั้น
- ศึกษาวิธีใช้งานซอฟต์แวร์ PC SCHEME
- ออกแบบและจัดสร้างระบบเล็กซ์เชลล์
- ทดสอบระบบ

ขั้นตอนที่ 2 ขั้นตอนของการพัฒนาไนแอมเบสฟิฟจีไอเอส

- ศึกษาทฤษฎีเกี่ยวกับไนแอมและเปลือกระบบผู้เชี่ยวชาญที่ใช้แนวคิดของฟิฟจีไอเอส
- ศึกษาภาษาโปรลอก
- ออกแบบและสร้างส่วนประกอบที่สำคัญทีละส่วนตั้งแต่ เมต้าคอนเซบชวลสกีมม่า ตัวตรวจสอบคอนเซบชวลสกีมม่า การแทนแบบแทนชนิดความจริง และ ซีไอพี
- ทดสอบระบบ

ขั้นตอนที่ 3 ขั้นตอนการพัฒนาการถ่ายทอดความรู้จากเลกซ์เชลล์ไปสู่ไนแอมเบสฟิฟจีไนเอส และการทดสอบผลดังกล่าว

ขั้นตอนที่ 4 จัดทำเอกสารของแต่ละระบบและเอกสารการใช้งานร่วมกันของทั้งสองระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9.2 ผลการทดลอง

ระบบเลกซ์เซลล์ที่ได้พัฒนาแล้วได้รวมเอาวิธีการปรับปรุงไอดีทรีเข้าไปด้วยคือ วิธีการตัดค่าของแอทริบิวท์บางค่าออก และการให้ผู้ใช้เลือกแอทริบิวท์เพื่อมาเป็นโหนดของต้นไม้สำหรับการตัดสินใจ แต่ไม่ได้รวมเอาวิธีการปรับปรุงไอดีทรีโดยวิธีการย้ายต้นไม้ย่อย เนื่องจากวิธีการนี้เหมาะกับต้นไม้บางรูปแบบตามที่ได้กล่าวถึงในวิธีการปรับปรุงไอดีทรีวิธีแรกเท่านั้น

ลักษณะการทำงานของเลกซ์เซลล์แบ่งเป็น 2 ส่วน ดังนี้

1. ส่วนจัดหาความรู้ (Acquisition Module) ประกอบ 3 ขั้นตอน คือ

ขั้นตอนที่ 1 ขั้นตอนการรับอินพุทชุดฝึกหัด ผู้ใช้สามารถให้ชุดฝึกหัดโดยใช้อีดิตเตอร์ (Editor)ที่ระบบจัดไว้ให้ หรือ โดยทางไฟล์ โดยมีเงื่อนไขว่า ชุดฝึกหัดที่ให้จะต้องเป็นไปตามรูปแบบของตารางชุดฝึกหัด

ขั้นตอนที่ 2 ขั้นตอนการอนุมานกฎจากชุดฝึกหัด ขั้นตอนนี้ระบบจะใช้วิธีการไอดีทรีที่ปรับปรุงแล้วในการสร้างต้นไม้สำหรับการตัดสินใจจากชุดฝึกหัด ในขณะที่ระบบสร้างต้นไม้อยู่นั้นอาจจะมีคำถามให้ผู้ใช้เลือกแอทริบิวท์เพื่อสร้างโหนดของต้นไม้ เมื่อสิ้นสุดขั้นตอนนี้ ระบบจะได้ต้นไม้สำหรับการตัดสินใจที่ได้จากการอุปมานอยู่ในฐานความรู้

ขั้นตอนที่ 3 ขั้นตอนการใช้ประโยชน์จากต้นไม้สำหรับการตัดสินใจจากฐานความรู้ ในขั้นตอนนี้ผู้ใช้สามารถเลือกให้ระบบทำการวินิจฉัยต้นไม้จากการอุปมานนี้ได้ โดยเข้าสู่ส่วนจัดการให้คำปรึกษา หรือผู้ใช้อาจต้องการกฎที่อุปมานได้ เพื่อนำไปใช้กับเปลือกระบบผู้เชี่ยวชาญอื่น ๆ ทำได้โดยเลือกว่าจะส่งกฎออกมาทางเอาท์พุทไหน ซึ่งมี 3 ทางด้วยกัน คือ ทางจอภาพ ทางไฟล์ และทางพริ้นเตอร์ หลังจากนั้นผู้ใช้ก็เลือกรูปแบบของกฎที่ต้องการ ซึ่งระบบมีให้เลือก 3 รูปแบบคือ รูปแบบสับเซตของภาษาธรรมชาติ รูปแบบของต้นไม้สำหรับการตัดสินใจ และรูปแบบของโปรลอกคลอส (PROLOG Clauses) หลังจากที่ผู้ใช้พิจารณาการวินิจฉัยต้นไม้แล้ว อาจจะทำการรับกฎต่อไป หรือ เมื่อผู้ใช้ได้รับกฎแล้วก็อาจจะต้องการให้ระบบวินิจฉัยกฎนั้นก็ได้นอกจากนี้ผู้ใช้ยังสามารถเลือกให้ระบบทำการอุปมานกฎจากชุดฝึกหัดใหม่ได้ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ส่วนจัดการให้คำปรึกษา (Consulting Session Interface) มีขั้นตอนการทำงาน 2 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ขั้นตอนการรับต้นไม้สำหรับการตัดสินใจจากไฟล์

ระบบจะทำการถามชื่อไฟล์ที่เก็บต้นไม้สำหรับการตัดสินใจที่ผู้ใช้ต้องการจะให้ระบบใช้ในการให้คำปรึกษากับผู้ใช้ ซึ่งต้นไม้สำหรับการตัดสินใจนี้ได้จากส่วนจัดหาความรู้เมื่อผู้ใช้เลือกที่จะรับกฎจากการอุปมานในรูปแบบของต้นไม้สำหรับการตัดสินใจทางไฟล์ ในกรณีที่ผู้ใช้เลือกใช้ส่วนจัดการให้คำปรึกษาหลังจากที่ได้กฎจากการอุปมาน ระบบจะผ่านขั้นตอนนี้ไป โดยส่วนจัดการให้คำปรึกษาจะใช้ต้นไม้สำหรับการตัดสินใจ ที่อยู่ในฐานความรู้ขณะนั้น

ขั้นตอนที่ 2 ขั้นตอนทำการวินิจฉัยต้นไม้สำหรับการตัดสินใจ

จะมีการทำงานร่วมกันของเครื่องจักรอินเฟอเร้นซ์ และส่วนจัดการให้คำปรึกษาเพื่อวินิจฉัย ต้นไม้สำหรับการตัดสินใจจากขั้นตอนที่ 1 ซึ่งได้ถูกเก็บไว้ในฐานความรู้ ระบบจะให้คำปรึกษาโดยมีการถามคำถามที่เกี่ยวข้องกับการววินิจฉัยจากผู้ใช้ และมีการให้คำอธิบายเกี่ยวกับข้อสรุปที่ระบบวินิจฉัยได้ และอาจจะมีภาพประกอบการบรรยายด้วย

ลักษณะการทำงานของไนแอมเบสฟิฟจีไอเอส

ประกอบด้วยลำดับการทำงาน 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ขั้นตอนการรับคอนเซบชวลสกีมม่าอินพุท

เมื่อเข้าสู่ระบบ ระบบจะทำการถามผู้ใช้เพื่อรับคอนเซบชวลสกีมม่าจากไฟล์มันในระบบ ผู้ออกแบบจะสามารถสร้างคอนเซบชวลสกีมม่าได้โดยอาศัยโปรแกรม เทกซ์อีดิตเตอร์ทั่วไป

ขั้นตอนที่ 2 ขั้นตอนกลั่นกรองการออกแบบคอนเซบชวลสกีมม่า

ตัวตรวจสอบคอนเซบชวลสกีมม่า จะทำหน้าที่ตรวจสอบกฎไวยากรณ์ไนแอมบนคอนเซบชวลสกีมม่าที่มีอยู่ในระบบ และฟ้องข้อผิดพลาดให้ผู้ออกแบบได้ทราบถ้าคอนเซบชวลสกีมม่าไม่เป็นไปตามกฎดังกล่าว ซึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กฎนี้ถูกกำหนดโดยเมต้าคอนเซบชวบสกีมม่าที่มีในระบบตั้งแต่เริ่มแรก ผู้ออกแบบจะต้องแก้ไขคอนเซบชวลสกีมม่าให้ถูกต้องเสียก่อน ระบบถึงจะทำตามขั้นตอนต่อไปได้

ขั้นตอนที่ 3 ขั้นตอนการทำงานของซีไอพี

ประกอบด้วยการทำงาน 10 ประเภทดังนี้

1. การรับคอนเซบชวลสกีมม่าตัวใหม่เข้าสู่ระบบแทนตัวเก่า ผู้ใช้สามารถเปลี่ยนยูไอดี ที่ระบบจะใช้อ้างอิงถึงได้จาก ซึ่งคอนเซบชวลสกีมม่าตัวใหม่นี้จะต้องผ่านขั้นตอนที่ 2 ก่อนหน้านี้แล้วเข้าสู่ขั้นตอนที่ 3 อีก

2. การเพิ่มความจริงในฐานข้อมูล ซีพีไอจะควบคุมการเพิ่มความจริงของผู้ใช้ในฐานข้อมูล ให้สอดคล้องกับคอนเซบชวลสกีมม่าในระบบ ถ้าเกิดขัดแย้งขึ้น ซีพีไอจะปฏิเสธการเพิ่มนั้น

3. การถามคำถามเกี่ยวกับความจริงในฐานข้อมูล ซีพีไอจะทำหน้าที่ตอบคำถาม โดยการค้นหาความจริงที่เป็นคำตอบในฐานข้อมูล

4. การลบความจริงในฐานข้อมูล ซีพีไอจะควบคุมการลบความจริงของผู้ใช้ในฐานข้อมูล ให้สอดคล้องกับคอนเซบชวลสกีมม่าในระบบ ถ้าเกิดขัดแย้งขึ้น ซีพีไอจะปฏิเสธการลบนั้น

5. การสร้างความจริงอนุมาน เป็นการให้สิทธิ์กับผู้ใช้สามารถเลือกสร้างความจริงอนุมาน จากชนิดความใดชนิดชนิดหนึ่ง ผู้ใช้จะต้องเป็นผู้ให้ตัวอย่างของความจริงที่คิดว่าเพียงพอที่ระบบจะสามารถเปลี่ยนให้อยู่ในรูปของชุดตัวอย่าง เพื่อให้เล็กซ์เชลล์ทำการอุปมานกฎอยู่ในรูปโปรลอกคลอส แล้วนำกฎนั้นมาเป็นความจริงอนุมานในระบบต่อไป

6. การแสดงชนิดความจริงทั้งหมดที่มีในคอนเซบชวลสกีมม่า ซีไอพีจะแสดงชนิดความจริงรวมทั้งข้อมูลเกี่ยวกับชนิดความจริงนั้น ได้แก่ บทบาททั้งหมดที่ประกอบกันเป็นชนิดความจริงนั้น กลุ่มบทบาทที่แทนคีย์บนชนิดความจริงนั้น กลุ่มบทบาทที่ถูกกำกับโดยโททอลโรลคอนสเตนท์ และชื่อโดเมนของเอนติตี้ที่เกี่ยวข้องกับบทบาทที่อยู่บนชนิดความจริงนี้ทั้งหมด

7. การแสดงตัวอย่างความจริงทั้งหมดในฐานข้อมูล แสดงตัวอย่างความจริงทั้งหมดในฐานข้อมูล

8. การแสดงกฎของการอนุมานความจริงอนุมาน แสดงกฎที่ได้จากการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เลือกการทำงานในข้อที่ 5

9. การเรียกใช้โปรลอกอินเตอร์พรีตเตอร์ เพื่อให้ผู้ใช้มีโอกาสในการใช้งานปรับปรุง และพัฒนาระบบในระดับล่าง ซึ่งเป็นโปรลอกอินเตอร์พรีตเตอร์

10. ออกจากระบบกลับสู่โอเอส (OS) เมื่อผู้ใช้ไม่ต้องการใช้ระบบอีกต่อไปแล้ว

9.3 ตัวอย่างการทำงานของระบบ

ตัวอย่างการทำงานของระบบจะอยู่ในคู่มือการใช้งานของโครงงาน รวมทั้งลิสติงโปรแกรมทั้งหมดของโครงงาน ซึ่งผู้สนใจสามารถติดต่อขอรายละเอียดได้จากภาควิชาวิศวกรรมคอมพิวเตอร์โดยตรง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 10

สรุปและวิจารณ์

ส่วนจัดหาความรู้ของเล็กซ์เชลล์มีความสามารถในการอุปมานกฎที่มีความถูกต้องสูงมาก โดยใช้อาศัยวิธีการของไอดีทรีที่ปรับปรุงแล้ว ในขณะที่วิธีการไอดีทรีเดิมมีข้อบกพร่องบางประการคือ

1. วิธีไอดีทรีดั้งเดิมไม่สามารถอุปมานกฎจากชุดฝึกหัดบางชุดได้ ในกรณีที่ชุดฝึกหัดนั้นถูกแบ่งเป็นชุดฝึกหัดย่อย เพื่อสร้างต้นไม้ย่อย แล้วปรากฏว่าตัวอย่างในนั้นมีค่าของแอทริบิวท์ไม่ครอบคลุมทุกค่าที่เป็นไปได้ของแอทริบิวท์นั้น

2. ในกรณีที่แอทริบิวท์ต่าง ๆ ที่ถูกเลือกมาเป็นกฎ มีโอกาสที่จะถูกเลือกมาเป็นโหนดเท่า ๆ กัน ไอดีทรีจะเลือกโหนดที่อาจจะสื่อความหมายในการวินิจฉัยน้อยกว่าโหนดอื่น ๆ ที่มีโอกาสถูกเลือกเท่ากันนั้น

ผู้จัดทำได้ปรับปรุงวิธีการไอดีทรีให้สามารถแก้ไขปัญหาเหล่านี้ได้ ถึงอย่างไรก็ตามวิธีการนี้ยังไม่สามารถอุปมานกฎ เมื่อค่าของแอทริบิวท์เป็นตัวเลขและกฎที่มีค่าความน่าจะเป็นเข้ามาเกี่ยวข้อง ซึ่งจะต้องมีการค้นคว้าทางทฤษฎีประกอบกับความเป็นไปได้ในทางปฏิบัติ เพื่อเอาชนะปัญหานี้ต่อไป

ขบวนการในการถ่ายทอดความรู้จากส่วนจัดหาความรู้ของเล็กซ์เชลล์ โดยให้ตัวแปลงกฎจากต้นไม้สำหรับการตัดสินใจไปเป็นโปรลอกคลอสมองคอนเซบชวลสกีมม่าร่วมกับไนแอมเบสฟิฟจีไอเอส เป็นวิธีการหนึ่งที่ขยายสถาปัตยกรรมของฟิฟจีไอเอส โดยให้ซีไอพีเตรียมชุดฝึกหัดเพื่อนำไปอุปมานกฎ รวมทั้งการรับกฎที่ได้จากการอุปมานแทนในคอนเซบชวลสกีมม่าเพื่อแทนความจริงอุปมานในระบบต่อไป

ถึงอย่างไรก็ตาม ชนิดความจริงที่ผู้ใช้จะเลือกสร้างความจริงอุปมานและชนิดความจริงที่เกี่ยวข้องนั้น มีรูปแบบบางประการที่ผู้ใช้จะต้องรู้ มิฉะนั้นแล้ว ซีไอพีจะไม่สามารถเตรียมชุดฝึกหัดเพื่อนำไปอุปมานกฎต่อไปได้ แนวทางในการพัฒนาการถ่ายทอดความรู้นี้ จะต้องหาวิธีการที่ทำให้ไม่จำกัดรูปแบบตามที่ได้กล่าวไปแล้วในบทที่ 8 เรื่องการถ่ายความรู้

สำหรับการทำงานอื่น ๆ ของซีไอพี ได้แก่

- การเพิ่มความจริงในฐานข้อมูล ควรที่จะเพิ่มความสามารถในการเพิ่มความจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในฐานข้อมูลในลักษณะเป็นชุดความจริง

- เพิ่มความสามารถในการติดต่อกับ ซีไอพี โดยอาศัยสับเซตของภาษาธรรมชาติ

ประการสุดท้ายที่สำคัญ สำหรับเมต้าคอนเซบชวลสกีมม่า ควรที่จะขยายเมต้าคอนเซบชวลสกีมม่าให้ครอบคลุมกฎไวยากรณ์ของไนแอม และ พัฒนาตัวกรองการออกแบบให้สามารถตรวจสอบหาข้อผิดพลาดต่าง ๆของการแทนคอนเซบชวลสกีมม่าด้วยไนแอมให้สมบูรณ์ยิ่งขึ้น ตลอดจนการแทนชนิดความจริงจะต้องพยายามหลีกเลี่ยงการค้นหาข้อมูลในเมต้าคอนเซบชวลสกีมม่าในกรณีที่ไม่จำเป็นให้มากที่สุด เพื่อจะทำให้ระบบทำงานได้เร็วขึ้น

กล่าวโดยสรุปก็คือ โครงงานนี้เป็นความพยายามที่จะเพิ่มความสามารถในการเรียนรู้ให้กับเปลือกระบบผู้เชี่ยวชาญ แสดงถึงวิธีการถ่ายความรู้จากเปลือกระบบผู้เชี่ยวชาญที่สามารถสรุปกฎจากตัวอย่างไปยังเปลือกระบบผู้เชี่ยวชาญอื่นโดยที่ทั้งสองระบบมีความรู้ร่วมกันอยู่ และเสนอแนวทางอันหนึ่งในการเชื่อมวิทยาการด้านการออกแบบและการจัดการฐานข้อมูลเข้ากับปัญญาประดิษฐ์ โดยวิธีการแทนความรู้โดยใช้แบบจำลองข้อมูลไนแอมและขบวนการอุปมานกฎโดยวิธีการไอดีทรี

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ผู้จัดทำขอขอบคุณ ดร. ศุภมิตร จิตตะยโศธร ด้วยความเคารพและนับถือเป็นอย่างยิ่ง ที่ท่านเป็นผู้คอยชี้แนะแนวทาง ให้กำลังใจ และ ติดตามโครงงานอย่างใกล้ชิดมาโดยตลอด ขอขอบคุณ ศจ.ดร. ไพรัช ธัชยพงษ์ ที่คอยให้คำปรึกษาโครงงาน ขอขอบคุณเพื่อน ๆ ทุกคนที่คอยให้กำลังใจอยู่เบื้องหลัง ตลอดจนให้ความช่วยเหลือ ทำให้โครงงานและวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี อนึ่ง อุปกรณ์ที่ใช้ในการทดลองได้รับความเอื้อเฟื้อจากห้องวิจัยปัญญาประดิษฐ์ สจล. สุดท้ายขอขอบคุณ อาจารย์ และ เจ้าหน้าที่ ภาควิชาวิศวกรรมคอมพิวเตอร์มา ณ โอกาสนี้ด้วย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] ครรชิต ไมตรี, "ทฤษฎีข่าวสาร", คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,2522

[2] Quinlan, J.R. , "Discovering rules by induction from large collections of examples" ,Expert Systems in the Micro Electronic Age,(D. Michie, Ed), Edinburgh University Press, 1979.

[3] Quinlan, J.R. , "Learning efficient classification procedures and their application to chess endgames" , Machine Learnign : an AI Approach,(R.S. Michalski, J.G. Carbonell and T.M. Mitchell, Eds), Palo Alto: Tioga Publishing Company, 1983.

[4] Quinlan, J.R. , "Induction of Decision Tree" , Technical report 85.6 School of Computing Science, New South Wales Institute of Technology, Australia, 1985.

[5] Hunt, E.B. , "Concept Learning: an Information Processing Problem" , New York: Wiley, 1962.

[6] Hunt, E.B., Marin, J. and Stone, P.J. ,"Experiments in Induction", New York: Academic Press, 1966.

[7] Feigenbaum, E.A. , "Expert systems in the 1980s" , State of the Art Report on Machine Intelligence , (A. Bond, Ed.), Maidenhead: Pergamon-Infotech, 1981.

[8] Sammut, C.A. , "Concept development for expert system knowledge bases" , Australian Computer Journal 17, 1, 1985.

[9] Date, C.J. , "An Introduction to Database System Vol. 1, 4 ed.", Massachusetts: Addison-Wesley, 1986.

[10] Halpin, T. , "CS112: Introduction to Information System" , Department of Computer Science, University of Queensland, 1986.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[11] Clocksin, W.F. and Mellish, C.S. , "Programming in PROLOG" , Berlin: Springer-Verlage, 1984.

[12] Shapiro, A.D. , "Structured Induction in Expert Systems, Massachusetts: Addison-Wesley, 1987.

[13] Bratko, I. , "PROLOG Programming for Artificial Intelligence" , Massachusetts: Addison-Wesley, 1986.

[14] Shapiro, E. and Sterling, L. , "The Art of PROLOG: Advanced Programming Techniques" , MIT Press, 1986.

[15] Steele, Guy L. Jr. , "COMMON LISP Reference Manual" , Digital Press , Bedford , Massachusetts , 1984.

[16] Abelson, H. and Sussman, G.J. , "Structure and Interpretation of Computer Programs" , Massachusetts: MIT Press, 1984.

[17] Texas Instruments, Inc. , "TI Scheme Language Reference Manual version 3.0" , 1987.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้