

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ระบบที่ปรึกษาเพื่อแก้ไขปัญหาระบบการอีทีแอล
โดยใช้วิธีการให้เหตุผลเชิงกรณี กรณีศึกษาแผนงานคลังข้อมูลขององค์กร
ธนาคารไทยพาณิชย์

A PROBLEM CONSULTING SYSTEM FOR ETL PROCESS.
USING CASE-BASED REASONING. A CASE STUDY OF
ENTERPRISE DATA WAREHOUSE DIVISION, SIAM
COMMERCIAL BANK



๑/พ
๑๔๙๔

เลขหมู่..... 2656
เลขทะเบียน..... 131405
วัน,เดือน,ปี..... 2... ๕... ๒๕๕๗

b..... ๗๒๖๐๑๔๙๓
i.....

รายงานนี้เป็นส่วนหนึ่งของวิชาการศึกษาระดับ 2
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 2 ปีการศึกษา 2555

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**A PROBLEM CONSULTING SYSTEM FOR ETL PROCESS.
USING CASE-BASED REASONING. A CASE STUDY OF
ENTERPRISE DATA WAREHOUSE DIVISION, SIAM
COMMERCIAL BANK**



**A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS OF THE COURSE
INDEPENDENT STUDY 2
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2/ 2012



COPYRIGHT 2013

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ ระบบที่ปรึกษาเพื่อแก้ไขปัญหากระบวนการอีทีแอล โดยใช้วิธีการ
ให้เหตุผลเชิงกรณี กรณีศึกษาแผนงานคลังข้อมูลขององค์กร
ธนาคารไทยพาณิชย์

นักศึกษา นายจิรพงษ์ โกกะพันธุ์

รหัสนักศึกษา 53660532

ปริญญา วิทยาศาสตรมหาบัณฑิต

สาขาวิชา เทคโนโลยีสารสนเทศ

แขนงวิชา เทคโนโลยีระบบสารสนเทศ

ปีการศึกษา 2555

อาจารย์ที่ปรึกษา รศ.ดร.พรฤดี เนติโสภาคกุล

บทคัดย่อ

การให้เหตุผลเชิงกรณีเป็นศาสตร์หนึ่งของปัญญาประดิษฐ์ ซึ่งเป็นการแก้ปัญหาโดยใช้
ประสบการณ์จากอดีตมาประยุกต์ใช้ในการแก้ปัญหาใหม่ที่มีลักษณะคล้ายคลึงกับปัญหาเดิม

โครงการนี้ได้นำเสนอการพัฒนากระบวนการให้เหตุผลเชิงกรณีเพื่อให้คำแนะนำหรือเป็นแนวทาง
ในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ของแผนงานคลังข้อมูลขององค์กร ธนาคาร
ไทยพาณิชย์ ทำให้ลดเวลาในการหาสาเหตุและแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด และยัง
เป็นการสร้างฐานความรู้ให้กับองค์กรอีกด้วย โดยนำปัญหาและแนวทางแก้ไขของกระบวนการอีที
แอลที่ผิดพลาด จากกรณีที่เกิดขึ้นมาประยุกต์ในการให้คำแนะนำตามวิธีการให้เหตุผลเชิงกรณีที่เป็น
การใช้ประสบการณ์ในอดีตมาแก้ไขปัญหาใหม่ที่มีลักษณะคล้ายคลึงกับปัญหาเดิม และระบบ
สามารถเรียนรู้กรณีของการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดที่ประสบความสำเร็จใน
การแก้ปัญหาได้ และจัดเก็บเป็นความรู้ใหม่ในคลังกรณีเพื่อใช้เป็นความรู้ต่อไป โดยโครงการนี้ได้
ใช้เครื่องมือพัฒนาระบบให้เหตุผลเชิงกรณี โพรเทจ (Protege) และ ปลั๊กอินมายซีบีอาร์ (MyCBR)
มาช่วยในการสร้างและออกแบบคลังกรณี และใช้ภาษา เจเอสพี (JSP) เชื่อมต่อกับ เอพีไอ (API)
ของมายซีบีอาร์ในการพัฒนาเว็บแอปพลิเคชัน

Title	A PROBLEM CONSULTING SYSTEM FOR ETL PROCESS. USING CASE-BASED REASONING. A CASE STUDY OF ENTERPRISE DATA WAREHOUSE DIVISION, SIAM COMMERCIAL BANK
Student	Mr. Jirapong Kokaphant
Student ID.	53660532
Degree	Master of Science
Program	Information Technology
Major	Information System Technology
Academic Year	2012
Advisor	Associate Professor Ponrudee Netisopakul, Ph.D.

ABSTRACT

Case-Based Reasoning is one of the artificial intelligent methodologies. It is a problem solving paradigm taking previously successful benefits of solution to similar problems.

This project propose developing Case-Based Reasoning system for provide problem solving advice of ETL process error in data warehouse division, Siam Commercial Bank to reduce the time to resolve the problem. Moreover, It is creating a knowledge base for the organization. The system can solve a new problem by the experience of the most similar problem is the reused to solve the new problem. In addition to system can learn new problem that successful to solve problem by storing new problem for using in the further. This project use Protégé and MyCBR tool to designing library case and web application is built by myCBR library integrate with java program (JSP).

กิตติกรรมประกาศ

โครงการพัฒนาระบบงานนี้สำเร็จได้เป็นอย่างดี ด้วยคำแนะนำและคำปรึกษาจาก รศ.ดร.พรฤดี เเนติโสภากุล ตลอดจนช่วยเหลือในการตรวจสอบและแก้ไขข้อพร่องต่างๆ ด้วยความเอาใจใส่อย่างดียิ่ง ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากอาจารย์ และขอขอบพระคุณเป็นอย่างสูง ไว้ ณ โอกาสนี้

ขอกราบพระคุณคณาจารย์ภาควิชาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ได้ถาม ให้คำแนะนำ และคอยให้กำลังใจเสมอมา

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่อง ทำให้ข้าพเจ้าสามารถทำโครงการพัฒนาระบบงานนี้สำเร็จลุล่วงด้วยดี

คุณค่าและประโยชน์อันพึงมาจากโครงการพัฒนาระบบงานนี้ ข้าพเจ้าขอมอบแด่ผู้มีพระคุณทุกท่าน

จิรพงษ์ โภกพันธ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการพัฒนาระบบงาน.....	3
1.3 ขอบเขตของโครงการ.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1 การให้เหตุผลเชิงกรณี (Case-Based Reasoning).....	5
2.2 วงจรการทำงานของระบบให้เหตุผลเชิงกรณี (Case-Base Reasoning Cycle).....	6
2.2.1 ฐานกรณี (Case-Based).....	7
2.2.2 การค้นคืนกรณี (Retrieve Case).....	8
2.2.3 การนำกรณีเก่ากลับมาใช้ใหม่ (Reuse Case).....	11
2.2.4 การปรับปรุงแนวทางการแก้ไขปัญหา (Revise Case).....	12
2.2.5 การจัดเก็บกรณี (Retain Case).....	12
2.3 โปรแกรมพัฒนาฐานกรณี โพรทิจ และ มายซีบีอาร์ (Protege and MyCBR).....	12
2.3.1 โพรทิจ (Protégé).....	12
2.3.2 แพลตฟอร์มโพรทิจสนับสนุนแบบจำลองออนโทโลยี.....	13
2.3.3 ลักษณะเด่นของแบบจำลองแบบเฟรมในโปรแกรมโพรทิจ.....	13
2.3.4 หน้าจอโปรแกรมโพรทิจของแบบจำลองแบบเฟรม.....	14
2.3.5 มายซีบีอาร์ (MyCBR).....	16
2.3.6 สถาปัตยกรรมของมายซีบีอาร์.....	17

IV

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.3.7 แบบจำลองการวัดค่าความคล้ายคลึงของมายชีปีอาร์.....	18
2.3.8 การทดสอบการทำงานของกรค้นคืน.....	22
2.3.9 การสร้างแอปพลิเคชันแบบสแตนด์อะโลน.....	23
2.3.10 ฟังก์ชันคำอธิบาย.....	24
2.4 เอ็นแกรม (Ngram).....	24
บทที่ 3 การออกแบบการให้เหตุผลเชิงกรณี.....	29
3.1 รวบรวมข้อมูล.....	29
3.2 การออกแบบคลังกรณี.....	30
3.3 การค้นคืนกรณี.....	32
บทที่ 4 การวิเคราะห์และการออกแบบระบบ.....	48
4.1 ยูสเคสไดอะแกรม.....	48
4.2 ซีแควนไดอะแกรม.....	54
4.3 คลาสไดอะแกรม.....	58
4.3 การออกแบบฐานข้อมูล.....	59
บทที่ 5 ผลการทดลองและประเมินประสิทธิผลของระบบ.....	64
บทที่ 6 บทสรุปและข้อเสนอแนะ.....	96
บรรณานุกรม.....	98
ภาคผนวก.....	99
ภาคผนวก ก. ตัวอย่างกรณีกระบวนการอีทีแอลที่ผิดพลาดในคลังกรณี.....	100

สารบัญตาราง

ตารางที่	หน้า
3.1 รายละเอียดของกระบวนการอีทีแอลที่ผิดพลาด.....	29
3.2 มาตรฐานการตั้งชื่องาน.....	34
3.3 ตารางเปรียบเทียบ 3 แกรมที่ String1 เหมือนกับ String2.....	37
3.4 ตารางเปรียบเทียบ 3 แกรมที่ String2 เหมือนกับ String1.....	38
3.5 ค่าความคล้ายคลึงประเภทของงาน.....	39
3.6 ประเภทข้อมูลแต่ละสล็อต.....	44
3.7 ค่าถ่วงน้ำหนักของแต่ละคุณลักษณะ.....	45
4.1 คำอธิบายยูสเคส Add Job Error.....	49
4.2 คำอธิบายยูสเคส Search Job Error.....	50
4.3 คำอธิบายยูสเคส Edit Job Error.....	50
4.4 คำอธิบายยูสเคส Ask For Advice Solution.....	51
4.5 คำอธิบายยูสเคส Approve Solution.....	52
4.6 คำอธิบายยูสเคส Retain New Solution.....	53
4.7 ตารางเก็บข้อมูลประเภทของผู้ใช้.....	60
4.8 ตารางเก็บข้อมูลประเภทของกระบวนการอีทีแอล.....	60
4.9 ตารางเก็บข้อมูลผู้ใช้ระบบ.....	61
4.10 ตารางเก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาด.....	61
4.11 ตารางเก็บข้อมูลรายละเอียดปัญหาของงานกระบวนการอีทีแอลที่ผิดพลาด.....	62
4.12 ตารางเก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาดที่ระบบมีอยู่ในคลังกรณี.....	63
5.1 การทดลองการวัดประสิทธิผลของการให้คำแนะนำ.....	65

สารบัญรูป

รูปที่	หน้า
2.1 แนวคิดของการให้เหตุผลเชิงกรณี	5
2.2 วงจรการทำงานของระบบให้เหตุผลเชิงกรณี.....	6
2.3 สมการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ	8
2.4 สมการวัดค่าความคล้ายคลึงในระดับกรณี	9
2.5 สมการค้นคืนกรณีที่คล้ายคลึง.....	9
2.6 ตัวอย่างการค้นคืนกรณีที่คล้ายคลึง	10
2.7 แท็บคลาส	14
2.8 แท็บฟอร์ม.....	15
2.9 แท็บอินสแตนซ์	16
2.10 สถาปัตยกรรมของมายซีปีอาร์	17
2.11 สมการวัดค่าความคล้ายคลึงในระดับกรณีแบบถ่วงน้ำหนักรวม.....	18
2.12 สมการวัดค่าความคล้ายคลึงในระดับคุณลักษณะประเภทตัวเลข.....	19
2.13 ตัวจัดการความคล้ายคลึงของแอดทริบิวต์ประเภทตัวเลข	19
2.14 ตัวจัดการความคล้ายคลึงของแอดทริบิวต์ประเภทสัญลักษณ์แบบตารางความคล้ายคลึง	20
2.15 ตัวจัดการความคล้ายคลึงของแอดทริบิวต์ประเภทสัญลักษณ์แบบการจัดกลุ่ม	20
2.16 ส่วนต่อประสานการค้นคืน	22
2.17 ตัวอย่างเว็บแอปพลิเคชันการให้เหตุผลเชิงกรณี	23
2.18 ผู้เกี่ยวข้องในสถานการณ์การอธิบาย	24
3.1 ตัวอย่างกรณี.....	30
3.2 ฐานกรณีที่สร้างโดยโพทิจและมายซีปีอาร์	31
3.3 ข้อมูลตัวอย่างฐานกรณี	32
3.4 โหมดการวัดความคล้ายคลึงของสล็อตชื่อระบบงาน	33
3.5 โหมดการวัดความคล้ายคลึงของสล็อตชื่องาน.....	34
3.6 โหมดการวัดความคล้ายคลึงของสล็อตประเภทของงาน	40
3.7 โหมดการวัดความคล้ายคลึงของสล็อตข้อความผิดพลาด	41
3.8 โหมดการวัดความคล้ายคลึงของสล็อตรหัสผิดพลาด	42
3.9 โหมดการวัดความคล้ายคลึงของสล็อตชื่องานย่อย	43

VII

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.10 คำน้ําหนักของคุณลักษณะและฟังก์ชันการวัดค่า ความคล้ายคลึงในระดับกรณีในมายซีปีอาร์.....	46
3.11 การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถามและกรณี 1 ในมายซีปีอาร์.....	46
3.12 การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถามและกรณี 2 ในมายซีปีอาร์.....	47
4.1 ยูสเคสไดอะแกรม.....	48
4.2 ซีเคเวนไดอะแกรมสำหรับการบันทึกรายการกระบวนการอีทีแอลที่ผิดพลาด.....	54
4.3 ซีเคเวนไดอะแกรมสำหรับค้นหารายการกระบวนการอีทีแอลที่ผิดพลาด.....	55
4.4 ซีเคเวนไดอะแกรมสำหรับแก้ไขกระบวนการอีทีแอลที่ผิดพลาด.....	56
4.5 ซีเคเวนไดอะแกรมสำหรับขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด.....	57
4.6 ซีเคเวนไดอะแกรมสำหรับอนุมัติและบันทึกกระบวนการอีทีแอลที่ผิดพลาดลงคลังกรณี.....	58
4.7 คลาสไดอะแกรมระบบที่ปรึกษาเพื่อแก้ไขปัญหากระบวนการอีทีแอล.....	59
4.8 แบบจำลองความสัมพันธ์ระหว่างเอนตีตี้ (ER-model).....	59

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

แผนกคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ เป็นหน่วยงานที่รับผิดชอบเกี่ยวกับงานกระบวนการคลังข้อมูล ซึ่งประกอบด้วยหน่วยงานย่อยๆ 3 หน่วยงาน ดังนี้ หน่วยงานการสร้างแบบจำลองข้อมูล(Data Modeling) หน่วยงานกระบวนการเคลื่อนย้ายข้อมูลอีทีแอล (ETL – Extract Transform Load) และ หน่วยงานปัญญาธุรกิจ (Business Intelligence)

กระบวนการเคลื่อนย้ายข้อมูลอีทีแอลหรือกระบวนการอีทีแอลของแผนกคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ เป็นกระบวนการแปลงและเคลื่อนย้ายข้อมูลจาก ข้อมูลต้นทาง(Data Source) ไปยัง ฐานข้อมูลคลังข้อมูล ประกอบด้วย 3 กระบวนการคือ 1.การโหลด(Load) 2.การแปลง(Transform) 3.การดึงออก(Extract) กระบวนการแต่ละกระบวนการเหล่านี้มีความสำคัญต่อการนำข้อมูลเข้าคลังข้อมูล อีกทั้งข้อมูลจากต้นทางของแผนกคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ นั้นมีจำนวนมากและหลายต้นทาง จึงมีการแปลงและเคลื่อนย้ายข้อมูลเข้าคลังข้อมูลตลอดเวลา รวมถึงข้อมูลในคลังข้อมูลที่เข้าไป ในแต่ละวันนั้นมีความสำคัญต่อการตัดสินใจทางธุรกิจขององค์กร ทำให้เมื่อเกิดปัญหากระบวนการอีทีแอลกระบวนการใดกระบวนการหนึ่งผิดพลาด(error) จะต้องได้รับการแก้ไขปัญหอย่างทันที

การแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดนั้น เป็นหน้าที่ของพนักงานแผนกคลังข้อมูลขององค์กรทุกคน ต้องแบ่งผลัดเวร สลับเปลี่ยนกันเพื่อแก้ไขปัญหา พนักงานที่มีประสบการณ์หรือเป็นผู้เชี่ยวชาญจะสามารถตัดสินใจในการแก้ไขปัญหาได้และทันเวลา เนื่องจากเคยพบเจอลักษณะของปัญหาแบบเดียวกันมาก่อน แต่พนักงานที่มีประสบการณ์น้อยหรือพนักงานใหม่ อาจใช้เวลาในการแก้ปัญหานาน เนื่องจากไม่รู้จะเริ่มต้นแก้ปัญหอย่างไรและอาจมองปัญหาไม่ถูกจุด อีกทั้งการบันทึกกระบวนการอีทีแอลที่ผิดพลาดที่เคยเกิดขึ้น จะถูกบันทึกเป็นเอกสารเอ็กเซล(Excel) ยากต่อการค้นหาและเข้าถึง ดังนั้นถ้ามีระบบเข้ามาช่วยในการจัดเก็บ ค้นหา การแก้ปัญหากระบวนการอีทีแอลที่ผิดพลาด ซึ่งถือเป็นความรู้อย่างหนึ่งขององค์กร จะสามารถช่วยให้พนักงานแผนกคลังข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขององค์กรทุกคน สามารถตัดสินใจในการแก้ปัญหากระบวนการอีทีแอลที่ผิดพลาดได้อย่างมีประสิทธิภาพมากขึ้น

การจัดการความรู้ เป็นการรวบรวมองค์ความรู้ที่มีอยู่ในองค์กร ซึ่งจัดกระจายตัวอยู่ในตัวบุคคลหรือเอกสาร และพัฒนาให้เป็นระบบ เพื่อให้ทุกคนในองค์กรสามารถเข้าถึงความรู้ พัฒนาตนเองให้เป็นผู้รู้ รวมทั้งปฏิบัติงานได้อย่างมีประสิทธิภาพ เทคโนโลยีหนึ่งที่ใช้ในการจัดการความรู้คือ เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ที่เกี่ยวข้องกับการออกแบบระบบคอมพิวเตอร์ให้มีความสามารถในการตัดสินใจใกล้เคียงกับมนุษย์ หรือสนับสนุนการตัดสินใจให้มีประสิทธิภาพและประสิทธิผลมากยิ่งขึ้น ซึ่งการให้เหตุผลเชิงกรณี(Case-Base Reasoning) ก็เป็นศาสตร์หนึ่งของปัญญาประดิษฐ์

การให้เหตุผลเชิงกรณีเป็นการแก้ปัญหาโดยใช้ประสบการณ์จากอดีตมาประยุกต์ใช้ในการแก้ปัญหาใหม่ที่มีลักษณะคล้ายคลึงกับปัญหาเดิม ปัญหาที่คล้ายคลึงกันจะใช้วิธีการแก้ปัญหาที่คล้ายคลึงกัน หากไม่สามารถประยุกต์ใช้กรณีที่เคยเกิดขึ้นได้มาแก้ไขได้ จะมีการดัดแปลงกรณีที่มีส่วนใกล้เคียงกับปัญหา เพื่อหาแนวทางแก้ไขปัญหานั้นๆ หลังจากที่ได้แนวทางการแก้ปัญหาแล้ว จะจัดเก็บไว้เป็นกรณีสำหรับใช้แก้ไขปัญหาที่จะเกิดขึ้นต่อไป

การนำการให้เหตุผลเชิงกรณี ไปประยุกต์ใช้ในการแก้ไขปัญหกระบวนการอีทีแอลที่ผิดพลาด จะช่วยให้พนักงานแผนกคลังข้อมูลขององค์กรทุกคนสามารถแก้ไขปัญหกระบวนการอีทีแอลที่ผิดพลาด ได้อย่างทันเวลาและมีประสิทธิภาพ เนื่องจากการแก้ไขปัญหกระบวนการอีทีแอลที่ผิดพลาดต้องอาศัยประสบการณ์และการเรียนรู้ที่เคยพบเจอ นอกจากนี้ยังเป็นสร้างฐานความรู้ในเรื่องการแก้ไขปัญหกระบวนการอีทีแอลที่ผิดพลาดให้กับองค์กรอีกด้วย

ดังนั้นในโครงการนี้ได้ทำการศึกษากระบวนการให้เหตุผลเชิงกรณีหรือระบบให้เหตุผลเชิงกรณี (Case-Based Reasoning Systems) เพื่อจัดทำระบบที่ปรึกษาเพื่อแก้ไขปัญหกระบวนการอีทีแอล โดยใช้วิธีการให้เหตุผลเชิงกรณี กรณีศึกษาแผนกคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของโครงการพัฒนาระบบงาน

1. ออกแบบโครงสร้างเชิงกรณี เพื่อระบุปัญหาและคำตอบของปัญหาในกระบวนการเคลื่อนย้ายข้อมูลอีทีแอลที่ผิดพลาดของแผนงานคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์
2. นำองค์ความรู้มาอิมพลีเมนต์ (Implement) เป็นฐานความรู้และกลไกการให้เหตุผลเพื่อสนับสนุนการตัดสินใจโดยใช้เครื่องมือ โพรเทจ (Protege) และ มายซีบีอาร์ (MyCBR)
3. พัฒนาระบบสารสนเทศเพื่อให้คำปรึกษาและสนับสนุนการตัดสินใจในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ของแผนงานคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์

1.3 ขอบเขตของโครงการ

โครงการพัฒนาระบบงานนี้ถูกพัฒนาขึ้นเพื่อให้คำปรึกษาและสนับสนุนการตัดสินใจในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ของแผนงานคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ ดังนั้นการทำงานของระบบครอบคลุมได้ดังนี้

1. สามารถจัดเก็บ แก้ไข และสอบถามข้อมูลกระบวนการอีทีแอลที่ผิดพลาดได้ เช่น ข้อมูลชื่อระบบต้นทาง วันที่ของข้อมูล ประเภทของกระบวนการ ประวัติการผิดพลาด และประวัติการแก้ปัญหา
2. สามารถค้นหากรณีที่เกี่ยวข้อง โดยใช้ข้อความผิดพลาด(error message) ประเภทของกระบวนการ จากกระบวนการอีทีแอลที่ผิดพลาดที่ไม่เคยมีในระบบ มาพิจารณากรณีที่คล้ายคลึง
3. สามารถให้คำแนะนำวิธีการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด รวมถึงกรณีที่ใกล้เคียง เพื่อนำไปใช้ในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด
4. สามารถเก็บกรณีใหม่ที่ได้นำไปใช้แล้วสามารถแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด เป็นความรู้ใหม่ของระบบได้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ระบบให้คำปรึกษาและสนับสนุนการตัดสินใจในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ของแผนงานคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ เพื่อนำไปใช้ในการวิเคราะห์ และเป็นแนวทางในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด
2. ลดเวลาในการหาสาเหตุกระบวนการอีทีแอลที่ผิดพลาด
3. ได้ฐานความรู้ สำหรับศึกษาค้นคว้าและเข้าถึงความรู้ เพื่อพัฒนาตนเองให้เป็นผู้รู้และปฏิบัติงานได้อย่างมีประสิทธิภาพ
4. ช่วยให้เกิดการเรียนรู้และแบ่งปันความรู้



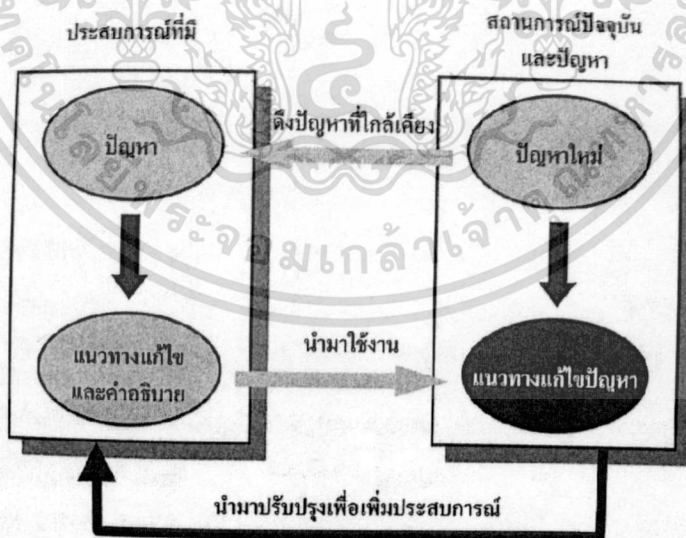
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

2.1 การให้เหตุผลเชิงกรณี (Case-Based Reasoning)

การให้เหตุผลเชิงกรณีเป็นวิธีหนึ่งที่น่าพื้นฐานทางความคิด และการเรียนรู้จากประสบการณ์ของมนุษย์มาใช้เป็นแนวทางการในการพัฒนา โดยการนำกรณีเก่าที่ถูกบันทึกลงฐานกรณีมาใช้เป็นองค์ความรู้เพื่อแก้ไขปัญหาต่างๆ ที่ระบบต้องเผชิญ และเมื่อพบกรณีใหม่ ระบบจะเรียนรู้และนำไปจัดเก็บ เพื่อเป็นแหล่งองค์ความรู้สำหรับใช้งานต่อไปในอนาคต

ระบบการให้เหตุผลเชิงกรณี (Case-Base Reasoning System) หมายถึงระบบที่มีกระบวนการแก้ไขปัญหาด้วยวิธีการสรุปความ และหาเหตุผลจากปัญหาหรือกรณีเดิมที่เคยพบในอดีต เพื่อนำมาใช้แก้ปัญหาใหม่ที่สนใจ องค์ความรู้ส่วนใหญ่จะอยู่ในรูปของกรณีที่มีความเกี่ยวข้องและสัมพันธ์กันระหว่างปัญหาใหม่กับปัญหาเก่า หรืออาจเป็นกรณีเดียวกันแต่ต่างกันที่ความทันสมัยของข้อมูล (Turban and Arunson, 2001)

แนวคิดของการให้เหตุผลเชิงกรณี คือการใช้ประสบการณ์ในอดีต มาแก้ปัญหาใหม่ที่มีลักษณะคล้ายคลึงกับปัญหาเดิม แสดงดังรูปที่ 2.1

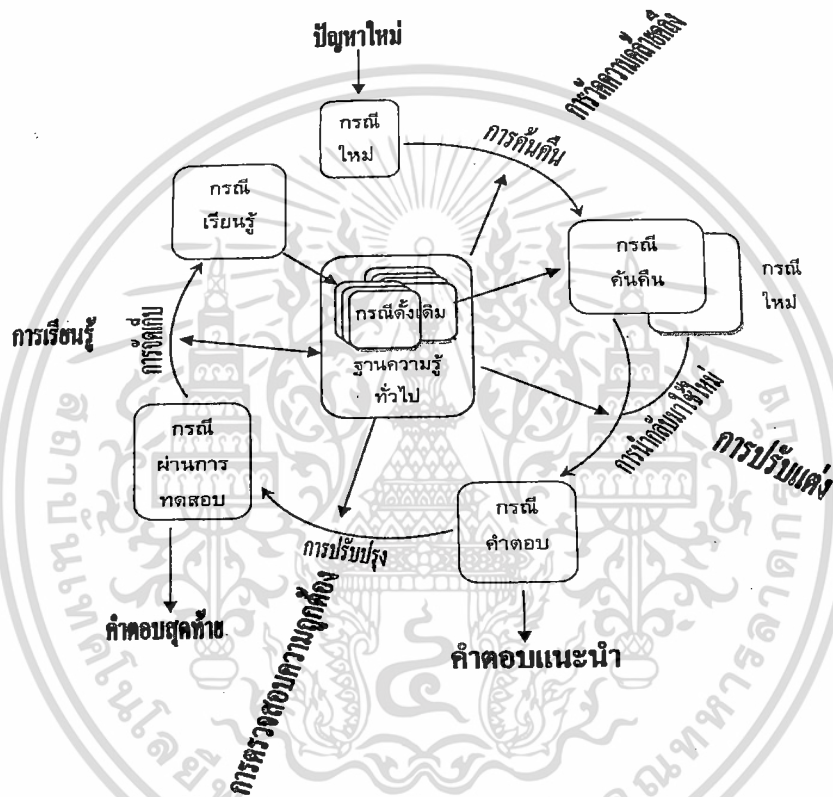


รูปที่ 2.1 แนวคิดของการให้เหตุผลเชิงกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.1 แสดงให้เห็นว่าการแก้ปัญหาต่างๆ ด้วยการให้เหตุผลเชิงกรณีมีปัจจัยสำคัญมาจากปัญหาที่เคยพบมาในอดีต ซึ่งขึ้นอยู่กับประสบการณ์ที่ผ่านมาว่าพบกับปัญหามากน้อยเพียงใด เมื่อพบปัญหาใหม่จะย้อนกลับไปพิจารณาประสบการณ์ทั้งหมด เพื่อหาปัญหาที่มีลักษณะใกล้เคียงกันและพิจารณาแนวทางการแก้ไขปัญหา ซึ่งจะถูกนำมาใช้ตามความเหมาะสม หลังจากแก้ไขปัญหาใหม่เรียบร้อยแล้วผลลัพธ์ที่ได้ยังช่วยเพิ่มพูนประสบการณ์ให้มากขึ้น

2.2 วงจรการทำงานของระบบให้เหตุผลเชิงกรณี (Case-Base Reasoning Cycle)



รูปที่ 2.2 วงจรการทำงานของระบบให้เหตุผลเชิงกรณี

จากรูป 2.2 วงจรการทำงานของระบบให้เหตุผลเชิงกรณี มี 4 ขั้นตอนสำคัญได้แก่

1. การค้นคืน (Retrieve) คือการดึงกรณีที่คล้ายคลึงกับกรณีใหม่จากแหล่งเก็บกรณีในอดีต ที่เรียกว่าคลังกรณี (Case Library)
2. การนำกลับมาใช้ใหม่ (Reuse) คือการนำวิธีการแก้ปัญหาในกรณีเก่ามาปรับใช้
3. การปรับปรุง (Revise) คืออาจมีการดัดแปลงวิธีการแก้ปัญหาให้เหมาะสมกับกรณีใหม่
4. การจัดเก็บ (Retain) คือการเพิ่มกรณีใหม่ที่ใช้งานได้ผลดี กลับเข้าไปในคลังกรณี

วงจรการทำงานของระบบให้เหตุผลเชิงกรณีเริ่มต้นจากการพบปัญหาหรือกรณีใหม่ที่ระบบสนใจ จากนั้นจึงเริ่มค้นหากรณีเก่าที่ได้จัดเก็บไว้ในฐานความรู้ที่มีความคล้ายคลึงกับกรณีเอกสารใหม่มากที่สุด (Retrieve) เมื่อพบกรณีที่มีความคล้ายคลึงกันจะนำวิธีการแก้ปัญหาของกรณีเก่ามาใช้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใช้เป็นแนวทางในการแก้ปัญหาของกรณีใหม่ (Reuse) จากนั้นจะเข้าสู่ขั้นตอนการปรับปรุง มีการตรวจสอบและแก้ไขแนวทางการแก้ปัญหาให้เหมาะสม (Revise) แล้วจึงทำการยืนยันความถูกต้องของผลลัพธ์ เมื่อมีการยืนยันความถูกต้องของแนวทางการแก้ปัญหาแล้ว จะถูกจัดเก็บลงในฐานความรู้ (Retain) เพื่อใช้ในการแก้ปัญหาใหม่อีกต่อไป

กระบวนการและองค์ประกอบของวงจรการทำงานของระบบให้เหตุผลเชิงกรณี มีดังนี้

2.2.1 ฐานกรณี (Case-Based)

ฐานกรณีเปรียบเสมือนฐานองค์ความรู้ที่จัดเก็บความรู้ไว้ แต่เป็นรูปแบบของกรณี (Case) ที่หมายถึงประสบการณ์หนึ่งครั้งที่เกิดขึ้น ณ. เวลาหนึ่ง และจะถูกเก็บแต่ละกรณีแยกกัน ไม่มีการเก็บเป็นกฎเหมือนฐานความรู้ โดยถูกรวบรวมไว้เพื่อใช้สนับสนุนการดึงองค์ความรู้ของระบบให้เหตุผลเชิงกรณี ยิ่งระบบมีกรณีมากเท่าไรยิ่งช่วยเพิ่มโอกาสในการแก้ไขกรณีใหม่ได้ง่ายขึ้น ข้อมูลภายในกรณีประกอบด้วยสองส่วนดังนี้

1. ส่วนคำอธิบายปัญหา (Problem Description) เป็นการอธิบายถึงลักษณะของปัญหา ซึ่งเป็นตัวบ่งบอกเป้าหมาย หรือวัตถุประสงค์ที่ระบบต้องดำเนินการแก้ไขให้สำเร็จ โดยปัญหาที่ระบุในกรณีจะมีรายละเอียดคุณลักษณะ (Feature) ต่างๆที่ให้เข้าใจปัญหา หรือสถานการณ์ว่ามีความต้องการสิ่งใด สิ่งเหล่านี้จะเป็นรายละเอียดคุณลักษณะที่ใช้ในการเปรียบเทียบกับส่วนอธิบายปัญหาของกรณีเก่าที่จัดเก็บไว้ เพื่อค้นหากรณีที่มีคุณลักษณะใกล้เคียงกับปัญหามากที่สุด ดังนั้นการแทนส่วนคำอธิบายปัญหาควรมีรายละเอียดคุณลักษณะเพียงพอ

2. ส่วนคำตอบ (Solution) เป็นแนวทางที่นำมาใช้ในการแก้ปัญหา แนวทางการแก้ปัญหาในกรณีเก่าที่จัดเก็บไว้จะเป็นตัวช่วยให้ระบบสามารถค้นหาแนวทางที่เหมาะสมในการแก้ปัญหาใหม่ที่เกิดขึ้นได้เป็นอย่างดี โดยรายละเอียดของแนวทางการแก้ไขปัญหามีได้หลายลักษณะ เช่น แผนการดำเนินงาน คำอธิบายปัญหา กระบวนการแก้ปัญหา และขั้นตอนการแก้ไขปัญหาโดยละเอียด เป็นต้น

ตัวอย่างของกรณี (Case) ในปัญหาการจำแนกผลไม้มัธยม หนึ่งในประสบการณ์จะแทนหนึ่งกรณี ดังนี้

ส่วนคำอธิบายปัญหา (Problem Description) :

รูปทรงของผลไม้ = รูปทรงกลม	}	Feature
สีของผลไม้ = สีส้ม		
จำนวนเมล็ด = มีหลายเมล็ด		

ส่วนของคำตอบ (Solution) :

ผลไม้ = ส้ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 การค้นคืนกรณี (Retrieve Case)

การค้นคืนกรณี (Retrieve Case) คือ การค้นหากรณีเก่าที่ถูกจัดเก็บไว้ในคลังกรณี เพื่อนำมาเปรียบเทียบกับกรณีใหม่ โดยใช้รายละเอียดคุณลักษณะ (feature) ของส่วนคำอธิบายปัญหา มาเปรียบเทียบกับ โดยกรณีเก่าจะต้องคล้ายคลึงกับกรณีใหม่มากที่สุด ซึ่งอาจมีได้มากกว่าหนึ่งกรณีก็ได้ โดยมีวิธีการนำมาใช้ในการค้นคืนกรณีนั้นมีอยู่หลายวิธีด้วยกัน ซึ่งโครงการนี้จะขอเสนอวิธีการวัดค่าความคล้ายคลึงระหว่างกรณีและการค้นคืนกรณีที่เป็นที่นิยม ดังนี้

1. การวัดค่าความคล้ายคลึง (Similarity Measurement) ระหว่างกรณี หมายถึง การมีฟังก์ชันที่แทนค่าความคล้ายคลึงของกรณีที่ค้นคืนมากับกรณีใหม่เป็นตัวเลข เพื่อตัดสินใจว่ากรณีใดมีความคล้ายคลึงกับปัญหาใหม่มากกว่ากัน

1.1 การวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) ในแต่ละกรณี จะประกอบด้วยคุณลักษณะ(feature)ของส่วนคำอธิบายปัญหา ซึ่งคุณลักษณะแต่ละตัวจะถูกนำมาใช้ในการวัดค่าความคล้ายคลึงของกรณีระหว่างปัญหาใหม่กับกรณีในคลังกรณีเพื่อค้นคืนกรณี ซึ่งเป็นการคำนวณค่า $sim(a, b_i)$ จากสมการด้านล่าง

$$Numeric \quad sim(a, b) = 1 - \frac{|a - b|}{range} \quad (2.1)$$

$$Symbolic. \text{ mono-valued} \quad sim(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \quad (2.2)$$

$$Symbolic. \text{ Multi-valued} \quad sim(a, b) = \frac{card(a) \cap card(b)}{card(a \cup b)} \quad (2.3)$$

$$sim(a, b) = \frac{card(a \cup b) - card(a \cap b)}{card(a \cup b)} \quad (2.4)$$

$$sim(a, b) = \frac{card(a \cup b) - card(a \cap b)}{card(a \cup b)} \quad (2.5)$$

$$Taxonomy \quad sim(a, b) = \frac{max(a \cup b) \cdot h(commonnode(a, b))}{\min(h(a), h(b))} \quad (2.6)$$

$$sim(a, b) = \frac{h(commonnode(a, b))}{h} \quad (2.7)$$

รูปที่ 2.3 สมการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ

เมื่อ a คือ เซตของคุณลักษณะที่ปรากฏในกรณี a

b คือ เซตของคุณลักษณะที่ปรากฏในกรณี b

range คือ ช่วงระหว่างค่าที่มากที่สุดและค่าที่น้อยที่สุดหรือพิสัยของข้อมูล

card คือ ค่าคาร์ดินัลลิตีหรือขนาดของเซต

h คือ ความสูงของ Taxonomy Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 การวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity) จะใช้ค่า $sim(a_i, b_i)$ จากการวัดค่าความคล้ายคลึงในระดับคุณลักษณะมาประเมินความคล้ายคลึงในระดับกรณี โดยสามารถเลือกใช้สมการตามอัลกอริทึมที่เหมาะสม ได้ดังต่อไปนี้

$$\text{Block - City} \quad sim(A, B) = \frac{1}{p} \sum_{i=1}^p sim_i(a_i, b_i) \quad (2.8)$$

$$\text{Wiegthed Block-City} \quad sim(A, B) = \sum_{i=1}^p \omega_i sim_i(a_i, b_i) \quad (2.9)$$

$$\text{Euclidean} \quad sim(A, B) = \left[\frac{1}{p} \sum_{i=1}^p [sim_i(a_i, b_i)]^2 \right]^{1/2} \quad (2.10)$$

$$\text{Minkowski} \quad sim(A, B) = \left[\frac{1}{p} \sum_{i=1}^p [sim_i(a_i, b_i)]^r \right]^{1/r} \quad (2.11)$$

$$\text{Weighted Minkowski} \quad sim(A, B) = \left[\frac{1}{p} \sum_{i=1}^p \omega_i [sim_i(a_i, b_i)]^r \right]^{1/r} \quad (2.12)$$

$$\text{Maximin biased} \quad sim(A, B) = \max_i \omega_i sim_i(a_i, b_i) \quad (2.13)$$

รูปที่ 2.4 สมการวัดค่าความคล้ายคลึงในระดับกรณี

เมื่อ p คือ จำนวนของคุณลักษณะ(feature)

ω_i คือ น้ำหนักของคุณสมบัติที่ i โดยที่ $\sum_{i=1}^p \omega_i = 1$

sim_i คือ ค่าความคล้ายคลึงในระดับคุณลักษณะ(Local Similarity) ตัวที่ i

1.3 การค้นคืนกรณีที่คล้ายคลึง (Nearest Neighbor) เป็นวิธีการค้นคืนอย่างง่าย โดยทำการคำนวณค่าความคล้ายคลึงระหว่างกรณีเก่ากับกรณีใหม่ โดยอาศัยการถ่วงน้ำหนักกับคุณลักษณะ(feature) และความแตกต่างระหว่างคุณลักษณะของกรณีใหม่กับกรณีเก่าที่อยู่ในคลังกรณี ถ้ากรณีใดมีความแตกต่างของคุณลักษณะน้อยที่สุดหรือมีความคล้ายคลึงมากที่สุด กรณีนั้นจะถูกคัดเลือก ซึ่งคำนวณได้จากสมการดังรูปที่ 2.5

$$\text{similarity}(Case_I, Case_R) = \frac{\sum_{i=1}^n w_i \times sim(a_i, b_i)}{\sum_{i=1}^n w_i}$$

รูปที่ 2.5 สมการค้นคืนกรณีที่คล้ายคลึง (Nearest Neighbor)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ $\text{Similarity}(\text{Case1}, \text{CaseR})$ คือ คะแนนความคล้ายคลึงระหว่าง Case1 กับ CaseR

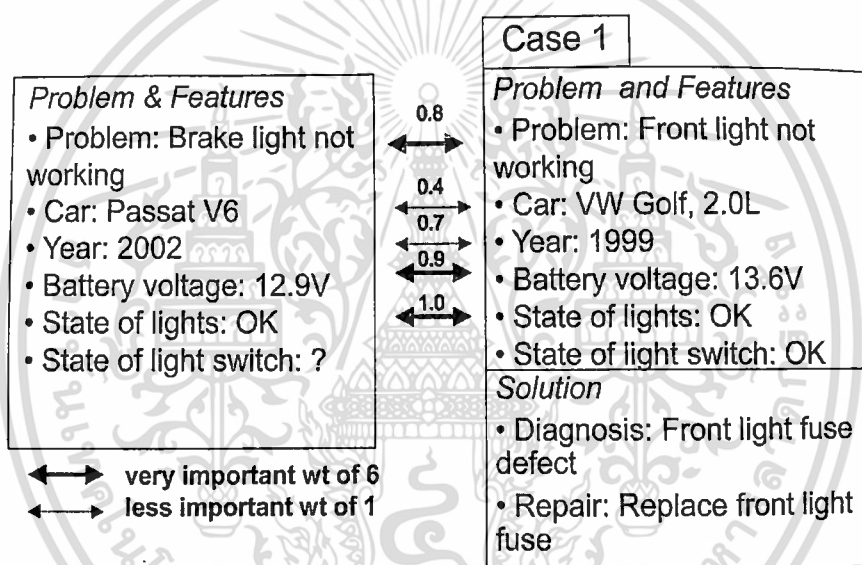
w_i คือ ค่าถ่วงน้ำหนักของแต่ละคุณลักษณะ(feature) ของกรณี ซึ่งสามารถประเมินได้จากถามสอบถามจากผู้เชี่ยวชาญหรือจากผู้ใช้

$\text{sim}(a_i, b_i)$ คือ ค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity)

a_i คือ คุณลักษณะตัวที่ i ใน Case1

b_i คือ คุณลักษณะตัวที่ i ใน CaseR

ตัวอย่างการค้นคืนกรณีที่คล้ายคลึง (Nearest Neighbor)



รูปที่ 2.6 ตัวอย่างการค้นคืนกรณีที่คล้ายคลึง

$$\text{Similarity}(\text{Case1}, \text{CaseR}) = \frac{(6 \times 0.8) + (1 \times 0.4) + (1 \times 0.7) + (6 \times 0.9) + (6 \times 1.0)}{6 + 1 + 1 + 6 + 6} = 0.87$$

จากรูปที่ 2.6 กรณีใหม่มีความคล้ายคลึงกับกรณีที่ 1 ในฐานกรณีอยู่ 0.87 ซึ่งคำนวณได้จากค่าเฉลี่ยถ่วงน้ำหนักของความคล้ายคลึงในระดับคุณลักษณะ(feature) โดยคุณลักษณะแต่ละตัวจะได้รับค่าถ่วงน้ำหนักตามความสำคัญของปัญหา

2.2.3 การนำกรณีเก่ากลับมาใช้ใหม่ (Reuse Case)

การนำกรณีเก่ากลับมาใช้ใหม่ (Reuse Case) เป็นขั้นตอนการนำแนวทางการแก้ปัญหาของกรณีเก่าที่ถูกค้นคืนมาใช้กับปัญหาใหม่ การปรับเปลี่ยนแนวทางการแก้ปัญหาของกรณีที่ถูกค้นคืนให้มีความเหมาะสมกับปัญหาใหม่มากที่สุดนั้นสำหรับบางโดเมนปัญหาอาจเป็นขั้นตอนที่ยากและไม่ชัดเจน ทำให้ต้องใช้ผู้เชี่ยวชาญมาช่วยพิจารณาแทนที่จะทำโดยอัตโนมัติ แต่วิธีในการปรับเปลี่ยนแนวทางการแก้ปัญหาแบบอัตโนมัติก็มีอยู่หลายวิธี ดังนี้

1. แทนค่าวัตถุใหม่ลงไปในการตอบจากกรณีเดิม (Reinstantiation) คำตอบจากกรณีเดิมอาจไม่รวมวัตถุใหม่ที่ต้องการ ดังนั้น อาจสร้างคำตอบใหม่ด้วยการแทนวัตถุเดิมด้วยวัตถุใหม่จากปัญหาใหม่ เช่น การจัดทำเมนูอาหารสำหรับงานเลี้ยง กรณีเดิมที่ค้นคืนอาจเหมาะสมทุกอย่าง ยกเว้นว่า มีแขกคนหนึ่งไม่รับประทานเนื้อวัว ซึ่งอาจแทนที่เนื้อวัวในกรณีเดิม ด้วยเนื้อจากกรณีใหม่ก็ได้

2. ปรับปรุงค่าพารามิเตอร์บางตัว (Parameter Adjustment) วิธีนี้ปรับแต่งค่าพารามิเตอร์ในการตอบจากกรณีเดิม ที่ไม่เหมาะสมกับปัญหาใหม่ให้เหมาะสม เช่น ลูกค้าขอกู้ยืมเงินรายใหม่ มีประวัติคล้ายคลึงกับลูกค้าจากกรณีเดิม ยกเว้นรายได้ ในกรณีเดิมคือ 50000 บาท ขอกู้ในวงเงิน 50000 บาท ส่วนลูกค้ารายใหม่ มีรายได้ 45000 บาท แต่ขอกู้ในวงเงิน 50000 บาทเท่ากัน ซึ่งอาจปรับค่าคำตอบจากกรณีเดิมให้กู้เพียง 45000 บาท

3. ค้นหาที่แทนที่จากโครงสร้างความรู้ต่างๆ (Search for substitutor) เพื่อหาค่าที่เหมาะสมมาแทนในบางส่วนของคำตอบจากกรณีเดิม ที่ใช้งานไม่ได้ในกรณีใหม่ เช่น การลงทะเบียนเรียนของนักศึกษาในเทอมปัจจุบัน อาจใช้กรณีเดียวกับนักศึกษารุ่นปีที่แล้วได้ แต่ในกรณีเดิมมีวิชาเลือกที่ไม่ได้เปิดในเทอมนี้ จึงต้องค้นหาวิชาเลือกใหม่ที่เปิดในเทอมนี้มาแทนที่

4. แทนที่บางขั้นตอนย่อยหรือบางกระบวนการย่อยที่เป็นอิสระต่อกัน (Case-Based Substitution) คำตอบที่ได้ อาจประกอบด้วยขั้นตอนหลายขั้นตอนที่เป็นอิสระต่อกัน ซึ่งบางขั้นตอนอาจไม่เหมาะสมกับปัญหาใหม่ จึงจำเป็นต้องค้นหากรณีอื่นๆ ที่มีขั้นตอนย่อยที่ใช้งานได้ มาแทนที่ขั้นตอนย่อยในคำตอบเดิม

5. เปลี่ยนรูปหรือแปลงรูปบางขั้นตอนย่อย (Transformation) ถ้าไม่มีกรณีใดที่มีขั้นตอนย่อยที่ใช้งานได้ ในกรณีนี้จำเป็นต้องสร้างขั้นตอนย่อยขึ้นมาเองใหม่ จากขั้นตอนย่อยของคำตอบเดิมที่ได้ โดยจำเป็นต้องใช้กฎฮิวริสติกส์ หรือฐานความรู้มาช่วยในการแปลงรูปคำตอบ

6. ซ่อมแซมโดยใช้แนวทางจากแบบจำลอง (Model-Guide Repair) แบบจำลองเชิงเหตุและผลอาจช่วยในกระบวนการแปลงรูปคำตอบได้ โดยการหาความแตกต่างระหว่างปัญหาเดิมกับปัญหาใหม่ และวิเคราะห์ลักษณะเด่นของความแตกต่างที่ส่งผลต่อคำตอบได้ จากนั้นทำการซ่อมแซมคำตอบเดิมด้วยความช่วยเหลือของกฎฮิวริสติกส์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4 การปรับปรุงแนวทางการแก้ไขปัญหา (Revise Case)

หลังจากได้นำเสนอแนวทางการแก้ไขปัญหาลแล้ว แต่มีข้อผิดพลาดหรือยังได้ผลลัพธ์ไม่เป็นที่น่าพอใจ ขั้นตอนนี้จะขั้นตอนแก้ไขปรับปรุงแนวทางการแก้ไขปัญหาให้ถูกต้องเหมาะสม มีการทำงานดังนี้

1. การประเมินแนวทางแก้ไขปัญหา (Evaluate Solution) จะเป็นการประเมินผลลัพธ์หลังจากที่ได้นำเอาแนวทางการแก้ไขปัญหาไปใช้งานว่าประสบความสำเร็จหรือไม่ หากสำเร็จและได้ผลลัพธ์เป็นที่น่าพอใจก็ไม่จำเป็นต้องผ่านกระบวนการแก้ไข โดยปกติแล้วการประเมินแนวทางแก้ไขปัญหาก็จะเกิดขึ้นภายนอกกระบวนการให้เหตุผลเชิงกรณี

2. การแก้ไขแนวทางแก้ไขปัญหาที่ผิดพลาด (Repair fault) หากประเมินแนวทางแก้ไขปัญหาลแล้วไม่เป็นที่น่าพอใจ จะทำการซ่อมแซมแนวทางการแก้ไขปัญหา ซึ่งในทางปฏิบัติกรณีที่เกิดผิดพลาดนี้จะถูกบันทึกไว้เป็นกรณีชั่วคราวที่จะสามารถซ่อมแซมแนวทางการแก้ไขปัญหานั้นกว่าจะเป็นที่ยอมรับได้

2.2.5 การจัดเก็บกรณี (Retain Case)

เป็นขั้นตอนสุดท้ายในวงจรการทำงานของระบบให้เหตุผลเชิงกรณี ซึ่งเกี่ยวข้องกับการจัดเก็บกรณีที่ผ่านการประเมินแล้วว่าประสบความสำเร็จในการแก้ไขปัญหาลไว้ในฐานกรณีหรือคลังกรณี เพื่อเป็นองค์ความรู้สำหรับการแก้ไขปัญหาลต่อไปในอนาคต

2.3 โปรแกรมพัฒนาฐานกรณี โฟทิจ และ มายซีบีอาร์ (Protégé and MyCBR)

2.3.1 โฟทิจ (Protégé)

โฟทิจ (Protégé) เป็นฟรีแพลตฟอร์มโอเพนซอร์สที่จัดเตรียมชุดเครื่องมือเพื่อสร้างแบบจำลอง(model) ภายในขอบเขตเพื่ออธิบายสิ่งที่เราสนใจ (domain) และสร้างแอปพลิเคชันฐานความรู้ด้วยออนโทโลยี (ontology) หัวใจหลักของโฟทิจ คือได้จัดเตรียมชุดโครงสร้างแบบจำลองความรู้และกิจกรรมต่างๆที่สนับสนุนการสร้างรูปภาพเสมือน และการจัดการของออนโทโลยีในรูปแบบต่างๆ นอกจากนี้ยังสามารถปรับแต่งเพื่อสนับสนุนขอบเขตสิ่งที่เราสนใจ (domain) สำหรับการสร้างความรู้และการป้อนข้อมูลความรู้

ออนโทโลยีคือการอธิบายแนวความคิดและความสัมพันธ์ที่มีความสำคัญในโดเมนเฉพาะที่เราสนใจ กำหนดคำศัพท์ที่จะอธิบายถึงขอบเขตและระบุให้คอมพิวเตอร์เข้าใจความหมาย ซึ่งจัดอยู่ในประเภทเดียวกับอนุกรมวิธาน(taxonomy)เป็นเทคนิคเกี่ยวกับการแบ่งประเภทและโครงสร้างฐานข้อมูลจะนำไปสู่ทฤษฎีที่เป็นจริง ปัจจุบันได้ถูกนำไปใช้ในทางวิทยาศาสตร์และธุรกิจจำนวนมากทำให้เกิดการแบ่งปัน และนำความรู้ในโดเมนที่สนใจกลับมาใช้ใหม่อีกครั้ง และตอนนี้ออน

โทโลยีได้เป็นหัวใจหลักของหลายๆ แอปพลิเคชัน เช่น ความรู้ทางวิทยาศาสตร์ การจัดการข้อมูล การรวมระบบ การพาณิชย์อิเล็กทรอนิกส์ และ บริการเว็บเชิงความหมาย

2.3.2 แพลตฟอร์มโทโลยีสนับสนุนแบบจำลองออนโทโลยี 2 แบบดังนี้

1. แบบเฟรม (Frame-based) ผู้ใช้สามารถสร้างและเพิ่มออนโทโลยีตามที่โพรโทคอลโอเคบีซี (OKBC – Open Knowledge Base Connectivity Protocol) กำหนด ในแบบจำลองนี้ออนโทโลยีประกอบด้วย

- ชุดของคลาสที่จัดอยู่ในลักษณะของลำดับชั้นเป็นตัวแทนของแนวคิดที่สำคัญของโดเมน
- ชุดของสล็อต (slot) ที่มีความเกี่ยวข้องกับคลาสเพื่ออธิบายคุณสมบัติและความสัมพันธ์ของคลาส

- ชุดของอินสแตนซ์ของคลาส อินสแตนซ์คือเนื้อข้อมูล

2. โอดับเบิลยูเอล (OWL) ผู้ใช้สามารถสร้างออนโทโลยีสำหรับเว็บเชิงความหมาย โดยเฉพาะ โอดับเบิลยูเอล (OWL - W3C's Web Ontology Language) โอดับเบิลยูเอลจะรวมถึงคำอธิบายของคลาส คุณสมบัติ (properties) และอินสแตนซ์

ในโครงการนี้ขอกกล่าวถึงแบบจำลองแบบเฟรม เนื่องจากเป็นโครงการพัฒนาฐานกรณี ซึ่งแบบจำลองแบบเฟรมเป็นแบบจำลองที่เหมาะสมในการพัฒนาฐานกรณี เพราะมีโครงสร้างคล้ายคลึงกันระหว่างเฟรม(frame)และกรณี(case)

2.3.3 ลักษณะเด่นของแบบจำลองแบบเฟรมในโปรแกรมโทโลยี

1. เป็นชุดขององค์ประกอบส่วนต่อประสานผู้ใช้ ที่สามารถปรับแต่งเพื่อให้ผู้ใช้โมเดลองค์ความรู้และป้อนข้อมูลได้อย่างง่าย

2. เป็นสถาปัตยกรรมแบบปลั๊กอินที่สามารถเพิ่มเติมได้ เช่น ส่วนประกอบกราฟิก กราฟ และตารางรูปแบบการจัดเก็บข้อมูล อาร์ดีเอฟ เอกซ์เอ็มแอล เอชทีเอ็มแอล ฐานข้อมูล และเครื่องมือสนับสนุนเพิ่มเติม สำหรับจัดการออนโทโลยี ภาพแสดงออนโทโลยี การอนุมานและการหาเหตุผล

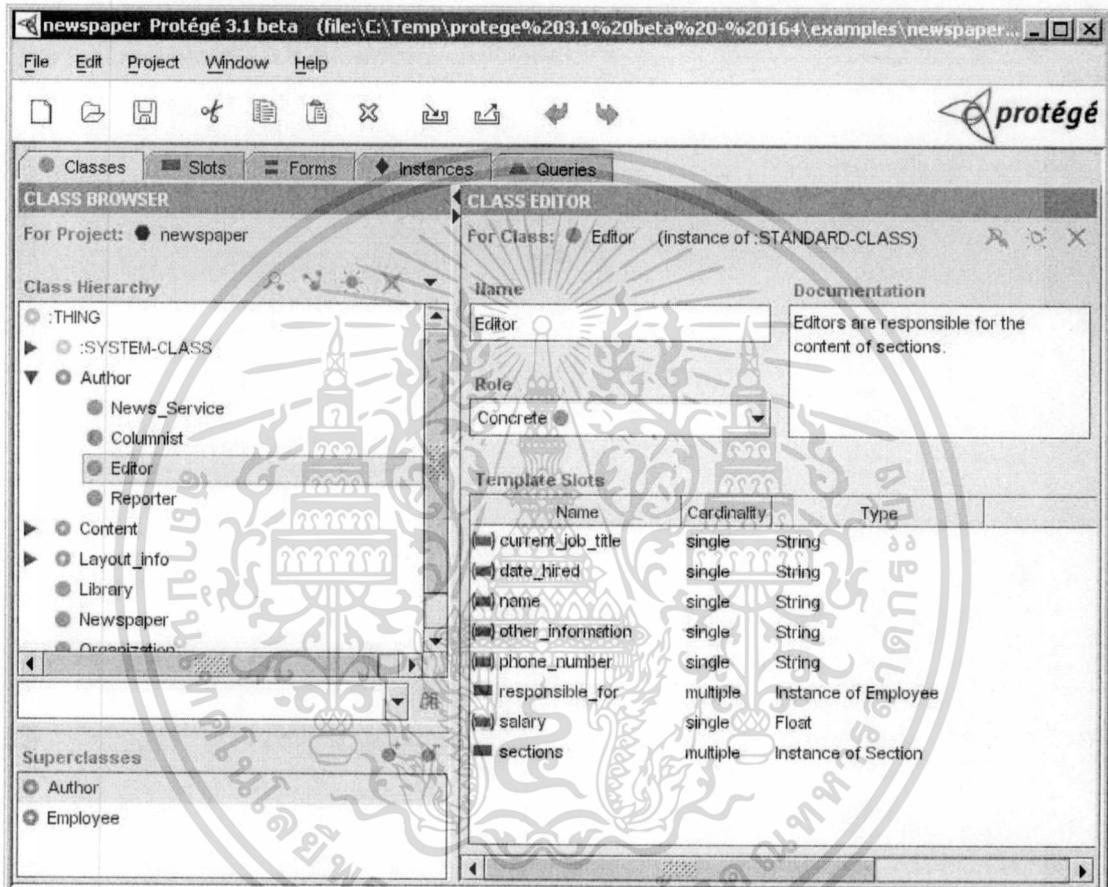
3. เป็นส่วนต่อประสานโปรแกรมโดยใช้จาวาสำหรับปลั๊กอินแอปพลิเคชันอื่นๆ เพื่อเข้าถึง เข้าใช้ และแสดงออนโทโลยีที่ถูกสร้างด้วยโปรแกรมโทโลยี

2.3.4 ภาพหน้าจอโปรแกรมโททิจของแบบจำลองแบบเฟรม

โททิจแบบจำลองแบบเฟรม ส่วนต่อประสานผู้ใช้จะถูกจัดตามชุดแท็บ (tab) ที่มุ่งเน้นมุมมองของการสร้างออนโทโลยี

แท็บคลาสเป็นหน้าจัดการออนโทโลยีซึ่งสามารถใช้นิยามคลาสและลำดับชั้นของคลาส สล็อตและข้อจำกัดของสล็อต ความสัมพันธ์ระหว่างคลาสและคุณสมบัติของความสัมพันธ์ ดังรูปที่

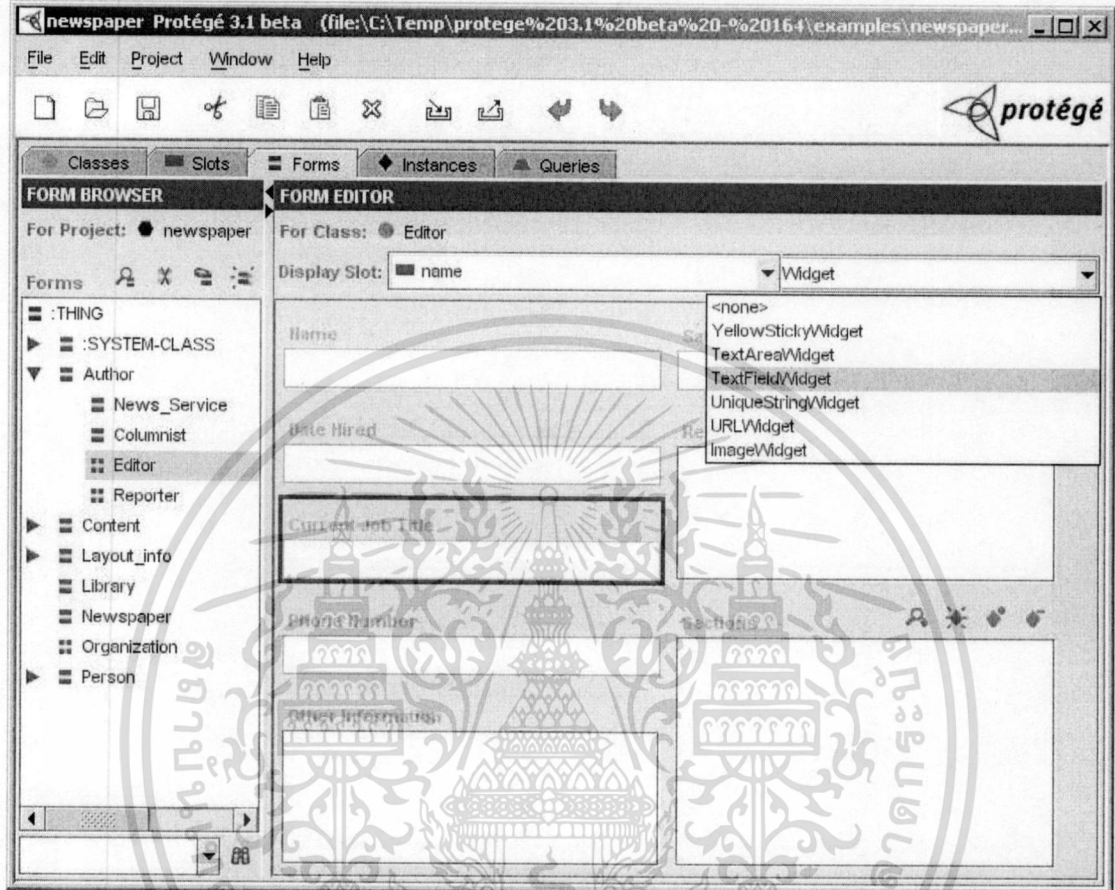
2.7



รูปที่ 2.7 แท็บคลาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

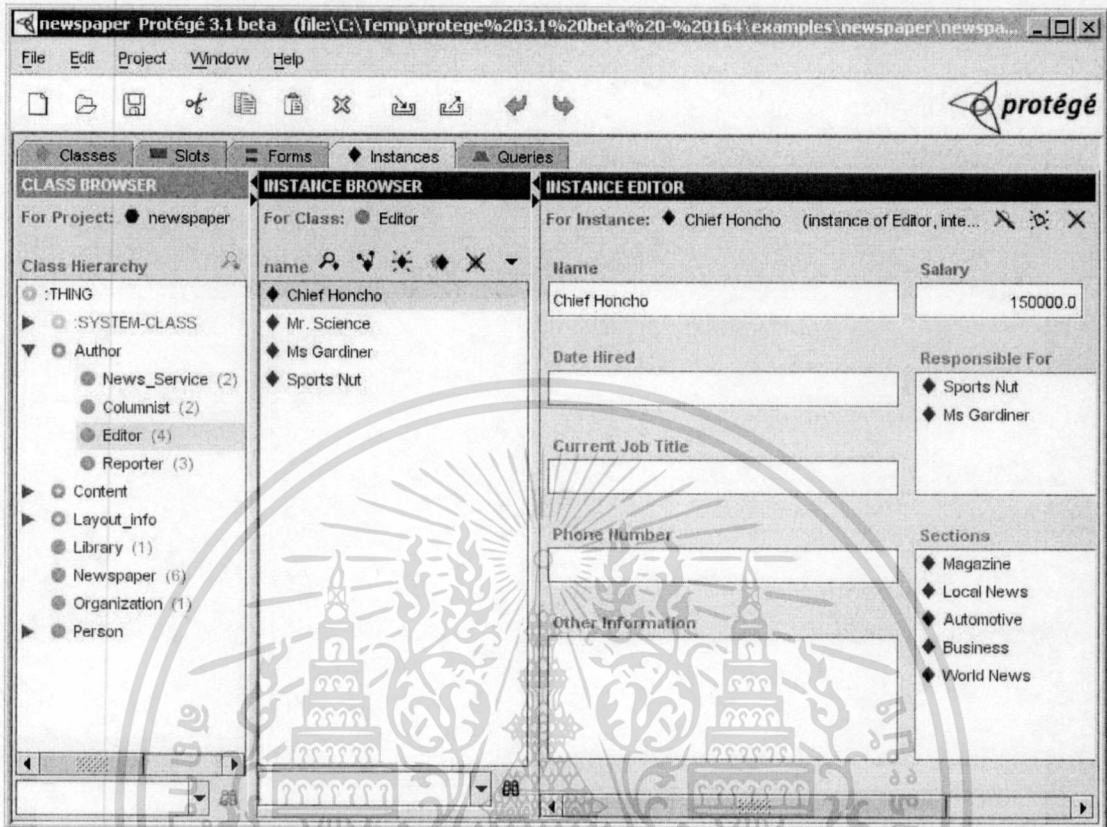
แท็บฟอร์ม(Form) เป็นหน้าฟอร์มสำหรับกรอกเพื่อแสดงอินสแตนซ์ในแท็บอินสแตนซ์ตามชนิดของสล็อตที่สร้างไว้ สามารถเปลี่ยนค่าเริ่มต้นของฟอร์มโดยการจัดเรียงฟิลด์บนหน้าจอ เปลี่ยนขนาด ป้ายชื่อ และคุณสมบัติอื่นๆสำหรับแต่ละสล็อตได้ ดังรูปที่ 2.8



รูปที่ 2.8 แท็บฟอร์ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แท็บอินสแตนซ์เป็นเครื่องมือที่ได้มาซึ่งความรู้ สามารถใช้สร้างอินสแตนซ์ของคลาสที่นิยามไว้ในออนโทโลยี ดังรูป 2.9



รูปที่ 2.9 แท็บอินสแตนซ์

2.3.5 มายซีบีอาร์ (MyCBR)

มายซีบีอาร์ (MyCBR) เวอร์ชัน 2.6.6 เป็นเครื่องมือพัฒนาระบบให้เหตุผลเชิงกรณี (Case-Based Reasoning) แบบ โอเพนซอร์สซึ่งสามารถใช้งานได้โดยไม่เสียค่าใช้จ่าย พัฒนาโดยศูนย์วิจัยเยอรมันเพื่อปัญญาประดิษฐ์ (the German Research Center for Artificial Intelligence) โดยแรงจูงใจสำคัญในการพัฒนามายซีบีอาร์คือความจำเป็นในการใช้เครื่องมือที่ง่ายต่อการสร้างแอปพลิเคชันต้นแบบการให้เหตุผลเชิงกรณีในการเรียน การทำวิจัย และ โครงการอุตสาหกรรมขนาดเล็กด้วยความพยายามที่น้อยที่สุด ปัจจุบันเวอร์ชันของมายซีบีอาร์ยังคงเน้นไปที่ขั้นตอนการค้นคืนโดยการวัดค่าความคล้ายคลึงในขั้นตอนของวงจรการทำงานของระบบให้เหตุผลเชิงกรณี

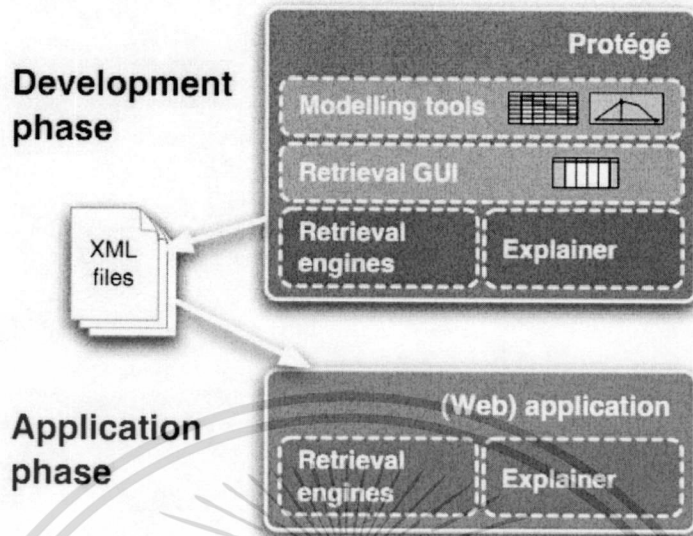
เป้าหมายหลักของมายซีบีอาร์คือการลดความพยายามให้เหลือน้อยที่สุดสำหรับสร้างแอปพลิเคชันการให้เหตุผลเชิงกรณีที่ต้องการการวัดค่าความคล้ายคลึงของความรู้ ดังนั้นจึงเตรียมส่วนต่อประสานผู้ใช้แบบกราฟิกที่สะดวกสำหรับการสร้างแบบจำลองที่เกี่ยวกับการวัดค่าความคล้ายคลึงของแอดทริบิวต์และการประเมินคุณภาพของผลลัพธ์ของการค้นคืน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาในเอกสารนี้ส่งถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.6 สถาปัตยกรรมของมายซีบีอาร์



รูปที่ 2.10 สถาปัตยกรรมของมายซีบีอาร์

สถาปัตยกรรมพื้นฐานของมายซีบีอาร์ ดังรูปที่ 2.10 ในระหว่างขั้นตอนการพัฒนาแอปพลิเคชันของการให้เหตุผลเชิงกรณี มายซีบีอาร์ทำงานเป็นปลั๊กอินภายในโททิจे ซึ่งประกอบด้วยโมดูลดังต่อไปนี้

1. เครื่องมือการสร้างแบบจำลอง (Modelling tool) เครื่องมือนี้เพิ่มฟังก์ชันการทำงานที่มีอยู่เดิมของโททิจेสำหรับสร้างแบบจำลองสิ่งที่สนใจ (domain) อินสแตนซ์ของกรณี และเพิ่มฟังก์ชันการทำงานที่ขาดหายไปสำหรับการกำหนดการวัดค่าความคล้ายคลึง
2. ส่วนต่อประสานผู้ใช้ของการค้นคืน (Retrieval GUI) ได้จัดเตรียมเครื่องมือที่มีประสิทธิภาพสำหรับวิเคราะห์คุณภาพของการกำหนดค่าการวัดค่าความคล้ายคลึง นอกจากนี้ยังสามารถใช้เป็นต้นแบบแรกของแอปพลิเคชันการให้เหตุผลเชิงกรณีได้อีกด้วย
3. เครื่องมือการค้นคืน (Retrieval engines) สำหรับการค้นคืนโดยใช้ค่าความคล้ายคลึง ได้จัดเตรียมเครื่องมือการค้นคืนไว้ให้
4. ส่วนอธิบาย (Explainer) ส่วนคำอธิบายได้จัดเตรียมเพื่อใช้เป็นข้อมูลสนับสนุนการสร้างแบบจำลองและคำอธิบายของผลการค้นคืนสำหรับการออกแบบและทดสอบ

สุดท้ายของการสร้างแบบจำลองและขั้นตอนการพัฒนา โดเมนที่สมบูรณ์และแบบจำลองความคล้ายคลึงกับฐานกรณีสามารถนำออกมาเป็นไฟล์เอกซ์เอ็มแอล (XML) และหลังจากการพัฒนาแอปพลิเคชันการให้เหตุผลเชิงกรณีโดยใช้ปลั๊กอินของโททิจе มายซีบีอาร์ยังสามารถใช้เป็นสแตนด์อโลน โมดูลจาวา เพื่อนำไปรวมกับแอปพลิเคชันตามต้องการ เช่น เว็บแอปพลิเคชันโดย

ใช้เจเอสพี เป็นต้น ในขั้นตอนการพัฒนาแอปพลิเคชัน เครื่องมือการค้นคืนของมายซีบีอาร์จะอ่านจากไฟล์เอกซ์เอ็มแอลที่สร้างขึ้นในระหว่างขั้นตอนการพัฒนาและค้นคืน โดยใช้ค่าความคล้ายคลึง

2.3.7 แบบจำลองการวัดค่าความคล้ายคลึงของมายซีบีอาร์

จุดเริ่มต้นของโครงการทำให้เหตุผลเชิงกรณีคือการสร้างกรณี ซึ่งจะต้องมีการกำหนดคลาสและแอตทริบิวต์ของคลาสในโพทิจเรียกว่าแอตทริบิวต์ว่า “สล็อต” และระบุช่วงค่าที่ถูกต้องสำหรับการวัดค่าความคล้ายคลึง หลังจากสร้างกรณีแล้วงานหลักๆของการสร้างแอปพลิเคชันการให้เหตุผลเชิงกรณีโดยใช้มายซีบีอาร์คือการกำหนดการวัดค่าความคล้ายคลึงที่เหมาะสม ตามระดับคุณลักษณะ(Local) และระดับกรณี (Global) ซึ่งแบ่งการกำหนดความคล้ายคลึงเป็น ชุดของการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) สำหรับแต่ละแอตทริบิวต์หรือสล็อต ชุดค่าถ่วงน้ำหนักของแอตทริบิวต์ และการวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity) สำหรับการคำนวณการวัดค่าความคล้ายคลึงโดยรวม กล่าวคือตัวแทนกรณี(Case)ประกอบด้วย n แอตทริบิวต์ และความคล้ายคลึงระหว่างการสืบค้น q และกรณี c อาจจะสามารถคำนวณได้ดังรูปที่ 2.11

$$sim(q, c) = \sum_{i=1}^n \omega_i sim_i(q_i, c_i)$$

รูปที่ 2.11 สมการวัดค่าความคล้ายคลึงในระดับกรณีแบบถ่วงน้ำหนักรวม

sim_i และ w_i แสดงถึงการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) และค่าถ่วงน้ำหนักของแอตทริบิวต์ที่ i ส่วน Sim แสดงถึงการวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity)

รูปที่ 2.11 เป็นการวัดค่าความคล้ายคลึงในระดับกรณีแบบผลรวมถ่วงน้ำหนัก (Weighted Sum) ผู้ใช้สามารถเลือกใช้ฟังก์ชันอื่นๆ เช่น ออกติน้อยที่สุด(minimum) ออกติมากที่สุด(maximum) และระยะห่างแบบยูคลิด (Euclidean distance) ทั้งการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) และการวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity) ผู้ใช้สามารถกำหนดและจัดการชุดของการวัดค่าความคล้ายคลึงที่แตกต่างกันได้ การวัดค่าความคล้ายคลึงที่ถูกเลือกจะถูกใช้ในการค้นคืน

การวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) ของมายซีบิวาร์จะมีให้ เลือกรูปแบบตามประเภทของแอตทริบิวต์ดังนี้

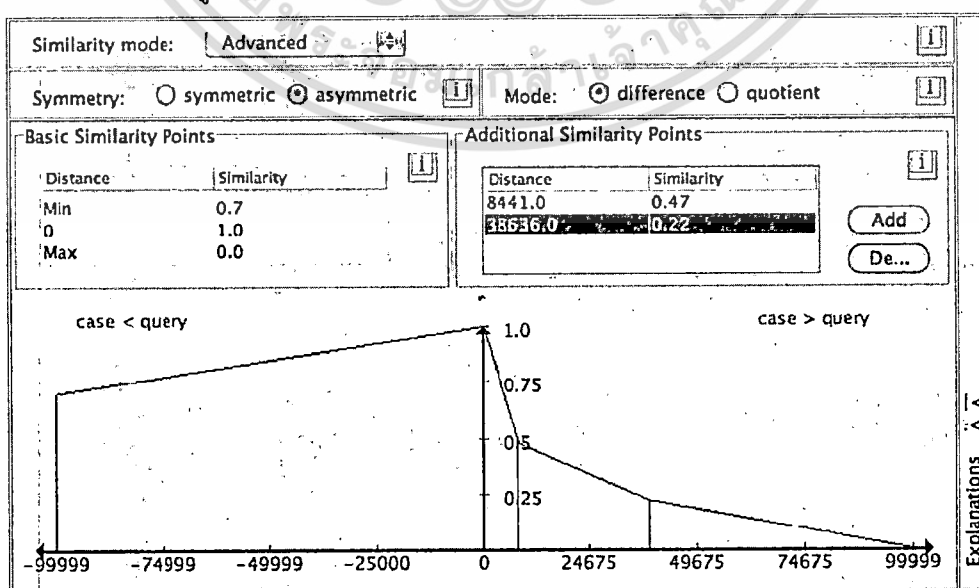
ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทตัวเลข (Similarity Editors for Numeric Attributes) สำหรับแอตทริบิวต์ประเภทตัวเลข การคำนวณการวัดความคล้ายคลึงเป็นการแมป (map) ระหว่างระยะห่างของค่าสองค่าที่เปรียบเทียบกับค่าความเหมือนตามที่ต้องการ ให้เป็น

$$sim_i(q_i, c_i) = f(d(q_i, c_i))$$

รูปที่ 2.12 สมการวัดค่าความคล้ายคลึงในระดับคุณลักษณะประเภทตัวเลข

รูปที่ 2.12 เป็นการกำหนดการแมปค่าของฟังก์ชัน f กับระยะห่างฟังก์ชัน d ที่กำหนด สำหรับ d มายซีบิวาร์ได้เตรียมสองทางเลือกในการคำนวณฟังก์ชัน d คือ แบบความแตกต่างสัมบูรณ์ (absolute difference) $d(q_i, c_i) = c_i - q_i$ หรือ แบบผลหาร (quotient) $d(q_i, c_i) = c_i/q_i$ ของสองค่า c, q

สำหรับการสร้างแบบจำลองการแมปฟังก์ชัน f มายซีบิวาร์ได้เตรียมตัวจัดการไว้ 2 โหมด คือ โหมดมาตรฐาน (standard mode) ผู้ใช้สามารถเลือกกระหว่างฟังก์ชันแบบทั่วไป (typical) และแบบปรับได้ (adjustable) เช่น แบบฟังก์ชันลดเชิงเส้นกำกับ (asymptotic decreasing) โหมดขั้นสูง (advanced mode) ผู้ใช้สามารถปรับเปลี่ยนฟังก์ชันการแมปได้ตามต้องการโดยใช้การประมาณเชิงเส้น โดยทำการระบุจุดของจุดที่ต้องการ ซึ่งจุดเหล่านี้สามารถสร้างได้ง่ายๆโดยใช้การลากและวางในโปรแกรมโพทเจ ดังรูปที่ 2.13



เอกสารนี้เป็นเอกสารที่รูปที่ 2.13 ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทตัวเลข ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทสัญลักษณ์ (Similarity Editors for Symbolic Attributes) สำหรับแอตทริบิวต์ประเภทสัญลักษณ์มีหลายความเป็นไปได้ที่จะจำลองความคล้ายคลึง โดยทั่วไปจะใช้ตารางความคล้ายคลึง (similarity table) ทุกคู่ของค่าความคล้ายคลึงจะพหุคูณกัน และจำแนกอย่างชัดเจน ดังภาพที่ 2.14 และมายซีบีอาร์จะทำการไฮไลต์ค่าความคล้ายคลึงให้ โดยน้ำหนัของสีแสดงถึงความคล้ายคลึงกัน

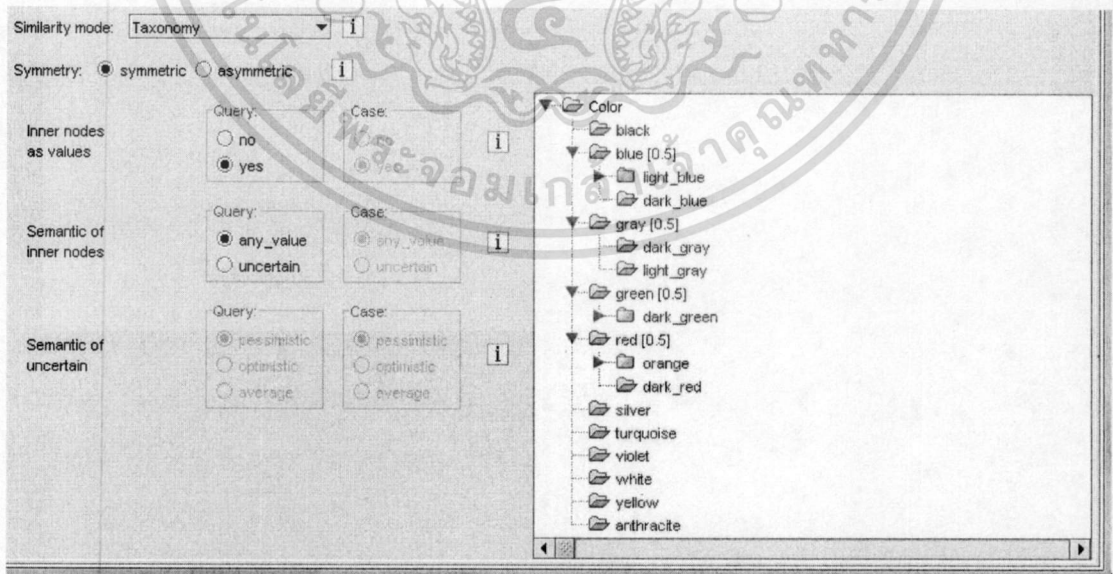
Similarity mode: Table

Symmetry: symmetric asymmetric

		Case Base Values					
Reset		convertible	coupe	fastback	roadster	sedan	station_wagon
convertible	1.0	0.0	0.3	0.7	0.0	0.0	
coupe	0.0	1.0	0.2	0.0	0.7	0.5	
fastback	0.3	0.2	1.0	0.5	0.0	0.0	
roadster	0.7	0.0	0.5	1.0	0.0	0.0	
sedan	0.0	0.7	0.0	0.0	1.0	0.8	
station_wagon	0.0	0.5	0.0	0.0	0.8	1.0	

รูปที่ 2.14 ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทสัญลักษณ์แบบตารางความคล้ายคลึง

อย่างไรก็ตามถ้าชุดของค่าแอตทริบิวต์มีจำนวนมากการกำหนดค่าความคล้ายคลึงโดยใช้ตารางความคล้ายคลึงจะเป็นงานที่ลำบากพอสมควร ดังนั้นมายซีบีอาร์ได้เตรียมการกำหนดค่าความคล้ายคลึงที่ง่ายกว่า คือ การจัดกลุ่ม (taxonomy) ลำดับแรกเป็นลำดับของสัญลักษณ์ที่ยอมให้กำหนดค่าความคล้ายคลึงโดยใช้ค่าตัวเลขและการจัดตำแหน่งในลำดับ หลังจากนั้นเป็นการจัดกลุ่มของสัญลักษณ์ที่เหมือนกันโดยใช้วิธีลากและวางไปใส่ในสัญลักษณ์ที่มีความคล้ายคลึงกัน ดังภาพที่ 2.15



รูปที่ 2.15 ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทสัญลักษณ์แบบการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทสายอักขระ (Similarity Editors for String Attributes) แม้ว่าข้อความของการให้เหตุผลเชิงกรณี ภาษามายซีบีอาร์จะไม่ได้ให้ความสำคัญเท่าไรนัก แต่ก็มีเครื่องมือการวัดค่าความคล้ายคลึงกันของการประมวลผลสายอักขระ ผู้ใช้สามารถเลือกระหว่างโหมดการประมวลผลของคำหรืออักขระและขึ้นอยู่กับโหมดการกำหนดค่าและวิธีการต่างๆที่เลือก เพื่อระบุการคำนวณค่าความคล้ายคลึง เช่น การเปรียบเทียบแบบเหมือน(exact) แบบเหมือนเฉพาะส่วน(partial match) แบบไตรแกรมหรือนิพจน์ปรกติ (regular expression)

ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ประเภทชุดข้อมูล (Similarity Editors for Set Attributes) แอตทริบิวต์ยอมให้มีค่าได้หลายๆนั้นการคำนวณค่าความคล้ายคลึงจะมีความซับซ้อนมากกว่าแบบที่ยอมให้มีแค่ค่าเดียว โดยทั่วไปความหมายของการเปรียบเทียบของชุดค่าเป็นการระบุเฉพาะเจาะจงสำหรับแอปพลิเคชันนั้นๆ โดยเฉพาะ เช่น ชุดของค่าอาจจะมีชนิดของความหมาย “และ” หรือชนิดของ “หรือ” นอกจากนี้ขนาดของชุดสอบถามและกรณีอาจมีอิทธิพลที่แตกต่างกันของค่าความคล้ายคลึง

ภาษามายซีบีอาร์ได้เตรียมตัวเลือกต่างๆเพื่อปรับเปลี่ยนการวัดค่าความคล้ายคลึงของชุดแอตทริบิวต์ ขึ้นอยู่กับการตั้งค่าที่เลือกแม่ว่าระหว่างค่าของชุดสอบถามและค่าของกรณีที่มีการคำนวณที่แตกต่างกัน เช่นอาจต้องการให้ตรงกับแต่ละค่าของข้อมูลสอบถามกับค่าของกรณีที่เหมาะสมที่สุดหรือในทางกลับกัน นอกจากนี้ค่าของข้อมูลสอบถามหรือกรณี เช่น ข้อมูลสอบถามมีจำนวนมากกว่าของกรณี อาจไม่สามารถจับคู่กับค่าของกรณีหรือข้อมูลสอบถามได้ อาจมีผลกระทบในการวัดค่าความคล้ายคลึง แต่เมื่อการแมปได้เลือกการคำนวณการวัดค่าความคล้ายคลึงเป็นแบบปกติสำหรับค่าเดี่ยวๆของชุดข้อมูล ตัวจัดการความคล้ายคลึงของแอตทริบิวต์ที่ได้อธิบายมาแล้วสามารถเลือกใช้ได้

การวัดค่าความคล้ายคลึงโดยใช้คำสั่งและการวัดค่าความคล้ายคลึงจากภายนอก (Script-base and external Similarity Measures) เพื่อความยืดหยุ่นในการวัดค่าความคล้ายคลึงสำหรับทุกประเภทของข้อมูลมีทางเลือกอีก 2 แนวทางที่ได้จัดเตรียมไว้

การวัดค่าความคล้ายคลึงโดยใช้คำสั่ง (Script) : ภาษามายซีบีอาร์ได้เตรียมภาษาไจทอนและตัวจัดการให้ผู้ใช้สามารถเขียนคำสั่งการวัดค่าความคล้ายคลึงได้ตามต้องการ

การวัดค่าความคล้ายคลึงจากภายนอก (External) : โหมดการวัดค่าความคล้ายคลึงนี้ อนุญาตให้ผู้ใช้สามารถเรียกโปรแกรมจากภายนอกได้ เช่น โปรแกรมที่เขียนโดยภาษา ซี หรือ ซีพลัสพลัส (C/C++) เพื่อคำนวณค่าความคล้ายคลึง ซึ่งจะเป็นประโยชน์โดยเฉพาะอย่างยิ่งในกรณีที่โปรแกรมโพทิจเไม่สนับสนุนประเภทของข้อมูลบางชนิด เช่น รูปภาพ ในกรณีนี้แอตทริบิวต์ในตัวแทนกรณีอาจระบุด้วยยูอาร์แอล (URL) ไปยังแหล่งข้อมูลภายนอกที่ใช้โปรแกรมภายนอกในการเข้าถึงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจัดการกับค่าที่ขาดหายไป (Dealing with Missing Values) ค่าแอตทริบิวต์ที่ขาดหายไปทั้งในข้อมูลสอบถามหรือในกรณี (Case) มักเป็นประเด็นที่สำคัญของการคำนวณค่าความคล้ายคลึงเพราะอาจทำให้เกิดอุปสรรคในการคำนวณการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) ทั้งนี้ขึ้นอยู่กับโดเมนของแอปพลิเคชัน ค่าที่ขาดหายไป (missing value) อาจมีความหมายที่ต่างกัน ยกตัวอย่างเช่น ในระบบแนะนำสินค้า ค่าที่หายไปของแอตทริบิวต์ข้อมูลสอบถามโดยทั่วไปจะหมายความว่าไม่สนใจข้อความของลูกค้า ในขณะที่ค่าที่หายไปของแอตทริบิวต์ของกรณีสอดคล้องกับไม่รู้หรือไม่มีคุณสมบัติของสินค้า

ในมายซีบีอาร์ค่าที่ขาดหายไปจะถูกแทนด้วยค่าเฉพาะ คือ “_ไม่นิยาม_” (undefined) อย่างไรก็ตามผู้ใช้สามารถระบุค่าของตัวเองเพิ่มเติมได้ เพื่อให้ครอบคลุมความต้องการของแอปพลิเคชันอิทธิพลของค่าเฉพาะในการคำนวณค่าความคล้ายคลึงสามารถปรับเปลี่ยนได้ในแต่ละรายการ

2.3.8 การทดสอบการทำงานของกรค้นคืน (Testing of Retrieval Functionality)

การกำหนดการวัดค่าความคล้ายคลึงให้ดีขึ้นมักเป็นงานที่ยุ้งยากและต้องใช้การทดสอบและปรับแต่งซ้ำๆ ด้วยเหตุผลนี้มายซีบีอาร์ได้รวมส่วนต่อประสานผู้ใช้แบบกราฟิกที่สะดวกสบายสำหรับการดำเนินการค้นคืนและวิเคราะห์ความสอดคล้องของผลลัพธ์ในรายละเอียด รูปที่ 2.16 ขวามือของหน้าต่างเป็นภาพรวมของผลการค้นคืนทั้งหมด ส่วนกลางของส่วนต่อประสานผู้ใช้แบบกราฟิกเป็นข้อมูลสอบถามและกรณีที่ถูกค้นคืนขึ้นมา โดยมีฟังก์ชันคำอธิบายและการไฮไลต์ค่าความคล้ายคลึง ซึ่งจะเห็นว่ามายซีบีอาร์สนับสนุนการวิเคราะห์ที่มีประสิทธิภาพของผลลัพธ์ในการคำนวณการวัดค่าความคล้ายคลึง

Attribute	Case 1 (455)	Case 2 (742)	Case 3 (905)	Similarity %
Color	black	black	dark_gray	50%
Manufacturer	audi	audi	audi	100%
Price	25000	25599	16099	97%
Body	sedan	station_wagon	station_wagon	80%
Extras	[Airbag, Central_Lockin...	[Airbag, ESP, Cen...	[ABS, ESP, Servo, ...	50%
Model	a6_2.8	a4_1.6	a4_1.8	25%
Sound	Subwoofer	undefined	undefined	undefined
Bass	undefined	undefined	undefined	undefined
Box_Brand	Bose	undefined	undefined	undefined
Box_amount	6	undefined	undefined	undefined
CD_Player	[Mp3CD]	[Changer, Mp3CD]	undefined	100%
Receiver	undefined	undefined	undefined	undefined
CCM	2800	1600	1800	undefined
Car Code	455	742	905	undefined
Doors	4	5	5	undefined
Gas	gasoline	gasoline	gasoline	100%
Miles	105674	85917	120734	undefined
Power	193	101	125	undefined
Speed	213	183	183	undefined
Year	1994.0	1995.0	1995.0	undefined
ZIP	4	6	0	undefined

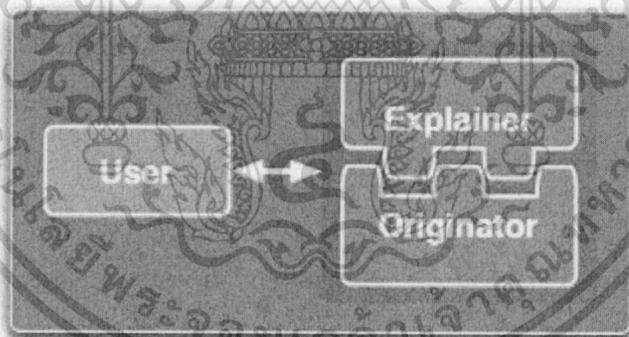
รูปที่ 2.16 ส่วนต่อประสานกรค้นคืน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.10 ฟังก์ชันคำอธิบาย (Explanation Functionality)

ความง่ายในการใช้งานของระบบซอฟต์แวร์ใดๆจะดีขึ้น โดยเพิ่มการทำความเข้าใจในระบบซอฟต์แวร์เข้าไป ซึ่งทำได้โดยให้ความช่วยเหลือในเรื่องการอธิบายสิ่งต่างๆในระบบซอฟต์แวร์อย่างเหมาะสม คำอธิบายเป็นวิธีที่มนุษย์ใช้มากที่สุดเพื่อทำความเข้าใจและการตัดสินใจในทุกๆวันมนุษย์สื่อสารกันโดยใช้คำอธิบายเป็นสื่อกลางสำคัญในการถ่ายทอดข้อมูลเพื่อให้เข้าใจซึ่งกันและกัน การอธิบายจะช่วยให้ความรู้กับผู้รับสารทำให้ผู้รับสารยอมรับข้อความที่เราสื่อสารได้ง่ายขึ้น เนื่องจากผู้รับสารเข้าใจมากขึ้นก็จะตัดสินใจได้ง่ายขึ้น

สถานการณ์การอธิบายของมุมมองการสื่อสารประกอบไปด้วย 3 ผู้เกี่ยวข้อง ดังรูปที่ 2.18 ผู้เริ่มต้น (originator) เป็นระบบหรือเอเจนต์ (agent) ที่เตรียมบางสิ่งบางอย่างเพื่อถูกอธิบาย เช่น การแก้ไขปัญหาของบางปัญหา อุปกรณ์ทางเทคนิค แผนและการตัดสินใจ ในกรณีของมายซีบีอาร์ ผู้เริ่มต้นประกอบด้วยเครื่องมือสร้างแบบจำลองต่างๆและเครื่องมือการค้นคืน ผู้ใช้ (user) เป็นผู้รับขอคำอธิบาย ผู้อธิบาย (explainer) เป็นผู้ทำการอธิบายผู้ใช้ เอเจนต์ใช้ในการถ่ายโอนความตั้งใจของผู้เริ่มต้นกับผู้ใช้ให้ถูกต้องที่สุดเท่าที่ทำได้ ผู้อธิบายเลือกรูปแบบของการอธิบายและรับผิดชอบมุมมองสำหรับการจับบทสนทนาถ้าจำเป็น ผู้เริ่มต้นและผู้อธิบายต้องทำงานร่วมกันเพื่อปรับปรุงการสื่อสารกับผู้ใช้ให้ดีขึ้น ผู้เริ่มต้นต้องให้ข้อมูลที่เหมาะสมเพื่อให้คนอื่นอธิบายสร้างการอธิบายด้วย



รูปที่ 2.18 ผู้เกี่ยวข้องในสถานการณ์การอธิบาย

เพื่อรองรับสถานการณ์การสื่อสารที่อธิบายข้างต้น มายซีบีอาร์ได้เตรียมรูปแบบของการอธิบายไว้สองชนิดคือ การอธิบายไปข้างหน้าและย้อนหลัง (forward and backward explanations) การอธิบายไปข้างหน้าเป็นการอธิบายโดยอ้อม ที่เสนอวิธีการที่แตกต่างของการเพิ่มประสิทธิภาพการได้มาซึ่งผลลัพธ์และความเป็นไปได้ของการวิเคราะห์อุปกรณ์หรือแอปพลิเคชัน การอธิบายย้อนหลังเป็นการอธิบายผลของกระบวนการและวิธีการที่สร้างขึ้น

เพื่อความถูกต้องของกระบวนการค้นคืน มายซีบีอาร์ได้สร้างวัตถุการอธิบายสำหรับแต่ละกรณีระหว่างการคำนวณความคล้ายคลึง เหมือนโครงสร้างข้อมูลแบบต้นไม้เก็บค่าความคล้ายคลึง เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในระดับคุณลักษณะ (Local Similarity) และค่าความคล้ายคลึงในระดับกรณี (Global Similarity) ตามความคิดเห็น (comment) ของแต่ละแอตทริบิวต์ รายละเอียดการค้นคืนเหล่านี้จะถูกแสดงเพื่อให้ผู้ใช้เห็นในส่วนต่อประสานผู้ใช้การค้นคืน ดังรูป 2.16 ไม่ว่าจะเป็นรายละเอียดของเครื่องมือต่างๆ หรือฟอร์มพร้อมค่าแอตทริบิวต์ของกรณี เช่น ราคารถ ไออีอาร์ 455 (26,899) เป็น 88% คล้ายคลึงกับราคารถยนต์ที่สอบถาม (25,000) อีกคุณสมบัติที่มีความสำคัญคือตัวเลือก (option) ที่จะหากรณีที่คล้ายคลึงกันมากที่สุดด้วยค่าแอตทริบิวต์เพียงค่าเดียวโดยคลิกที่ชื่อของแอตทริบิวต์ ในกรณีที่แอตทริบิวต์ของกรณีมีมากเกินไปอาจต้องการเรียงลำดับค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) ของกรณีหนึ่งจากน้อยไปมากหรือมากไปน้อย ก็สามารถทำได้โดยการคลิกที่ชื่อของแต่ละกรณีที่ถูกค้นคืนขึ้นมา

ขณะที่กำลังพัฒนาระบบให้เหตุผลเชิงกรณีคำถามที่สำคัญคือการวัดค่าความคล้ายคลึงจะนำไปสู่กรณีที่เหมาะสมสำหรับการสอบถาม การอธิบายไปยังข้างหน้าช่วยทำนายพฤติกรรมของระบบระหว่างการสร้างแบบจำลองและอธิบายระหว่างการวัดค่าความคล้ายคลึงและฐานกรณี สำหรับองค์ประกอบอธิบายจะวิเคราะห์ฐานกรณีและรวบรวมข้อมูลทางสถิติ การกระจายของค่าในฐานกรณีจะมีประโยชน์และอาจแสดงให้เห็นส่วนของการวัดค่าความคล้ายคลึงที่ในความเป็นจริงไม่เคยได้ใช้ หรือแสดงให้เห็นเป็นกรณีที่เกิดพลาดซึ่งเป็นสิ่งสำคัญสำหรับการยกเว้น

2.4 เอ็นแกรม (Ngram)

เอ็นแกรม คือ แบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของสายอักขระ (character sequence) ที่เกิดขึ้นร่วมกันเป็นคำ หรือค่าความน่าจะเป็นของสายคำ (word sequence) ที่เกิดขึ้นร่วมกันเป็นประโยค โดยค่าความน่าจะเป็นของสายอักขระหรือคำ ประมาณได้จากคลังข้อมูลที่สร้างไว้

แกรม คือ หน่วยที่ใช้ในการสร้างแบบจำลอง อาจจะเป็นเสียง คำ หรือ อักขระก็ได้ และแกรมมีได้หลายขนาดแล้วแต่จะกำหนด ตั้งแต่ 1 จนถึง n

ในแบบจำลองเอ็นแกรมนี้ใช้ความยาวของสายอักขระและสายคำแตกต่างกัน ได้แก่ 2-แกรม 3-แกรม 4-แกรม ฯลฯ ถ้าจะประมาณค่าความน่าจะเป็นของสายคำหรือสายอักขระจากคลังข้อมูลโดยใช้วิธีเอ็นแกรม ผลที่ได้มีดังนี้

การประมาณค่าด้วย 2-แกรม (Probability bigram) คือ การประมาณค่าความน่าจะเป็นของสายอักขระ (สายคำ) ที่เกิดขึ้นร่วมกันว่ามีค่าเท่ากับผลคูณของความน่าจะเป็นที่จะพบอักขระ (คำ) ที่ละ 2 ตัว (คำ) ติดกันในสายอักขระ (สายคำ) นั้น

การประมาณค่าด้วย 3-แกรม (Probability trigram) คือ การประมาณค่าความน่าจะเป็น ของสายอักขระ (สายคำ) ที่เกิดขึ้นร่วมกันว่ามีค่าเท่ากับผลคูณของความน่าจะเป็นที่จะพบอักขระ (คำ) ทีละ 3 ตัว (คำ) ติดกันในสายอักขระ (สายคำ) นั้น

การประมาณค่าด้วย 4-แกรม (Probability quadigram) คือ การประมาณค่าความน่าจะเป็น ของสายอักขระ (สายคำ) ที่เกิดขึ้นร่วมกันว่ามีค่าเท่ากับผลคูณของความน่าจะเป็นที่จะพบอักขระ (คำ) ทีละ 4 ตัว (คำ) ติดกันในสายอักขระ (สายคำ) นั้น

หรืออาจประมาณค่าความน่าจะเป็นจากความยาวของเอ็นแกรมมากกว่า 4-แกรม ก็ได้ ขึ้นอยู่กับความจำเป็นในการทดลอง แต่ระบบของเอ็นแกรมก็ยิ่งซับซ้อนมากขึ้นตามลำดับ

การประมาณค่าความน่าจะเป็นของสายอักขระ โดยการใช้เอ็นแกรมดังที่กล่าวมา คือ การใช้สมมติฐานของมาร์คอฟ (Markov assumption) ว่า การปรากฏของตัวอักษรตัวหนึ่งขึ้นกับตัวอักษรก่อนหน้าเพียง $n-1$ ตัว เราจึงเรียกแบบจำลองลักษณะนี้ว่าเป็นแบบจำลองเอ็นแกรม (n -gram model) หรือแบบจำลองมาร์คอฟ ซึ่งวิธีนี้มักนิยมใช้ในงานระบุภาษาของข้อความกันมาก เนื่องจากสามารถใช้เพื่อระบุภาษาได้อย่างมีประสิทธิภาพและเรียบง่ายกว่า โดยสามารถประมาณได้ดังนี้

$$\text{ไบแกรม } P(c_1c_2c_3\dots c_n) = P(c_1) P(c_2|c_1) P(c_3|c_2) \dots P(c_n|c_{n-1})$$

$$\text{ไตรแกรม } P(c_1c_2c_3\dots c_n) = P(c_1) P(c_2|c_1) P(c_3|c_1c_2) \dots P(c_n|c_{n-2}c_{n-1})$$

$$\text{ควอดริแกรม } P(c_1c_2c_3\dots c_n) = P(c_1) P(c_2|c_1) P(c_3|c_1c_2) P(c_4|c_1c_2c_3) \dots P(c_n|c_{n-3}c_{n-2}c_{n-1})$$

หมายเหตุ: P แทน ค่าความน่าจะเป็น, c แทน อักขระหรือตัวอักษร และ $(c_1c_2c_3\dots c_n)$ แทน สายอักขระที่ประกอบด้วยอักขระตั้งแต่ 3 ตัวขึ้นไปจนถึง n ตัว

ส่วนความน่าจะเป็นของสายคำที่รวมกันเป็นประโยค $w_1w_2w_3\dots w_n$ หากประมาณค่าด้วยเอ็นแกรมต่างๆ ผลที่ได้ดังนี้ โดย w แทน คำ n แทน จำนวนนับต่อไป P แทน ค่าความน่าจะเป็น และ $(w_1w_2w_3\dots w_n)$ แทน สายคำที่ประกอบด้วยคำมากกว่า 3 คำขึ้นไป

- ความน่าจะเป็นของประโยคโดยใช้วิธี 2-แกรม คือ

$$P(w_1w_2w_3\dots w_n) = P(w_1) P(w_2|w_1) P(w_3|w_2) \dots P(w_n|w_{n-1})$$

- ความน่าจะเป็นของประโยคโดยใช้วิธี 3-แกรม คือ

$$P(w_1w_2w_3\dots w_n) = P(w_1) P(w_2|w_1) P(w_3|w_1w_2) \dots P(w_n|w_{n-2}w_{n-1})$$

- ความน่าจะเป็นของประโยคโดยใช้วิธี 4-แกรม คือ

$$P(w_1w_2w_3\dots w_n) = P(w_1) P(w_2|w_1) P(w_3|w_1w_2) P(w_4|w_1w_2w_3) \dots P(w_n|w_{n-3}w_{n-2}w_{n-1})$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง การประมาณค่าความน่าจะเป็นของประโยค “He like to eat”

ใช้สมการประมาณค่าความน่าจะเป็นพื้นฐานดังนี้

$$P(w_1 w_2 K w_r) = \prod_{i=1}^T P(w_i | w_1 K w_{i-1})$$

โดยที่ P คือ ค่าความน่าจะเป็น (Probability) ซึ่งประมาณได้จากคลังข้อมูล

T คือ จำนวนของคำ

i คือ ลำดับของคำโดยเริ่มต้นที่ลำดับที่ 1

$P(w_i | w_1 \dots w_{i-1})$ คือ ความน่าจะเป็นของคำ w_i หลังจากเกิดคำ $w_1 w_2 \dots w_{i-1}$ ก่อนหน้านี้อันแล้ว ความน่าจะเป็นของประโยคนี้ $P(w_1 w_2 \dots w_r)$ สามารถประมาณได้ โดยถือว่าการปรากฏของคำ w_i นั้นขึ้นอยู่กับจำนวนคำข้างหน้า $n-1$ ตัวเท่านั้นหรือขึ้นอยู่กับขนาดของเอ็นแกรม ดังนั้นถ้าหากประมาณค่าความน่าจะเป็นของประโยคนี้ โดยใช้ 2-แกรม จะปรับเปลี่ยนสมการดังนี้

$$P(w_1 w_2 K w_r) = P(w_1 | < s >) P(w_2 | w_1) K P(w_i | w_{i-1})$$

$P(w_1 | < s >)$ หมายถึง ความน่าจะเป็นของคำที่หนึ่งเมื่อเกิดเป็นคำแรกของประโยค ซึ่งในที่นี้คือช่องว่าง

$P(w_2 | w_1)$ หมายถึง ความน่าจะเป็นของคำ w_2 หลังจากเกิดคำ w_1

$P(w_i | w_{i-1})$ หมายถึง ความน่าจะเป็นของคำ w_i หลังจากเกิดคำ w_{i-1}

ดังนั้น จากสูตรประมาณค่าความน่าจะเป็นด้วย 2-แกรม หากต้องการหาความน่าจะเป็นของประโยค “He like to eat” โดยใช้เอ็นแกรมในระดับของคำ ผลออกมาคือ $P(\text{He} | < s >) P(\text{like} | \text{He}) P(\text{to} | \text{like}) P(\text{eat} | \text{to})$

ส่วนค่าความน่าจะเป็นจะหาได้จากคลังข้อมูล เช่น

$$P(\text{like} | \text{He}) = \frac{c(\text{He-like})}{c(\text{He})}$$

จากสูตรนี้หมายถึง ค่าความน่าจะเป็นของ like เมื่อเกิดร่วมกับคำว่า He จำนวนได้จากนำจำนวนนับของ He ที่เกิดร่วมกับคำว่า like หารด้วยจำนวนนับของการเกิด He เดี่ยวๆ

วิธีการใช้แบบจำลองเอ็นแกรมนี้เป็นวิธีทางสถิติที่นิยมใช้กันมากที่สุด เพราะเป็นวิธีที่เรียบง่าย มีประสิทธิภาพสูง และเหมาะสำหรับวิเคราะห์ภาษา สามารถใช้ระบุภาษาได้ดีกว่าวิธีอื่นๆ

ปี พ.ศ. 2548 อัครพล ได้ยกตัวอย่างงานวิจัยที่ใช้แบบจำลองแบบเอ็นแกรม ในวิทยานิพนธ์การระบุคำไทยและคำทับศัพท์ด้วยแบบจำลองเอ็นแกรม ได้แสดงให้เห็นว่าประสิทธิภาพของเอ็นแกรมที่ดีจะอยู่ที่ 3 แกรม ซึ่งเมื่อทำการเพิ่มจำนวนเอ็นแกรม จะทำให้เกิดความซับซ้อนมากยิ่งขึ้นและอาจเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำให้เกิดปัญหาข้อมูลในคลังข้อมูลน้อยเกินไป (sparse data problem) เมื่อเพิ่ม n ให้สูงขึ้น เนื่องจากมีโอกาสสูงขึ้นที่จะหาข้อมูลในคลังข้อมูลไม่พบมากขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบการให้เหตุผลเชิงกรณี

การออกแบบการให้เหตุผลเชิงกรณีเพื่อใช้ในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดของแผนกคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ มีขั้นตอนดังนี้

3.1 รวบรวมข้อมูล

การรวบรวมข้อมูลเป็นการรวบรวมข้อมูลเพื่อนำมาสร้างเป็นฐานความรู้ ซึ่งในโครงการนี้ ได้ใช้ข้อมูลจากเอกสารอิเล็กทรอนิกส์ที่บันทึกรายละเอียดของกระบวนการอีทีแอลที่ผิดพลาดโดยพนักงานแผนกคลังข้อมูลขององค์กร ในช่วงระยะเวลาตั้งแต่ 1 พฤษภาคม 2555 บางส่วนมาสร้างเป็นฐานความรู้ โดยที่ข้อมูลรายละเอียดของกระบวนการอีทีแอลที่ผิดพลาดมีดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดของกระบวนการอีทีแอลที่ผิดพลาด

Error Date	วันที่ทำงานผิดพลาด
Business Date	วันที่ข้อมูลระบบ
Job Type	ประเภทของงาน
System Name	ชื่อระบบงาน
Job Name	ชื่องาน
Sub Task Name	ชื่องานย่อย
Error Message	ข้อความผิดพลาด
Error Code	รหัสผิดพลาด
Problem	ปัญหา
File / Table / View	ไฟล์, ตาราง และวิวฐานข้อมูล ที่ผิดพลาด
Field	คอลัมน์ที่ผิดพลาด
Standby L1	ชื่อพนักงานคอยแก้ปัญหาระดับหนึ่ง
Standby L2	ชื่อพนักงานคอยแก้ปัญหาระดับสอง
Solution	แนวทางการแก้ไขปัญหา
Action	สิ่งที่ต้องทำหลังจากหาสาเหตุของปัญหาพบ
Status	สถานะของงาน
Closed Date	วันที่ปิดสถานะของงาน
Note	หมายเหตุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การออกแบบคลังกรณี

การออกแบบคลังกรณี หรืออาจเรียกว่าการวิเคราะห์ฐานกรณี (Case-Based Diagnosis) คือ การตัดสินใจว่าคุณลักษณะใดของปัญหามีความสำคัญต่อผลลัพธ์หรือคำตอบที่ได้เพื่อจัดเก็บลงในฐานความรู้ ซึ่งรายละเอียดของกรณีหรือฐานกรณีอธิบายได้ดังต่อไปนี้และแสดงดังรูปที่ 3.1



1. ส่วนคำอธิบายปัญหา (Problem Description) แสดงคุณลักษณะสำคัญของปัญหาของกระบวนการอีทีแอลที่ผิดพลาดประกอบไปด้วยรายละเอียดของตารางที่ 3.1 บางส่วนที่มีความสำคัญต่อผลลัพธ์หรือคำตอบที่ได้ดังนี้ ชื่อระบบงานคือชื่อระบบของกระบวนการอีทีแอลที่ผิดพลาดชื่องานคือชื่อของกระบวนการอีทีแอลที่ผิดพลาดประเภทของงานคือประเภทของกระบวนการอีทีแอลที่ผิดพลาด ซึ่งมีด้วยกันอยู่ 6 ประเภทคือ โหลด (Load) แปลง (Transform) บีคีย์ (Bkey) บีแมป (Bmap) ส่งออก (Export) และเปลี่ยนวันที่ (SetDate)ข้อความผิดพลาดคือข้อความที่กระบวนการอีทีแอลแสดงเมื่อเกิดความผิดพลาดรหัสผิดพลาดคือรหัสที่กระบวนการอีทีแอลแสดงเมื่อเกิดความผิดพลาด และชื่องานย่อยคือชื่อขั้นตอนงานย่อยภายในกระบวนการอีทีแอลที่เกิดความผิดพลาด

2. ส่วนคำตอบ (Solution) แสดงแนวทางที่นำมาใช้ในการแก้ปัญหาโดยประกอบไปด้วยรายละเอียดของตารางที่ 3.1 บางส่วนที่เป็นผลลัพธ์หรือคำตอบดังนี้ ปัญหาคือผลที่เกิดขึ้นเมื่อกระบวนการอีทีแอลผิดพลาดตามข้อผิดพลาดนั้นๆ แนวทางการแก้ไขปัญหาคือวิธีการหาสาเหตุหรือวิธีการแก้ไขปัญหา และสิ่งที่ต้องทำหลังจากหาสาเหตุของปัญหาพบ

กรณี (Case) : EDW_ETL_JOB
ส่วนคำอธิบายปัญหา (Problem Description) >ชื่อระบบงาน (System Name) : ALM >ชื่องาน (Job Name) : SCN_L_ALM_RCPARMET >ประเภทของงาน (Job Type) : Load >ข้อความผิดพลาด (Error Message) : Load Error >รหัสผิดพลาด (Error Code) : 2633 >ชื่องานย่อย (Sub Task Name) : Run_MultiLoad
ส่วนคำตอบ (Solution) >ปัญหาหรือคำวินิจฉัย (Problem/Diagnosis) : Server too many load/unload tasks running >แนวทางการแก้ไขปัญหาหรือข้อเสนอแนะ (Solution/Suggestion) : Wait a moment & rerun job >สิ่งที่ต้องทำหลังจากหาสาเหตุของปัญหาพบ (Action) : Call operation term to rerun job.


รูปที่ 3.1 ตัวอย่างกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสร้างฐานกรณีโดยใช้โปรแกรมโททิจและมายซีบีอาร์จะใช้น้ำต่างแท็บคลาสที่ได้กล่าวไว้ในหัวข้อที่ 2.3.4 ในการสร้างคลาสหรือฐานกรณีขึ้นมาโดยใช้เครื่องมือ  เป็นเครื่องมือสร้างคลาส หลังจากนั้นทำการสร้างแอตทริบิวต์หรือสล็อต โดยใช้เครื่องมือ  เป็นเครื่องมือสร้างสล็อต โดยมีการกำหนดจำนวนสมาชิกของสล็อต (Cardinality) ประเภทของสล็อตด้วย ส่วนการตั้งชื่อสล็อต มีการขึ้นต้นด้วย พี(p)ตามด้วยตัวเลขลำดับของสล็อต เพื่อบ่งบอกว่าสล็อตนั้นเป็นส่วนคำอธิบายปัญหา และขึ้นต้นด้วย เอส(s) ตามด้วยตัวเลขลำดับของสล็อต เพื่อบ่งบอกว่าสล็อตนั้นเป็นส่วนคำตอบ แสดงดังรูปที่ 3.2

CLASS EDITOR

For Class: ● EDW_ETL_JOB (Instance of :STANDARD-CLASS)

Name: EDW_ETL_JOB Documentation: ETL Job of SCBEDW Constraints: 

Role: Concrete ●

Template Slots

Name	Cardinality	Type	Other Facets
■ P00_System_Name	single	String	
■ P01_Job_Name	single	String	
■ P02_Job_Type	required s...	Symbol	load,Transform,Bkey,Bmap,Export,SetC...
■ P03_Error_Message	single	String	
■ P04_Error_Code	required s...	Symbol	unknown,2794,UTY4016,2793,6706,26...
■ P05_Sub_Task_Name	required s...	Symbol	allowed-values=(Read_Control_File_Info_&_Updat_CSFE...
■ S01_Diagnosis	single	String	
■ S02_Suggestion	single	String	
■ S03_Action	single	String	

ส่วนคำอธิบายปัญหา (Problem Description)

ส่วนคำตอบ (Solution)

รูปที่ 3.2 ฐานกรณีที่สร้างโดยโททิจและมายซีบีอาร์

3.3 การค้นคืนกรณี

การค้นคืนกรณีเป็นการค้นคืนกรณีเก่าที่มีความคล้ายคลึงกับปัญหาใหม่ และเคยประสบความสำเร็จในการแก้ปัญหามาแล้ว โดยแบ่งขั้นตอนวัดค่าความคล้ายคลึงเป็น 2 ระดับดังนี้

1. การวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity)

ขั้นตอนแรกของการค้นคืนกรณีเป็นการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (feature) ของปัญหาใหม่กับกรณีเก่าในคลังกรณี และเป็นการคำนวณค่า $sim(a,b_i)$ จากสมการในรูป 2.3 สามารถอธิบายได้จากข้อมูลตัวอย่างดังรูป 3.3

ปัญหาใหม่หรือข้อมูลสอบถาม (New Problem or Query)	
คำอธิบายงานอีทีแอลที่ผิดพลาด >ชื่อระบบงาน : ESL >ชื่องาน : SCN_T_ESL_EDW_EVENT_TR >ประเภทของงาน : Transform >ข้อความผิดพลาด : Null Pointer >รหัสผิดพลาด : Unknown >ชื่องานย่อย : Assigning_Data	
ฐานกรณีที่ 1 (Case1)	ฐานกรณีที่ 2 (Case2)
คำอธิบายงานอีทีแอลที่ผิดพลาด >ชื่อระบบงาน : TF >ชื่องาน : SCN_T_TF_EVENTPARTYPA_01 >ประเภทของงาน : Transform >ข้อความผิดพลาด : Null Pointer >รหัสผิดพลาด : Unknown >ชื่องานย่อย : Assigning_Data	คำอธิบายงานอีทีแอลที่ผิดพลาด >ชื่อระบบงาน : TF >ชื่องาน : SCN_T_TF_IMCOFL_1_0_M09700 >ประเภทของงาน : Bmap >ข้อความผิดพลาด : NameError: ChkCnt >รหัสผิดพลาด : Unknown >ชื่องานย่อย : Insert_Reference_Code

รูปที่ 3.3 ข้อมูลตัวอย่างฐานกรณี

จากรูปที่ 3.3 การวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (feature) ในมายชีปียาร์ที่เลือกใช้ในแต่ละสล็อตมีดังนี้

- ชื่อระบบงาน (System Name) เป็นประเภทสายอักขระใช้โหมดการวัดความคล้ายคลึงแบบใช้ตัวอักษร (character-based) และฟังก์ชันการวัดค่าความคล้ายคลึงแบบเหมือนทั้งหมด (exact mode) ดังรูปที่ 3.4 เนื่องจากชื่อระบบงานมีลักษณะเป็นตัวอักษรที่บ่งบอกชื่อระบบงานของร้านค้า
- ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

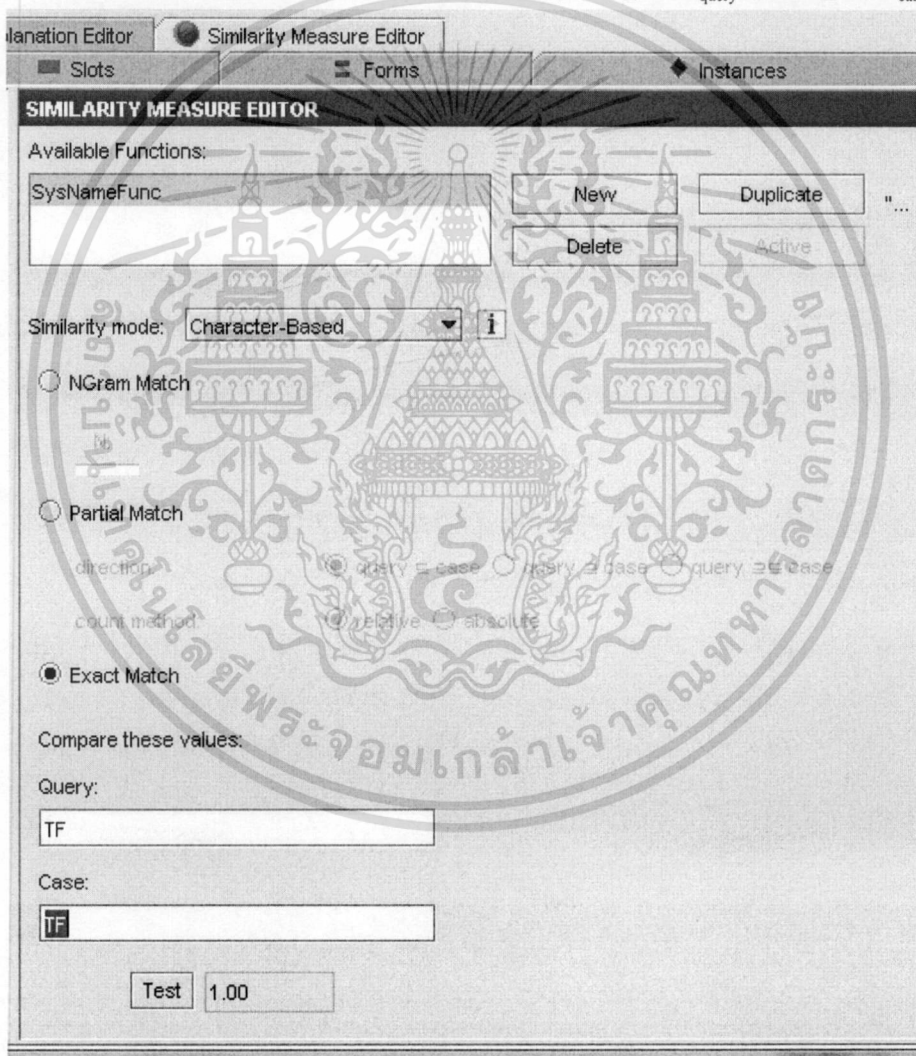
งานอีทีแอลนั้นๆ ดังนั้นการวัดความคล้ายคลึงของชื่อระบบนั้นจะต้องมีตัวอักษรและจำนวนตัวอักษรระหว่างข้อมูลสอบถามและข้อมูลกรณีเหมือนกันทั้งหมดจึงจะถือว่าเป็นระบบเดียวกันซึ่งวัดค่าความคล้ายคลึงได้ดังนี้

$$\text{sim}(\text{System_Name}_{\text{query}}, \text{System_Name}_{\text{case1}}) = \begin{cases} 1 & \text{if } \text{ESL} = \text{TF} \\ 0 & \text{if } \text{ESL} \neq \text{TF} \end{cases} = 0$$

และ

$$\text{sim}(\text{System_Name}_{\text{query}}, \text{System_Name}_{\text{case2}}) = \begin{cases} 1 & \text{if } \text{ESL} = \text{TF} \\ 0 & \text{if } \text{ESL} \neq \text{TF} \end{cases} = 0$$

ค่าความคล้ายคลึงของสตริงชื่อระบบงาน(System Name) ระหว่างข้อมูลสอบถาม(query) คือ ESL และกรณี1 คือ TF มีค่าเป็น $\text{sim}(\text{system_name}_{\text{query}}, \text{system_name}_{\text{case1}}) = 0$ และข้อมูลสอบถาม(query) คือ ESLและกรณี2 คือ TF มีค่าเป็น $\text{sim}(\text{system_name}_{\text{query}}, \text{system_name}_{\text{case2}}) = 0$



รูปที่ 3.4 โหมดการวัดความคล้ายคลึงของสตริงชื่อระบบงาน

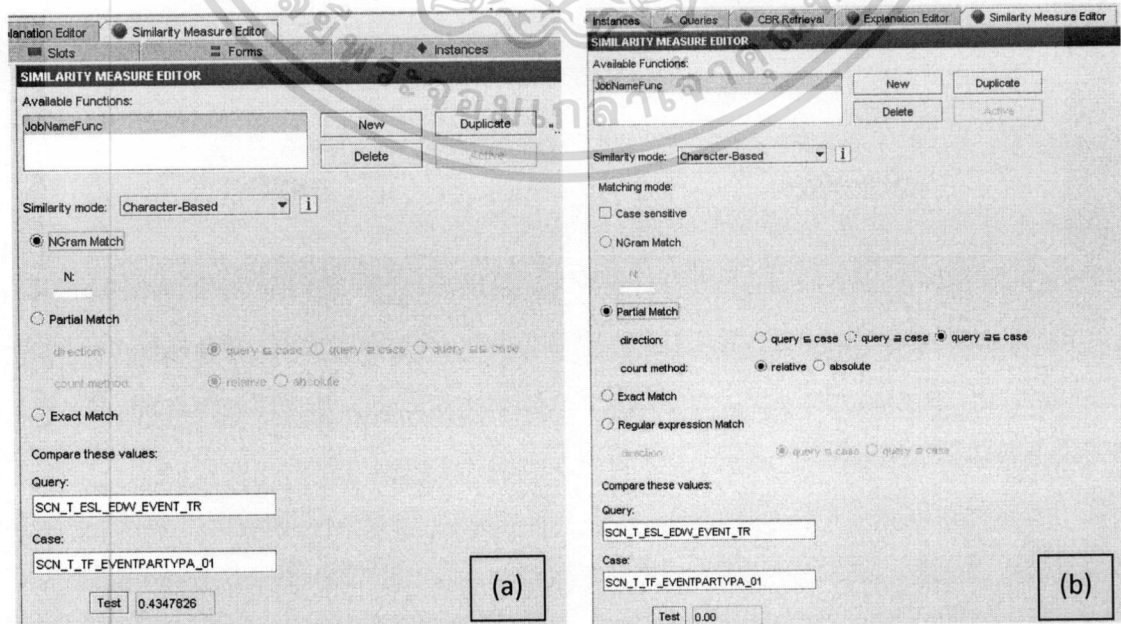
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ชื่องาน (Job Name) เป็นประเภทสายอักขระใช้โหมคการวัดความคล้ายคลึงแบบใช้ตัวอักษร (character-based) และฟังก์ชันการวัดค่าความคล้ายคลึงแบบเอ็นแกรม (ngram) โดยจะใช้ไตรแกรม(trigram matching) ดังรูปที่ 3.5 (a) เนื่องจากข้อมูลชื่องานมีลักษณะเป็นตัวอักษรและมีมาตรฐานการตั้งชื่องานดังนี้

ตารางที่ 3.2 มาตรฐานการตั้งชื่องาน

ประเภทงาน	ขึ้นต้นด้วย	ตัวอย่างข้อมูล
โหลด	SCN_L_(ชื่อระบบงาน)_(ชื่องาน)	SCN_L_TR_EDW_LIMITPTY
แปลง, บีคีย์, บีแมป ส่งออก(รันที่ข้อมูล)	SCN_T_(ชื่อระบบงาน)_(ชื่องาน)	SCN_T_TF_EVENTPARTYPA_01
ส่งออก(ไฟล์)	SCN_E_(ชื่อระบบงาน)_(ชื่องาน)	SCN_E_CRM_TRANS_INC_EXP_M

จากตารางที่ 3.2 จะเห็นได้ว่าความคล้ายคลึงของรูปแบบชื่องานมีผลต่อการวัดค่าความคล้ายคลึงด้วย ดังนั้นฟังก์ชันการวัดค่าความคล้ายคลึงแบบเอ็นแกรมจึงมีความเหมาะสม เนื่องจากเอ็นแกรมนอกจากใช้คำนวณความน่าจะเป็นของสายอักขระ (character sequence) ที่เกิดขึ้นแล้วยังสามารถใช้ในการประมาณค่าความใกล้เคียงกันของสายอักขระได้ด้วย โดยการแปลงลำดับของอักขระเป็นเซตของ เอ็นแกรม (n-gram) โดยแบ่งลำดับของอักขระเพื่อเปรียบเทียบกับลำดับของอักขระอื่นๆ ทำให้มีประสิทธิภาพมากขึ้นในการประมาณค่าความใกล้เคียงของสายอักขระและมีประสิทธิภาพมากกว่าฟังก์ชันการวัดค่าความคล้ายคลึงแบบเหมือนบางส่วนที่คำนวณค่าความใกล้เคียงกันของสายอักขระได้ 0 ตามรูปที่ 3.5 (b)



เอกสารรูปที่ 3.5 โหมคการวัดค่าความคล้ายคลึงของสตริงชื่องาน (a) แบบเอ็นแกรม; (b) แบบเหมือนบางส่วน
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความคล้ายคลึงของสล็อตชื่องาน(Job Name) ระหว่างข้อมูลสอบถาม(query) คือ SCN_T_ESL_EDW_EVENT_TR และกรณี1 คือ SCN_T_TF_EVENTPARTYPA_01 มีค่าเป็น $\text{sim}(\text{Job_Name}_{\text{query}}, \text{Job_Name}_{\text{case1}}) = 0.4347826$ และข้อมูลสอบถาม(query) คือ SCN_T_ESL_EDW_EVENT_TR และกรณี2 คือ SCN_T_TF_IMCOFL_1_0_M09700 มีค่าเป็น $\text{sim}(\text{Job_Name}_{\text{query}}, \text{Job_Name}_{\text{case2}}) = 0.25$

จากการวัดค่าความคล้ายคลึงแบบไทรแกรมสามารถอ้างอิงจากซอร์สโค้ดของมายซีบีอาร์ เวอร์ชัน 2.6.6 ดังต่อไปนี้

```
/**
 * คำนวณค่าความคล้ายคลึงแบบไทรแกรมระหว่าง 2 สายอักขระ (String)
 *
 * @param s1 String string1
 * @param s2 String string2
 * @return double between 0 and 1.
 */
public static double ngram (String s1, String s2, int n,
boolean caseSensitive) {
    //กรณี No Case Sensitive ทำการแปลง 2 สายอักขระ เป็นตัวอักษรตัวเล็ก
    if(!caseSensitive) {
        s1 = s1.toLowerCase();
        s2 = s2.toLowerCase();
    }
    //ทำการแบ่งลำดับของอักขระเป็นเซตของแกรมในที่นี้จะเท่ากับ 3 แกรม
    ArrayList<String> triples_s1 = getTuples(s1, 3);
    ArrayList<String> triples_s2 = getTuples(s2, 3);

    int matches = 0;
    //ทำการวนรอบเพื่อนับจำนวนเทอมที่เหมือนกันของการเปรียบเทียบระหว่าง สายอักขระ s1 กับ สาย
    อักขระ s2
    for (int i=0; i<triples_s1.size(); i++) {
        for (int j=0; j<triples_s2.size(); j++) {
            if (triples_s1.get(i).equals(triples_s2.get(j))) {
                matches++;
            }
        }
    }
    //คำนวณค่าความคล้ายคลึง
    double sim =
    ((double)2*matches)/((double)(triples_s1.size()+triples_s2.size()));
    return sim;
}
```

```

/**
 * ทำการแบ่งลำดับของอักขระเป็นเซตของแกรม
 * @return ArrayList triples (of type String)
 */
private static ArrayList<String> getTuples(String s, int
length) {
    s = "_" + s.trim() + "_";
    while (s.length() < length) {
        s += "_";
    }

    ArrayList<String> triples = new ArrayList<String>();
    int lastIndex = s.length()-length;
    for (int i=0; i<=lastIndex; i++) {
        String tmp = s.substring(i, i+length);
        if(!triples.contains(tmp)){
            triples.add(s.substring(i, i+length));
        }
    }
    return triples;
}
}

```

ตัวอย่างการวัดค่าความคล้ายคลึงแบบไตรแกรมจากซอร์สโค้ดของมายซีบีอาร์ เวอร์ชัน

2.6.6

วัดค่าความคล้ายคลึงแบบไตรแกรมระหว่างชื่องาน

- CCAP_EDW_CCAC กำหนดให้เป็น String1

- CCAP_EDW_SYSCHK กำหนดให้เป็น String2

สามารถแสดงได้ดังนี้

1. ทำการแบ่งลำดับของอักขระเป็นเซตของแกรมในที่นี้จะเท่ากับ 3 แกรมจาก String1 และ String2 จะได้เซตของ 3 แกรมดังนี้

- String1 = { _CC, CCA, CAP, AP_, P_E, _ED, EDW, DW_, W_C, CAC, AC_ }

- String2 = { _CC, CCA, CAP, AP_, P_E, _ED, EDW, DW_, W_S, _SY, SYS, YSC, SCH, CHK, HK_ }

2. เปรียบเทียบ 3 แกรมของ String1 กับ 3 แกรมของ String2 โดยนับจำนวนเทอมที่เหมือนกันดังต่อไปนี้

ตารางที่ 3.3 ตารางเปรียบเทียบ 3 แกรมที่ String1 เหมือนกับ String2

(_cc , _cc) = 1	(cap , ap_) = 0	(p_e , edw) = 0	(edw , _sy) = 0	(w_c , sch) = 0
(_cc , cca) = 0	(cap , p_e) = 0	(p_e , dw_) = 0	(edw , sys) = 0	(w_c , chk) = 0
(_cc , cap) = 0	(cap , _ed) = 0	(p_e , w_s) = 0	(edw , ysc) = 0	(w_c , hk_) = 0
(_cc , ap_) = 0	(cap , edw) = 0	(p_e , _sy) = 0	(edw , sch) = 0	(cac , _cc) = 0
(_cc , p_e) = 0	(cap , dw_) = 0	(p_e , sys) = 0	(edw , chk) = 0	(cac , cca) = 0
(_cc , _ed) = 0	(cap , w_s) = 0	(p_e , ysc) = 0	(edw , hk_) = 0	(cac , cap) = 0
(_cc , edw) = 0	(cap , _sy) = 0	(p_e , sch) = 0	(dw_ , _cc) = 0	(cac , ap_) = 0
(_cc , dw_) = 0	(cap , sys) = 0	(p_e , chk) = 0	(dw_ , cca) = 0	(cac , p_e) = 0
(_cc , w_s) = 0	(cap , ysc) = 0	(p_e , hk_) = 0	(dw_ , cap) = 0	(cac , _ed) = 0
(_cc , _sy) = 0	(cap , sch) = 0	(_ed , _cc) = 0	(dw_ , ap_) = 0	(cac , edw) = 0
(_cc , sys) = 0	(cap , chk) = 0	(_ed , cca) = 0	(dw_ , p_e) = 0	(cac , dw_) = 0
(_cc , ysc) = 0	(cap , hk_) = 0	(_ed , cap) = 0	(dw_ , _ed) = 0	(cac , w_s) = 0
(_cc , sch) = 0	(ap_ , _cc) = 0	(_ed , ap_) = 0	(dw_ , edw) = 0	(cac , _sy) = 0
(_cc , chk) = 0	(ap_ , cca) = 0	(_ed , p_e) = 0	(dw_ , dw_) = 1	(cac , sys) = 0
(_cc , hk_) = 0	(ap_ , cap) = 0	(_ed , _ed) = 1	(dw_ , w_s) = 0	(cac , ysc) = 0
(cca , _cc) = 0	(ap_ , ap_) = 1	(_ed , edw) = 0	(dw_ , _sy) = 0	(cac , sch) = 0
(cca , cca) = 1	(ap_ , p_e) = 0	(_ed , dw_) = 0	(dw_ , sys) = 0	(cac , chk) = 0
(cca , cap) = 0	(ap_ , _ed) = 0	(_ed , w_s) = 0	(dw_ , ysc) = 0	(cac , hk_) = 0
(cca , ap_) = 0	(ap_ , edw) = 0	(_ed , _sy) = 0	(dw_ , sch) = 0	(ac_ , _cc) = 0
(cca , p_e) = 0	(ap_ , dw_) = 0	(_ed , sys) = 0	(dw_ , chk) = 0	(ac_ , cca) = 0
(cca , _ed) = 0	(ap_ , w_s) = 0	(_ed , ysc) = 0	(dw_ , hk_) = 0	(ac_ , cap) = 0
(cca , edw) = 0	(ap_ , _sy) = 0	(_ed , sch) = 0	(w_c , _cc) = 0	(ac_ , ap_) = 0
(cca , dw_) = 0	(ap_ , sys) = 0	(_ed , chk) = 0	(w_c , cca) = 0	(ac_ , p_e) = 0
(cca , w_s) = 0	(ap_ , ysc) = 0	(_ed , hk_) = 0	(w_c , cap) = 0	(ac_ , _ed) = 0
(cca , _sy) = 0	(ap_ , sch) = 0	(edw , _cc) = 0	(w_c , ap_) = 0	(ac_ , edw) = 0
(cca , sys) = 0	(ap_ , chk) = 0	(edw , cca) = 0	(w_c , p_e) = 0	(ac_ , dw_) = 0
(cca , ysc) = 0	(ap_ , hk_) = 0	(edw , cap) = 0	(w_c , _ed) = 0	(ac_ , w_s) = 0
(cca , sch) = 0	(p_e , _cc) = 0	(edw , ap_) = 0	(w_c , edw) = 0	(ac_ , _sy) = 0
(cca , chk) = 0	(p_e , cca) = 0	(edw , p_e) = 0	(w_c , dw_) = 0	(ac_ , sys) = 0
(cca , hk_) = 0	(p_e , cap) = 0	(edw , _ed) = 0	(w_c , w_s) = 0	(ac_ , ysc) = 0
(cap , _cc) = 0	(p_e , ap_) = 0	(edw , edw) = 1	(w_c , _sy) = 0	(ac_ , sch) = 0
(cap , cca) = 0	(p_e , p_e) = 1	(edw , dw_) = 0	(w_c , sys) = 0	(ac_ , chk) = 0
(cap , cap) = 1	(p_e , _ed) = 0	(edw , w_s) = 0	(w_c , ysc) = 0	(ac_ , hk_) = 0

String2 มีค่าความคล้ายคลึงแบบไตรแกรมกับ String1

$$= (2 * \text{จำนวนที่เหมือนกัน}) / (\text{จำนวนแกรม String1} + \text{จำนวนแกรม String2})$$

$$= (2 * 8) / (11 + 15) = 0.6154$$

เปรียบเทียบ 3 แกรมของ String2 กับ 3 แกรมของ String1 โดยนับจำนวนทอมที่เหมือนกันดังตารางที่ 3.4 เพื่อทดสอบค่าความคล้ายคลึงว่าเท่ากับการเปรียบเทียบ 3 แกรมของ String1 กับ 3 แกรมของ String2 หรือไม่

ตารางที่ 3.4 ตารางเปรียบเทียบ 3 แกรมที่ String2 เหมือนกับ String1

(_cc , _cc) = 1	(ap_ , _cc) = 0	(edw , _cc) = 0	(_sy , _cc) = 0	(sch , _cc) = 0
(_cc , cca) = 0	(ap_ , cca) = 0	(edw , cca) = 0	(_sy , cca) = 0	(sch , cca) = 0
(_cc , cap) = 0	(ap_ , cap) = 0	(edw , cap) = 0	(_sy , cap) = 0	(sch , cap) = 0
(_cc , ap_) = 0	(ap_ , ap_) = 1	(edw , ap_) = 0	(_sy , ap_) = 0	(sch , ap_) = 0
(_cc , p_e) = 0	(ap_ , p_e) = 0	(edw , p_e) = 0	(_sy , p_e) = 0	(sch , p_e) = 0
(_cc , _ed) = 0	(ap_ , _ed) = 0	(edw , _ed) = 0	(_sy , _ed) = 0	(sch , _ed) = 0
(_cc , edw) = 0	(ap_ , edw) = 0	(edw , edw) = 1	(_sy , edw) = 0	(sch , edw) = 0
(_cc , dw_) = 0	(ap_ , dw_) = 0	(edw , dw_) = 0	(_sy , dw_) = 0	(sch , dw_) = 0
(_cc , w_c) = 0	(ap_ , w_c) = 0	(edw , w_c) = 0	(_sy , w_c) = 0	(sch , w_c) = 0
(_cc , cac) = 0	(ap_ , cac) = 0	(edw , cac) = 0	(_sy , cac) = 0	(sch , cac) = 0
(_cc , ac_) = 0	(ap_ , ac_) = 0	(edw , ac_) = 0	(_sy , ac_) = 0	(sch , ac_) = 0
(cca , _cc) = 0	(p_e , _cc) = 0	(dw_ , _cc) = 0	(sys , _cc) = 0	(chk , _cc) = 0
(cca , cca) = 1	(p_e , cca) = 0	(dw_ , cca) = 0	(sys , cca) = 0	(chk , cca) = 0
(cca , cap) = 0	(p_e , cap) = 0	(dw_ , cap) = 0	(sys , cap) = 0	(chk , cap) = 0
(cca , ap_) = 0	(p_e , ap_) = 0	(dw_ , ap_) = 0	(sys , ap_) = 0	(chk , ap_) = 0
(cca , p_e) = 0	(p_e , p_e) = 1	(dw_ , p_e) = 0	(sys , p_e) = 0	(chk , p_e) = 0
(cca , _ed) = 0	(p_e , _ed) = 0	(dw_ , _ed) = 0	(sys , _ed) = 0	(chk , _ed) = 0
(cca , edw) = 0	(p_e , edw) = 0	(dw_ , edw) = 0	(sys , edw) = 0	(chk , edw) = 0
(cca , dw_) = 0	(p_e , dw_) = 0	(dw_ , dw_) = 1	(sys , dw_) = 0	(chk , dw_) = 0
(cca , w_c) = 0	(p_e , w_c) = 0	(dw_ , w_c) = 0	(sys , w_c) = 0	(chk , w_c) = 0
(cca , cac) = 0	(p_e , cac) = 0	(dw_ , cac) = 0	(sys , cac) = 0	(chk , cac) = 0
(cca , ac_) = 0	(p_e , ac_) = 0	(dw_ , ac_) = 0	(sys , ac_) = 0	(chk , ac_) = 0
(cap , _cc) = 0	(_ed , _cc) = 0	(w_s , _cc) = 0	(ysc , _cc) = 0	(hk_ , _cc) = 0
(cap , cca) = 0	(_ed , cca) = 0	(w_s , cca) = 0	(ysc , cca) = 0	(hk_ , cca) = 0
(cap , cap) = 1	(_ed , cap) = 0	(w_s , cap) = 0	(ysc , cap) = 0	(hk_ , cap) = 0
(cap , ap_) = 0	(_ed , ap_) = 0	(w_s , ap_) = 0	(ysc , ap_) = 0	(hk_ , ap_) = 0
(cap , p_e) = 0	(_ed , p_e) = 0	(w_s , p_e) = 0	(ysc , p_e) = 0	(hk_ , p_e) = 0

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่โดยไม่ได้รับอนุญาตในการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 (ต่อ)

(cap, _ed) = 0	(_ed, _ed) = 1	(w_s, _ed) = 0	(ysc, _ed) = 0	(hk, _ed) = 0
(cap, edw) = 0	(_ed, edw) = 0	(w_s, edw) = 0	(ysc, edw) = 0	(hk, edw) = 0
(cap, dw_) = 0	(_ed, dw_) = 0	(w_s, dw_) = 0	(ysc, dw_) = 0	(hk, dw_) = 0
(cap, w_c) = 0	(_ed, w_c) = 0	(w_s, w_c) = 0	(ysc, w_c) = 0	(hk, w_c) = 0
(cap, cac) = 0	(_ed, cac) = 0	(w_s, cac) = 0	(ysc, cac) = 0	(hk, cac) = 0
(cap, ac_) = 0	(_ed, ac_) = 0	(w_s, ac_) = 0	(ysc, ac_) = 0	(hk, ac_) = 0

String1 มีค่าความคล้ายคลึงแบบไตรแกรมกับ String2

$$= (2 * \text{จำนวนที่เหมือนกัน}) / (\text{จำนวนแกรม Sting1} + \text{จำนวนแกรม String2})$$

$$= (2 * 8) / (15 + 11) = 0.6154$$

จากตารางที่ 3.3 และ ตารางที่ 3.4 จะเห็นได้ว่า ค่าความคล้ายคลึงของการเปรียบเทียบ 3 แกรมของ String1 กับ 3 แกรมของ String2 และ ค่าความคล้ายคลึงของการเปรียบเทียบ 3 แกรมของ String2 กับ 3 แกรมของ String1 นั้นมีค่าเท่ากัน

- **ประเภทของงาน (Job Type)** เป็นประเภทสัญลักษณ์ใช้โหมมการวัดความคล้ายคลึงแบบตารางและกำหนดค่าความคล้ายคลึงของแต่ละประเภทของงานลงในตาราง ดังรูปที่ 3.6 และ ตารางที่ 3.5

ตารางที่ 3.5 ค่าความคล้ายคลึงประเภทของงาน

ประเภทงาน	Transform	Bkey	Bmap	Load	Export	SetDate
Load	0.0	0.0	0.0	1.0	0.0	0.0
Transform	1.0	0.3	0.3	0.0	0.0	0.0
Bkey	0.3	1.0	0.6	0.0	0.0	0.0
Bmap	0.3	0.6	1.0	0.0	0.0	0.0
Export	0.0	0.0	0.0	0.0	1.0	0.0
SetDate	0.0	0.0	0.0	0.0	0.0	1.0

จากตารางที่ 3.5 เป็นตารางค่าความคล้ายคลึงของประเภทงานแต่ละคู่ ซึ่งค่าความคล้ายคลึงของแต่ละประเภทจะถูกกำหนดโดยผู้เชี่ยวชาญตามความเหมาะสมค่าความคล้ายคลึงของสล็อตประเภทของงาน (Job Type) ระหว่างข้อมูลสอบถาม(query) คือ Transform และกรณี 1 คือ Transform

เอกสารนี้เป็นลิขสิทธิ์ของกรมการขนส่งทางบก ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มีค่าเป็น $\text{sim}(\text{Job_Type}_{\text{query}}, \text{Job_Type}_{\text{case1}}) = 1$ และระหว่างข้อมูลสอบถาม(query) คือ Transform และกรณี2 คือ Bmap มีค่าเป็น $\text{sim}(\text{Job_Type}_{\text{query}}, \text{Job_Type}_{\text{case2}}) = 0.3$

The screenshot shows the 'SIMILARITY MEASURE EDITOR' window. It includes a 'JobTypeFunc' field, 'New', 'Duplicate', 'Delete', and 'Apply' buttons. Below, there are 'Similarity mode' (set to 'Table') and 'Symmetry' options (symmetric and asymmetric). The main part is a table titled 'Case Base Values' with columns for 'Reset', 'Transform', 'Bkey', 'Bmap', 'Load', 'Export', and 'SetDate'. The rows represent different job types with their respective percentages and similarity values.

	Reset	Transform	Bkey	Bmap	Load	Export	SetDate
Load	53%	0.0	0.0	0.0	1.0	0.0	0.0
Transform	16%	1.0	0.3	0.3	0.0	0.0	0.0
Bkey	0%	0.3	1.0	0.6	0.0	0.0	0.0
Bmap	6%	0.3	0.6	1.0	0.0	0.0	0.0
Export	20%	0.0	0.0	0.0	0.0	1.0	0.0
SetDate	3%	0.0	0.0	0.0	0.0	0.0	1.0

รูปที่ 3.6 โหมดการวัดความคล้ายคลึงของสล็อตประเภทของงาน

- ข้อความผิดพลาด (Error Message) เป็นประเภทสายอักขระ ใช้โหมดการวัดความคล้ายคลึงแบบใช้ตัวอักษร (character-based) และฟังก์ชันการวัดค่าความคล้ายคลึง แบบเหมือนทั้งหมด (exact mode) ดังรูปที่ 3.7 เนื่องจากข้อมูลข้อความผิดพลาดมีลักษณะเป็นตัวอักษรที่บ่งบอกข้อผิดพลาดของงานอีทีแอลอื่นๆ ดังนั้นการวัดความคล้ายคลึงของข้อความผิดพลาดระหว่างข้อมูลสอบถามและข้อมูลกรณีเหมือนกันทั้งหมดจึงจะถือว่าเป็นข้อผิดพลาดของงานอีทีแอลเดียวกันซึ่งวัดค่าความคล้ายคลึงได้ดังนี้

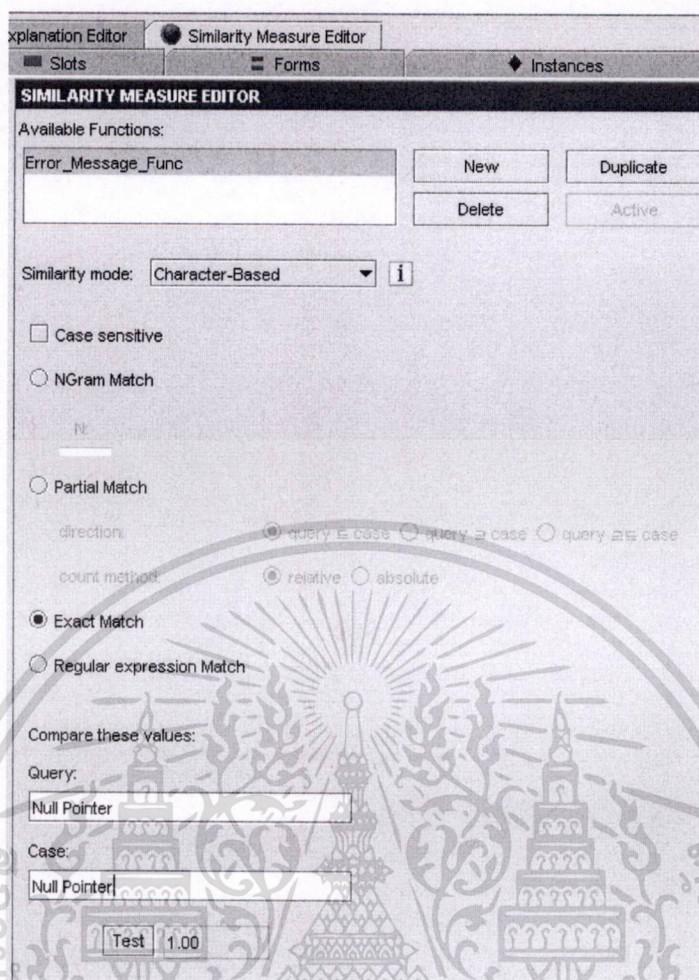
$$\text{sim}(\text{Error_Message}_{\text{query}}, \text{Error_Message}_{\text{case1}}) = \begin{cases} 1 & \text{if Null Pointer} = \text{Null Pointer} \\ 0 & \text{if Null Pointer} \neq \text{Null Pointer} \end{cases} = 1$$

และ

$$\text{sim}(\text{Error_Message}_{\text{case1}}, \text{Error_Message}_{\text{case2}}) = \begin{cases} 1 & \text{if Null Pointer} = \text{NameError: ChkCnt} \\ 0 & \text{if Null Pointer} \neq \text{NameError: ChkCnt} \end{cases} = 0$$

ค่าความคล้ายคลึงของสล็อตข้อความผิดพลาด (Error Message) ระหว่างข้อมูลสอบถาม (query) คือ Null Pointer และกรณี1 คือ Null Pointer มีค่าเป็น 1 และ ระหว่างข้อมูลสอบถาม(query) คือ Null Pointer และกรณี2 คือ NameError: ChkCnt มีค่าเป็น 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 โหมดการวัดความคล้ายคลึงของสตริงข้อความผิดพลาด

- รหัสผิดพลาด (**Error Code**) เป็นประเภทสายอักขระ ใช้โหมดการวัดความคล้ายคลึงแบบใช้ตัวอักษร (character-based) และฟังก์ชันการวัดค่าความคล้ายคลึงแบบเหมือนทั้งหมด (exact mode) ดังรูปที่ 3.8 เนื่องจากข้อมูลรหัสผิดพลาดมีลักษณะเป็นตัวอักษรหรือตัวเลขที่บ่งบอกข้อผิดพลาดของงานอีทีแอลนั้นๆ และบางครั้งการผิดพลาดของงานอีทีแอลไม่ได้แสดงรหัสผิดพลาดจะให้มีค่าเป็น ไม่ทราบ(Unknown) เพื่อให้สามารถวัดค่าความคล้ายคลึงได้ ดังนั้นการวัดความคล้ายคลึงของรหัสผิดพลาดระหว่างข้อมูลสอบถามและข้อมูลกรณีต้องเหมือนกันทั้งหมดจึงจะถือว่าเป็นรหัสผิดพลาดของงานอีทีแอลเดียวกันซึ่งวัดค่าความคล้ายคลึงได้ดังนี้

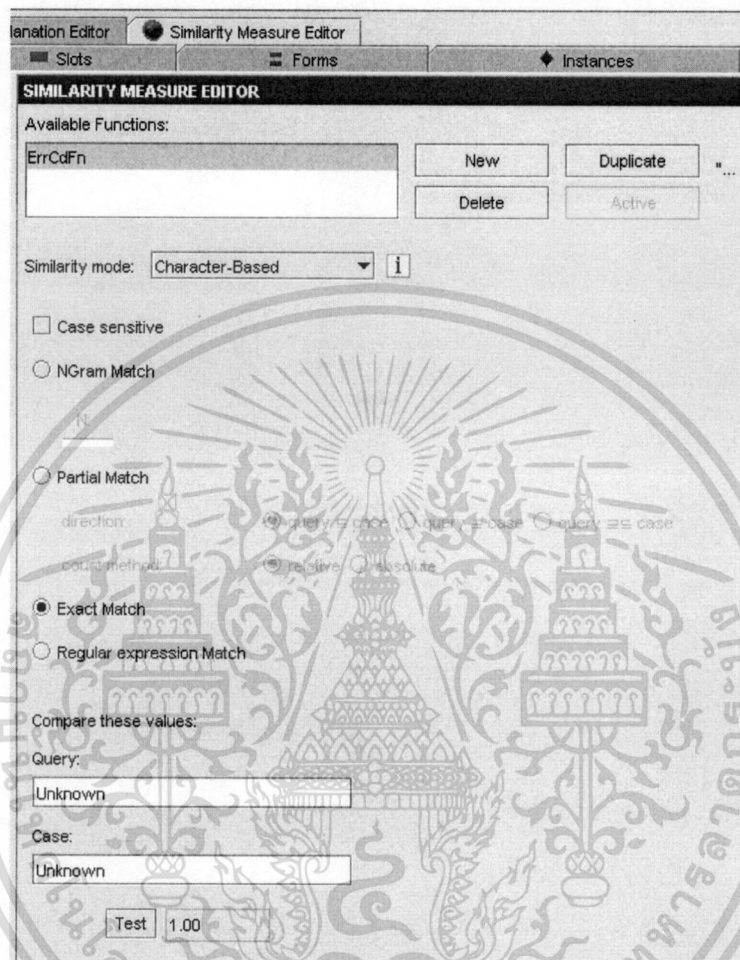
$$\text{sim}(\text{Error_Code}_{\text{query}}, \text{Error_Code}_{\text{case1}}) = \begin{cases} 1 & \text{if } \text{Unknown} = \text{Unknown} \\ 0 & \text{if } \text{Unknown} \neq \text{Unknown} \end{cases} = 1$$

และ

$$\text{sim}(\text{Error_Code}_{\text{case1}}, \text{Error_Code}_{\text{case2}}) = \begin{cases} 1 & \text{if } \text{Unknown} = \text{Unknown} \\ 0 & \text{if } \text{Unknown} \neq \text{Unknown} \end{cases} = 1$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความคล้ายคลึงของสล็อตรหัสผิดพลาด (Error Code) ระหว่างข้อมูลสอบถาม(query) คือ Unknown และกรณี1 คือ Unknown มีค่าเป็น 1 และ ระหว่างข้อมูลสอบถาม(query) คือ Unknown และ กรณี2 คือ Unknown มีค่าเป็น 1



รูปที่ 3.8 โหมดการวัดความคล้ายคลึงของสล็อตรหัสผิดพลาด

- ชื่องานย่อย (Sub Task Name) เป็นประเภทสายอักขระ ใช้โหมดการวัดความคล้ายคลึงแบบใช้ตัวอักษร (character-based) และฟังก์ชันการวัดค่าความคล้ายคลึงแบบเหมือนทั้งหมด (exact mode) ดังรูปที่ 3.9 เนื่องจากข้อมูลชื่องานย่อยมีลักษณะเป็นตัวอักษรที่บ่งบอกชื่องานย่อยของงานอีทีแอลที่ผิดพลาดนั้นๆ ดังนั้นการวัดความคล้ายคลึงของชื่องานย่อยระหว่างข้อมูลสอบถามและข้อมูลกรณีจะต้องเหมือนกันทั้งหมดจึงจะถือว่าเป็นชื่องานย่อยของงานอีทีแอลเดียวกันซึ่งวัดค่าความคล้ายคลึงได้ดังนี้

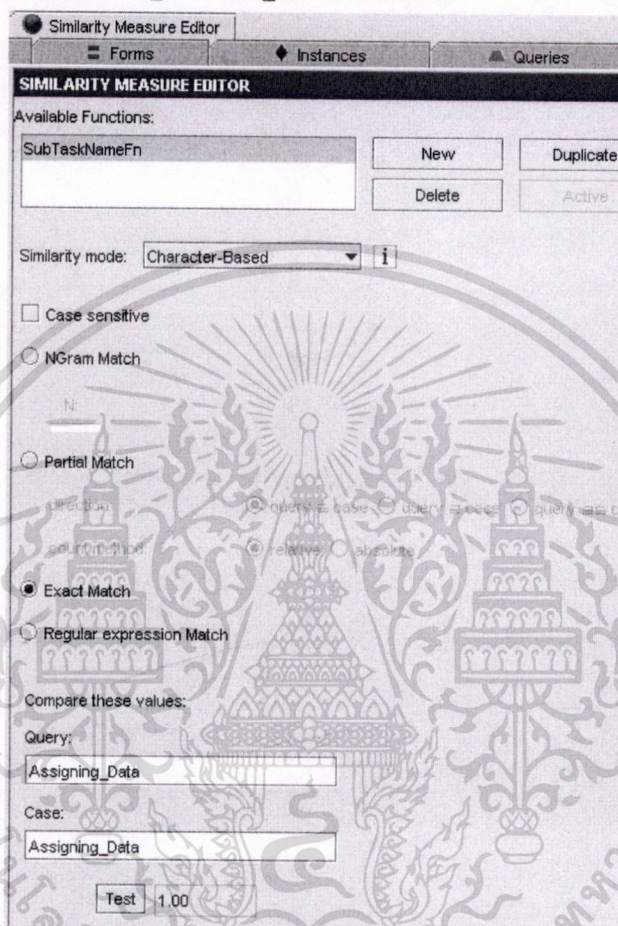
$$\begin{aligned} \text{sim}(\text{Sub_Task_Name}_{\text{query}}, \text{Sub_Task_Name}_{\text{case1}}) \\ = \begin{cases} 1 & \text{if Assigning_Data} = \text{Assigning_Data} \\ 0 & \text{if Assigning_Data} \neq \text{Assigning_Data} \end{cases} = 1 \end{aligned}$$

และ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{aligned} & \text{sim}(\text{Sub_Task_Name}_{\text{case1}}, \text{Sub_Task_Name}_{\text{case2}}) \\ & = \begin{cases} 1 & \text{if } \text{Assigning_Data} = \text{Insert_Reference_Code} \\ 0 & \text{if } \text{Assigning_Data} \neq \text{Insert_Reference_Code} \end{cases} = 0 \end{aligned}$$

ค่าความคล้ายคลึงของสตริงชื่องานย่อย (Sub Task Name) ระหว่างข้อมูลสอบถาม(query) คือ Assigning_Data และกรณี1 คือ Assigning_Data มีค่าเป็น 1 และ ระหว่างข้อมูลสอบถาม(query) คือ Assigning_Data และกรณี2 คือ Insert_Reference_Code มีค่าเป็น 0



รูปที่ 3.9 โหมดการวัดความคล้ายคลึงของสตริงชื่องานย่อย

การคำนวณค่า $\text{sim}(a, b)$ จากสมการในรูปที่ 2.3 ในโครงการนี้จะเลือกใช้เฉพาะสมการที่เหมาะสมกับประเภทข้อมูล (value type) ของสตริงที่ได้กล่าวไปแล้วซึ่งประเภทข้อมูลแต่ละสตริงเป็นดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 ประเภทข้อมูลแต่ละสล็อต

ชื่อสล็อต	ประเภทข้อมูล(Value Type)	จำนวนสมาชิก (Cardinality)
ชื่อระบบงาน (System Name)	สายอักขระ (String)	เดี่ยว (single)
ชื่องาน (Job Name)	สายอักขระ (String)	เดี่ยว (single)
ประเภทของงาน (Job Type)	สัญลักษณ์ (Symbol)	เดี่ยว (single)
ข้อความผิดพลาด (Error Message)	สายอักขระ (String)	เดี่ยว (single)
รหัสผิดพลาด (Error Code)	สายอักขระ (String)	เดี่ยว (single)
ชื่องานย่อย (Sub Task Name)	สายอักขระ (String)	เดี่ยว (single)

จากตารางที่ 3.6 ประเภทของข้อมูลแต่ละสล็อตของโครงการนี้เป็นประเภทสายอักขระ (String) หรือสัญลักษณ์ (Symbol) และมีจำนวนสมาชิก(Cardinality) เป็นค่าเดี่ยว (single) เท่านั้น จึงไม่เหมาะสมกับ สมการที่ได้กล่าวไปแล้วในหัวข้อ 2.2.2 การค้นคืนกรณี ในเรื่องการวัดค่าความคล้ายคลึงในระดับคุณลักษณะดังนี้

- สมการ (2.1) Numeric ที่ใช้ในการคำนวณการวัดค่าความคล้ายคลึงของประเภทข้อมูลประเภทตัวเลข
- สมการ (2.3),(2.4),(2.5) Symbolic, Multi-valued ที่ใช้ในการคำนวณการวัดค่าความคล้ายคลึงประเภทข้อมูลประเภทสัญลักษณ์ที่เป็นแบบหลายค่า
- สมการ (2.6),(2.7) Taxonomy ที่ใช้ในการคำนวณการวัดค่าความคล้ายคลึงของประเภทข้อมูลที่มีการจัดกลุ่มเป็นลำดับชั้น

โครงการนี้จึงเลือกใช้การวัดค่าความคล้ายคลึงในระดับคุณลักษณะแบบ เหมือนทั้งหมด (exact mode) , กำหนดความคล้ายคลึงโดยใช้ตาราง และ เอ็นแกรม ตามที่ได้อธิบายในแต่ละสล็อตก่อนหน้า

2. การวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity)

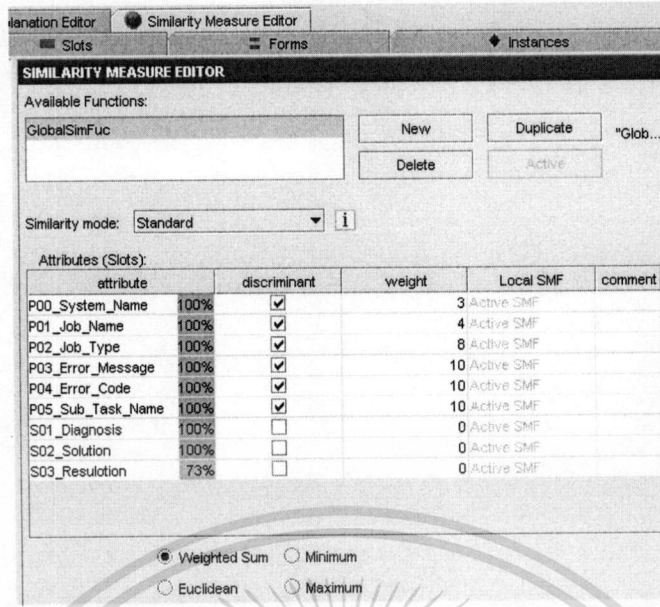
หลังจากทำการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ (Local Similarity) แล้วจะทำการวัดค่าความคล้ายคลึงในระดับกรณี (Global Similarity)เป็นการวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถาม(query) และกรณี (Case)โดยใช้ค่า $\text{sim}(a,b)$ จากการวัดค่าความคล้ายคลึงในระดับคุณลักษณะ มาใช้ในการประเมินความคล้ายคลึง และมีการใช้ค่าถ่วงน้ำหนักของแต่ละคุณลักษณะ มาใช้ในการคำนวณ ซึ่งค่าถ่วงน้ำหนักของแต่ละคุณลักษณะได้มาจากการกำหนดและสอบถามผู้เชี่ยวชาญดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 ค่าถ่วงน้ำหนักของแต่ละคุณลักษณะ

ส่วนคำอธิบายปัญหา (Problem Description)	ผู้เชี่ยวชาญ			ค่าถ่วงน้ำหนักเฉลี่ย (บ้ดเศษ)
	a	b	c	
ชื่อระบบงาน (System Name)	3	3	4	3
ชื่องาน (Job Name)	4	5	3	4
ประเภทของงาน (Job Type)	7	9	9	8
ข้อความผิดพลาด (Error Message)	10	10	10	10
รหัสผิดพลาด (Error Code)	10	10	10	10
ชื่องานย่อย (Sub Task Name)	10	10	10	10

จากตารางที่ 3.7 เนื่องจากโครงการนี้มุ่งเน้นที่การแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดผู้เชี่ยวชาญจึงให้น้ำหนักคุณลักษณะ (Feature) ที่เกี่ยวข้องกับการแสดงข้อผิดพลาดของกระบวนการอีทีแอล เช่น ข้อความผิดพลาด รหัสผิดพลาด และ ชื่องานย่อย มากกว่า ชื่อระบบงาน และชื่องาน ส่วนประเภทของงานมีความสำคัญในการแยกงานว่าอยู่ในประเภทเดียวกันหรือไม่จึงให้น้ำหนักในระดับสูงเช่นกันสาเหตุที่ให้ผู้เชี่ยวชาญเป็นผู้ให้น้ำหนักเนื่องจากผู้เชี่ยวชาญเป็นผู้ที่มีประสบการณ์หรือมีความรู้ในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดจึงเป็นผู้ที่รู้ความสำคัญและสามารถให้น้ำหนักของรายละเอียดคุณลักษณะ (feature) แต่ละตัวในส่วนคำอธิบายปัญหาได้อย่างเหมาะสม โดยดูว่าคุณลักษณะใดมีความสำคัญในการใช้พิจารณาการแก้ไขปัญหากระบวนการอีทีแอลและให้น้ำหนักคุณลักษณะนั้นๆจากการสอบถามรูปที่ 3.10 เป็นการกำหนดน้ำหนักของคุณลักษณะ (Feature) และฟังก์ชันการวัดค่าความคล้ายคลึงในระดับกรณี



รูปที่ 3.10 คำนวณน้ำหนักของคุณลักษณะและฟังก์ชันการวัดค่าความคล้ายคลึงระดับกรณีในมายซีปียอร์

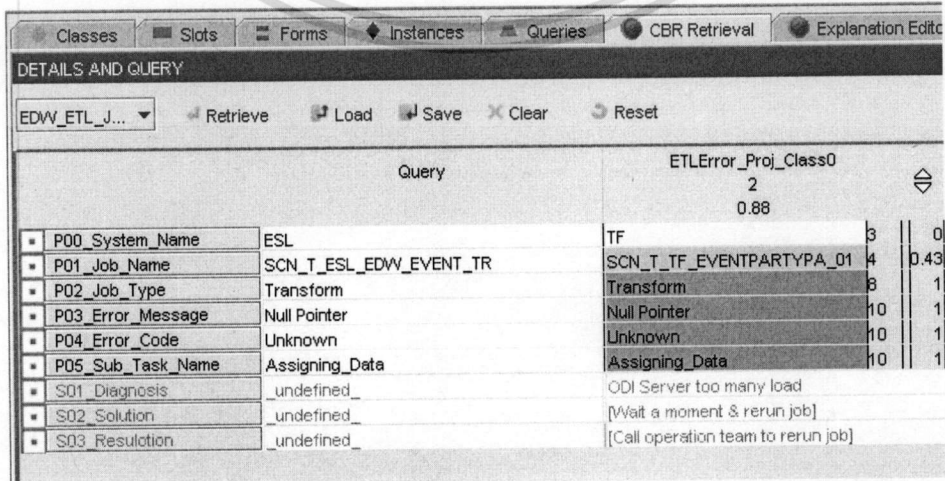
ดังนั้นการวัดค่าความคล้ายคลึงในระดับกรณีเป็นการคำนวณค่า similarity(Case1,CaseR) จากสมการในรูป 2.5 ดังนี้

- การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถาม(query)และกรณี1(Case1)

$$\text{similarity}(\text{query}, \text{case1}) = \frac{(W_{\text{system_name}} * \text{sim}_{\text{system_name}}) + (W_{\text{job_name}} * \text{sim}_{\text{job_name}}) + (W_{\text{job_type}} * \text{sim}_{\text{job_type}}) + (W_{\text{error_message}} * \text{sim}_{\text{error_message}}) + (W_{\text{error_code}} * \text{sim}_{\text{error_code}}) + (W_{\text{Sub_Task_Name}} * \text{sim}_{\text{Sub_Task_Name}})}{W_{\text{system_name}} + W_{\text{job_name}} + W_{\text{job_type}} + W_{\text{error_message}} + W_{\text{error_code}} + W_{\text{Sub_Task_Name}}}$$

$$\text{similarity}(\text{query}, \text{case1}) = \frac{(3 * 0) + (4 * 0.4347826) + (8 * 1) + (10 * 1) + (10 * 1) + (10 * 1)}{3 + 4 + 8 + 10 + 10 + 10}$$

$$\text{similarity}(\text{query}, \text{case1}) = 0.88$$



รูปที่ 3.11 การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถามและกรณี1ในมายซีปียอร์

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของบริษัทฯ ขอสงวนสิทธิ์ในเนื้อหา ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถาม(queryและกรณี2(Case2)

$$\text{similarity}(\text{query}, \text{case2}) = \frac{(W_{\text{system_name}} * \text{sim}_{\text{system_name}}) + (W_{\text{job_name}} * \text{sim}_{\text{job_name}}) + (W_{\text{job_type}} * \text{sim}_{\text{job_type}}) + (W_{\text{error_message}} * \text{sim}_{\text{error_message}}) + (W_{\text{error_code}} * \text{sim}_{\text{error_code}}) + (W_{\text{Sub_Task_Name}} * \text{sim}_{\text{Sub_Task_Name}})}{W_{\text{system_name}} + W_{\text{job_name}} + W_{\text{job_type}} + W_{\text{error_message}} + W_{\text{error_code}} + W_{\text{Sub_Task_Name}}}$$

$$\text{similarity}(\text{query}, \text{case2}) = \frac{(3 * 0) + (4 * 0.25) + (8 * 0.3) + (10 * 0) + (10 * 1) + (10 * 0)}{3 + 4 + 8 + 10 + 10 + 10}$$

$$\text{similarity}(\text{query}, \text{case2}) = 0.30$$

Query	ETLError_Proj_Class70001			
P00_System_Name	ESL	TF	3	0
P01_Job_Name	SCN_T_ESL_EDW_EVENT_TR	SCN_T_TF_IMCOFL_1_0_M09700	4	0.25
P02_Job_Type	Transform	Bmap	8	0.3
P03_Error_Message	Null Pointer	NameError: ChkCnt	10	0
P04_Error_Code	Unknown	Unknown	10	1
P05_Sub_Task_Name	Assigning_Data	Insert_Reference_Code	10	0
S01_Diagnosis	[undefined]	Script jython source side not pass param...		
S02_Solution	[undefined]	[Rerun Job]		
S03_Resolution	[undefined]	[Call operation to rerun]		

รูปที่ 3.12 การวัดค่าความคล้ายคลึงระหว่างข้อมูลสอบถามและกรณี2ในมายซีบีอาร์

ดังนั้นสามารถสรุปได้ว่ากรณี1 มีค่าความคล้ายคลึงกับกรณีสอบถามหรือปัญหามากกว่ากรณี2 เนื่องจากมีค่า $\text{similarity}(A,B)$ มากกว่ากรณี2

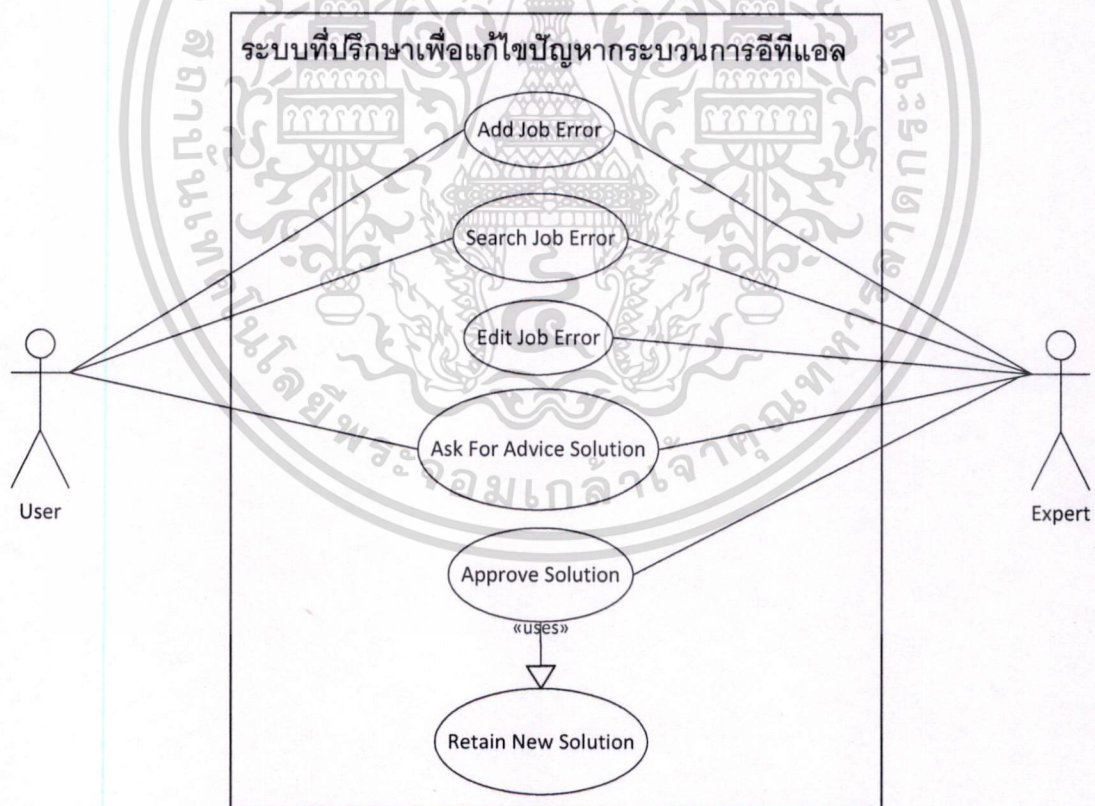
บทที่ 4

การวิเคราะห์และการออกแบบระบบ

ในการวิเคราะห์และออกแบบระบบที่ปรึกษาเพื่อแก้ไขปัญหากระบวนการอีทีแอล โดยใช้วิธีการให้เหตุผลเชิงกรณี กรณีศึกษาแผนกงานคลังข้อมูลขององค์กร ธนาคารไทยพาณิชย์ จะใช้การวิเคราะห์และออกแบบตามหลักการของ ยูเอ็มแอล (Unified Modeling Language : UML) โดยจะนำแผนภาพยูเอ็มแอลมาอธิบายการทำงานของระบบ และลำดับของกิจกรรมในการทำงาน ซึ่งมีดังต่อไปนี้

4.1 ยูสเคสไดอะแกรม (Use case diagram)

จากการศึกษาความต้องการของระบบ สามารถสร้างยูสเคสไดอะแกรม ที่แสดงฟังก์ชันหลักของระบบ และแสดงผู้ใช้ระบบ (Actor) ที่มีความสัมพันธ์กับยูสเคส ดังรูป 4.1



รูปที่ 4.1 ยูสเคสไดอะแกรม (Use Case Diagram)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้ใช้ระบบ (Actor) มี 2 ประเภท คือ ผู้ใช้ทั่วไป (User) และ ผู้เชี่ยวชาญ (Expert)

4.1.1 ผู้ใช้ทั่วไป (User) คือ ผู้ที่เข้ามาใช้งานระบบทั่วไป ซึ่งในที่นี้คือพนักงานในแผนกคลังข้อมูลขององค์การธนาคารไทยพาณิชย์ทั่วไป สามารถค้นหา เพิ่มเติม กระบวนการอีทีแอลที่ผิดพลาด และขอคำแนะนำแนวทางการแก้ปัญหาจากระบบได้

4.1.2 ผู้เชี่ยวชาญ (Expert) คือ ผู้ที่เข้ามาใช้งานระบบที่มีความเชี่ยวชาญ ซึ่งในที่นี้คือพนักงานในแผนกคลังข้อมูลขององค์การธนาคารไทยพาณิชย์ที่มีความเชี่ยวชาญในการแก้ปัญหา กระบวนการอีทีแอลที่ผิดพลาด สามารถค้นหา เพิ่มเติม แก้ไข ขอคำแนะนำแนวทางการแก้ปัญหา กระบวนการอีทีแอลที่ผิดพลาด นอกจากนี้ผู้เชี่ยวชาญยังต้องทำการตรวจสอบการบันทึก กระบวนการอีทีแอลที่ผิดพลาดของผู้ใช้งานทั่วไปว่าถูกต้องหรือไม่ และทำการอนุมัติกระบวนการอีทีแอลที่ผิดพลาดและแนวทางการแก้ไขที่ผู้ใช้ทั่วไปบันทึก แล้วระบบจะทำการบันทึกลงคลังกรณีถ้ากรณีนั้นเป็นกรณีที่ระบบไม่เคยมีมาก่อน

จากยูสเคสโคดอะแกรมตามภาพที่ 4.1 สามารถอธิบายรายละเอียดแต่ละยูสเคสได้ดังตารางที่ 4.1 ถึงตารางที่ 4.6 ดังนี้

ตารางที่ 4.1 คำอธิบายยูสเคส Add Job Error

Use Case Name	Add Job Error
Brief Description	เป็นยูสเคสสำหรับบันทึกการกระบวนการอีทีแอลที่ผิดพลาด
Primary Actor	User และ Expert
Pre-conditions	จะต้องมีข้อมูลสำคัญของกระบวนการอีทีแอลที่ผิดพลาด
Basic Flow	<ol style="list-style-type: none"> 1) ระบบ เตรียมหน้าจอสําหรับกรอกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด 2) แอกเตอร์ กรอกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด 3) แอกเตอร์ กดปุ่ม ตกลง เพื่อบันทึกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด 4) ระบบ ตรวจสอบการกรอกข้อมูล 5) ระบบ บันทึกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด 6) ระบบ แจ้งเตือนผลการบันทึก
Alternative Flow	<ol style="list-style-type: none"> 4a) กรอกข้อมูลไม่ครบถ้วน <ol style="list-style-type: none"> 4a1) ระบบแสดงข้อความเตือน 4a2) กลับไปทำข้อสอง 5a) ระบบไม่สามารถบันทึกข้อมูลได้ <ol style="list-style-type: none"> 5a1) ระบบแสดงข้อความผิดพลาด 5a1) กลับไปทำข้อหนึ่ง
Post-Condition	ระบบบันทึก ข้อมูลกระบวนการอีทีแอลที่ผิดพลาด

เอกสารที่สงวนลิขสิทธิ์ของธนาคารไทยพาณิชย์ จำกัด (มหาชน) ให้เข้าไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 คำอธิบายยูสเคส Search Job Error

Use Case Name	Search Job Error
Brief Description	เป็นยูสเคสสำหรับค้นหารายการกระบวนกรออีทีแอลที่ผิดพลาด ที่มีอยู่ในระบบ
Primary Actor	User และ Expert
Pre-conditions	จะต้องมีข้อมูลสำคัญสำหรับการค้นหากระบวนกรออีทีแอลที่ผิดพลาด
Basic Flow	1) ระบบ เตรียมหน้าจอสำหรับค้นหาข้อมูลกระบวนกรออีทีแอลที่ผิดพลาด 2) แอกเตอร์ กรอกข้อมูลเพื่อใช้ในการค้นหา 3) แอกเตอร์ กดปุ่มค้นหา 4) ระบบ แสดงรายการกระบวนกรออีทีแอลที่ผิดพลาดที่พบในระบบ
Alternative Flow	4a) ระบบ ไม่พบรายการกระบวนกรออีทีแอลที่ผิดพลาด 4a1) ระบบ แจ้งผลการค้นหา
Post-Condition	ระบบ แสดงรายการกระบวนกรออีทีแอลที่ผิดพลาดตามที่ได้ค้นหา

ตารางที่ 4.3 คำอธิบายยูสเคส Edit Job Error

Use Case Name	Edit Job Error
Brief Description	เป็นยูสเคสสำหรับแก้ไขกระบวนกรออีทีแอลที่ผิดพลาด
Primary Actor	Expert
Pre-conditions	จะต้องมีข้อมูลสำคัญสำหรับการค้นหากระบวนกรออีทีแอลที่ผิดพลาดที่ต้องการแก้ไข
Basic Flow	1) ระบบ เตรียมหน้าจอสำหรับค้นหาข้อมูลกระบวนกรออีทีแอลที่ผิดพลาดที่ต้องการแก้ไข 2) แอกเตอร์ กรอกข้อมูลเพื่อใช้ในการค้นหา 3) แอกเตอร์ กดปุ่มค้นหา 4) ระบบ แสดงรายการกระบวนกรออีทีแอลที่ผิดพลาดที่พบในระบบ 5) แอกเตอร์ แก้ไขข้อมูล 6) แอกเตอร์ ปุ่มแก้ไข เพื่อบันทึกข้อมูลลงระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 คำอธิบายยูสเคส Edit Job Error (ต่อ)

Alternative Flow	4a) ระบบไม่พบรายการกระบวนการอีทีแอลที่ผิดพลาด 4a1) ระบบแจ้งผลการค้นหา 6a) ระบบไม่สามารถบันทึกข้อมูลได้ 6a1) ระบบแจ้งข้อความผิดพลาด
Post-Condition	ระบบ บันทึกข้อมูลกระบวนการอีทีแอลที่ผิดพลาดที่ต้องการแก้ไขระบบ

ตารางที่ 4.4 คำอธิบายยูสเคส Ask For Advice Solution

Use Case Name	Ask For Advice Solution
Brief Description	เป็นยูสเคสสำหรับขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาดจากระบบ
Primary Actor	User และ Expert
Pre-conditions	จะต้องมีข้อมูลสำคัญสำหรับการขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด คือ ประเภทของงาน, ชื่อระบบงาน, ชื่องาน, ชื่องานย่อย, ข้อความผิดพลาด, รหัสผิดพลาด
Basic Flow	1) ระบบ เตรียมหน้าจอสําหรับการขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด 2) แอกเตอร์ กรอกข้อมูลเพื่อขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด 3) แอกเตอร์ กดปุ่มค้นหา 4) ระบบ แสดงรายการคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด 5) แอกเตอร์ เลือกใช้แนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด
Alternative Flow	4a) ระบบ ไม่พบรายการคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด 4a1) ระบบแจ้งผลการค้นหา 5a) กรณีที่แอกเตอร์ กดปุ่ม ใช้แนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาดจะใช้งาน (Uses) <i>Use Case Add Job Error</i>
Post-Condition	ระบบ แสดงรายการคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 คำอธิบายยูสเคส Approve Solution

Use Case Name	Approve Solution
Brief Description	เป็นยูสเคสสำหรับอนุมัติกระบวนการอีทีแอลที่ผิดพลาดและแนวทางการแก้ไขกระบวนการอีทีแอลที่แอกเตอร์ผู้ใช้ทั่วไปและผู้เชี่ยวชาญบันทึกเข้ามาในระบบ
Primary Actor	Expert
Including	Retain New Solution
Pre-conditions	จะต้องมีรายการบันทึกกระบวนการอีทีแอลที่ผิดพลาดเข้ามาในระบบ
Basic Flow	1) ระบบ เตรียมหน้าจอรายการบันทึกกระบวนการอีทีแอลที่ผิดพลาดเข้ามาในระบบ 2) แอกเตอร์ กดปุ่มอนุมัติ รายการกระบวนการอีทีแอลที่ผิดพลาด 3) ระบบ บันทึกข้อมูลการอนุมัติ
Alternative Flow	3a) ระบบ ไม่สามารถบันทึกข้อมูลได้ 3a1) ระบบ แจ้งข้อความผิดพลาด 3b) กรณีที่ระบบบันทึกข้อมูลแล้วจากเรียกใช้ (Uses) <i>Use Case Retain New Solution</i>
Post-Condition	ระบบ อนุมัติรายการบันทึกกระบวนการอีทีแอลที่ผิดพลาดเข้ามาในระบบที่แอกเตอร์เลือก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 คำอธิบายยูสเคส Retain New Solution

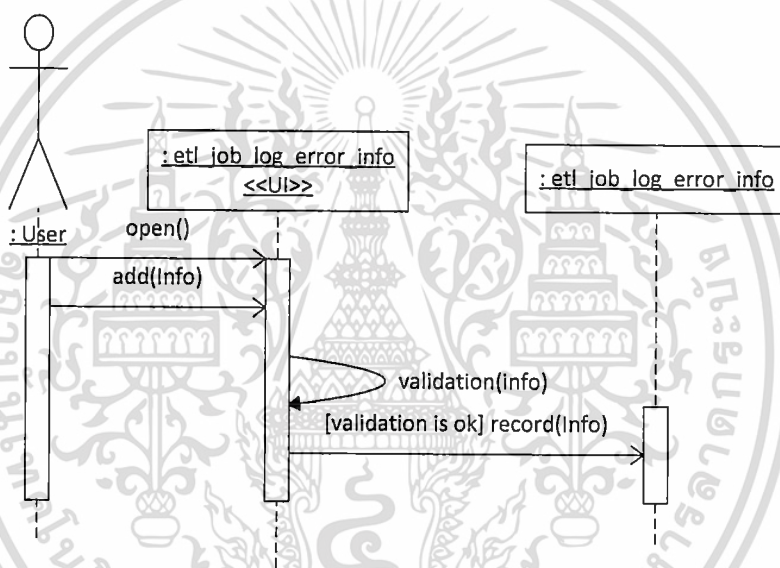
Use Case Name	Retain New Solution
Brief Description	เป็นยูสเคสสำหรับการบันทึกกระบวนการอีทีแอลที่ผิดพลาดที่ไม่เคยมีอยู่ในระบบมาก่อน ลงคลังกรณี
Primary Actor	System
Pre-conditions	จะต้องมีข้อมูลสำคัญสำหรับการบันทึกกระบวนการอีทีแอลที่ผิดพลาดที่ไม่เคยมีอยู่ในระบบมาก่อน โดยแอกเตอร์ผู้เชี่ยวชาญ (Expert) ต้องเรียกใช้ <i>Use Case Approve Solution</i>
Basic Flow	1) ระบบ ทำการตรวจสอบกระบวนการอีทีแอลที่ผิดพลาด 2) ระบบ บันทึกกระบวนการอีทีแอลที่ผิดพลาดลงคลังกรณี
Alternative Flow	1a) กรณีที่กระบวนการอีทีแอลที่ผิดพลาดนั้นอยู่ในระบบอยู่แล้ว ระบบจะไม่ทำการบันทึกลงคลังกรณี 1b) กรณีที่กระบวนการอีทีแอลที่ผิดพลาดนั้น ไม่เคยมีอยู่ในระบบ ระบบจะทำการบันทึกลงคลังกรณี
Post-Condition	ระบบ บันทึกกระบวนการอีทีแอลที่ผิดพลาดและแนวทางการแก้ไขที่ไม่เคยมีอยู่ในระบบมาก่อนลงคลังกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ซีแควนไคอะแกรม (Sequence Diagram)

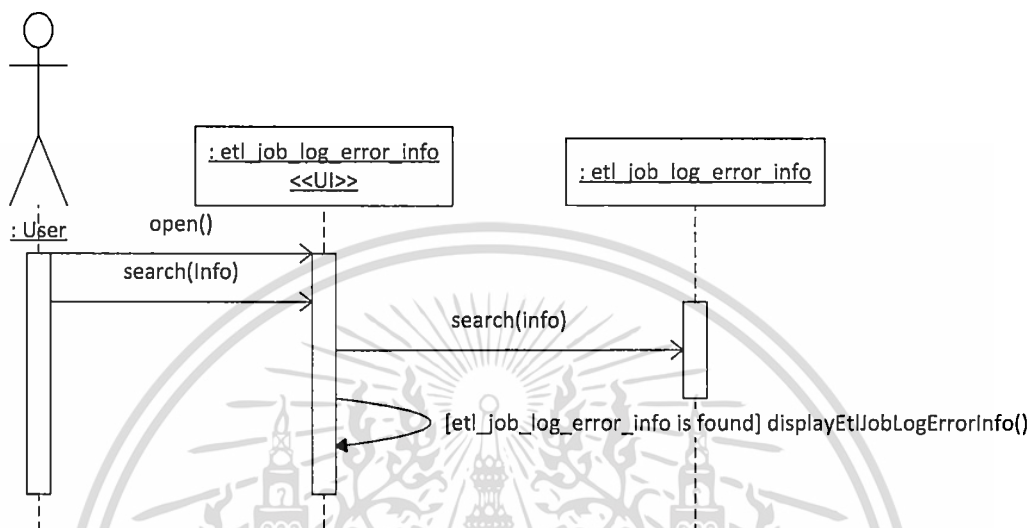
ซีแควนไคอะแกรมจะอธิบายลำดับขั้นตอนการทำงานของเหตุการณ์หรือกิจกรรมที่เกิดขึ้นของระบบ โดยกิจกรรมเหล่านี้เกิดขึ้นจากการมีปฏิสัมพันธ์ (Interaction) หรือการติดต่อสื่อสารกันระหว่างออบเจกต์ โดยมีการกำหนดลำดับการเกิดขึ้นของการสื่อสารระหว่างออบเจกต์ ซึ่งสามารถแสดงซีแควนไคอะแกรมของกระบวนการในระบบงานได้ดังนี้

กิจกรรมการบันทึกรายการกระบวนการอีทีแอลที่ผิดพลาด (Add Job Error) คือ การที่ผู้ใช้กรอกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด ผ่านหน้าจอจากนั้นระบบตรวจสอบการกรอกข้อมูล ถ้ากรอกข้อมูลครบถ้วน ระบบจะทำการบันทึกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด ดังภาพที่ 4.2



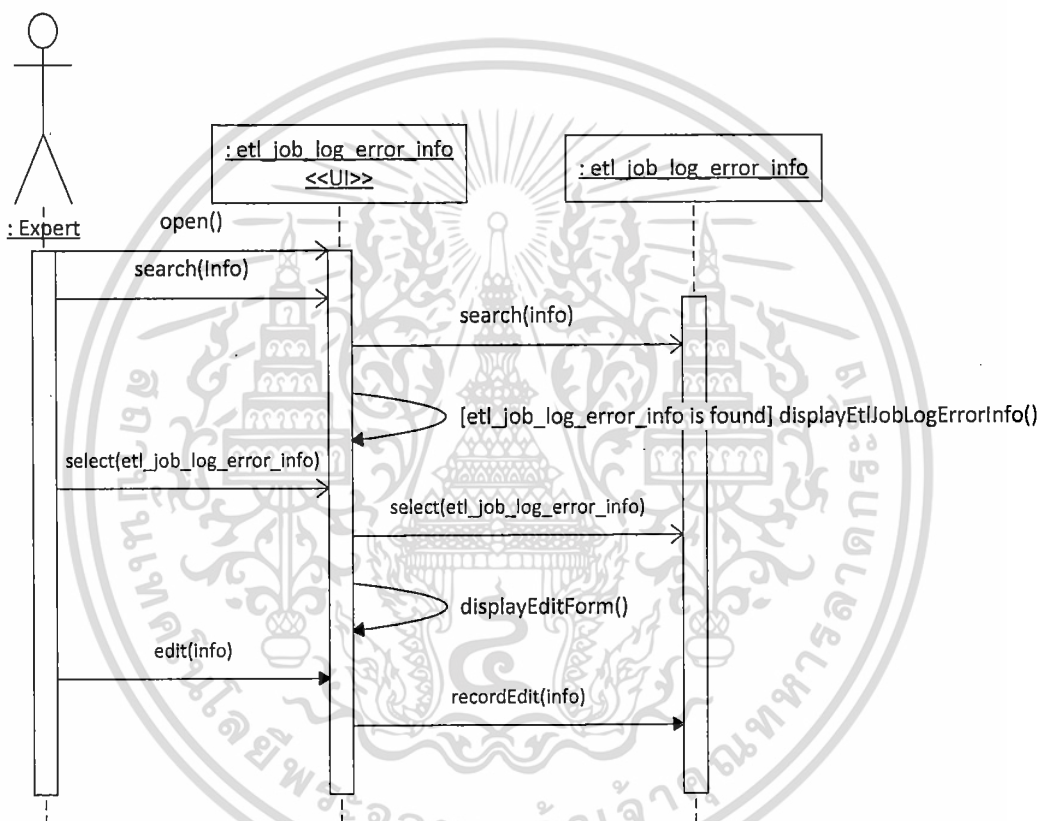
รูปที่ 4.2 ซีแควนไคอะแกรมสำหรับการบันทึกรายการกระบวนการอีทีแอลที่ผิดพลาด

กิจกรรมสำหรับค้นหารายการกระบวนการอีทีแอลที่ผิดพลาด (Search Job Error) คือ การที่ผู้ใช้กรอกข้อมูลเพื่อใช้ในการค้นหากระบวนการอีทีแอลที่ผิดพลาดตามที่ใช้ต้องการ ผ่าน หน้าจอจากนั้นระบบจะทำการค้นหากระบวนการอีทีแอลที่ผิดพลาด และแสดงรายการ กระบวนการอีทีแอลที่ผิดพลาดที่พบในระบบผ่านทางหน้าจอ ดังภาพที่ 4.3



รูปที่ 4.3 ซีควเอนโคอะแกรมสำหรับค้นหารายการกระบวนการอีทีแอลที่ผิดพลาด

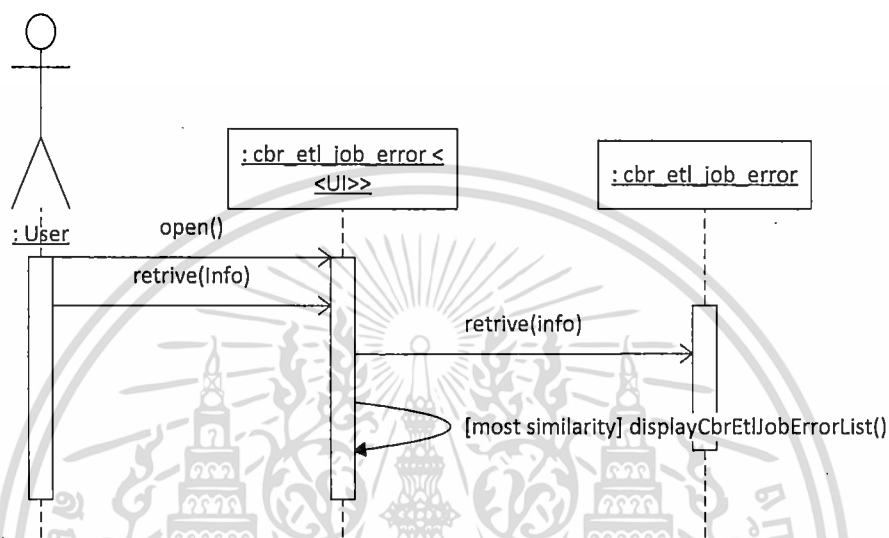
กิจกรรมสำหรับแก้ไขกระบวนการอีทีแอลที่ผิดพลาด (Edit Job Error) คือ การที่ผู้ใช้ประเภทผู้เชี่ยวชาญ (Expert) กรอกข้อมูลเพื่อใช้ในการค้นหากระบวนการอีทีแอลที่ผิดพลาดตามที่ผู้ใช้ต้องการแก้ไข ผ่านหน้าจอจากนั้นระบบจะทำการค้นหากระบวนการอีทีแอลที่ผิดพลาด และแสดงรายการกระบวนการอีทีแอลที่ผิดพลาดที่พบในระบบผ่านทางหน้าจอ หลังจากนั้นผู้ใช้ทำการเลือกกระบวนการอีทีแอลที่ผิดพลาดที่ต้องการแก้ไขผ่านทางจอ ระบบจะส่งเข้าสู่หน้าจอแก้ไขกระบวนการอีทีแอลที่ผิดพลาด เมื่อผู้ใช้แก้ไขกระบวนการอีทีแอลที่ผิดพลาดผ่านทางจอ และกดปุ่มแก้ไขระบบจะทำการบันทึกข้อมูล ดังภาพที่ 4.4



รูปที่ 4.4 ซีควเอน ไคอะแกรมสำหรับแก้ไขกระบวนการอีทีแอลที่ผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

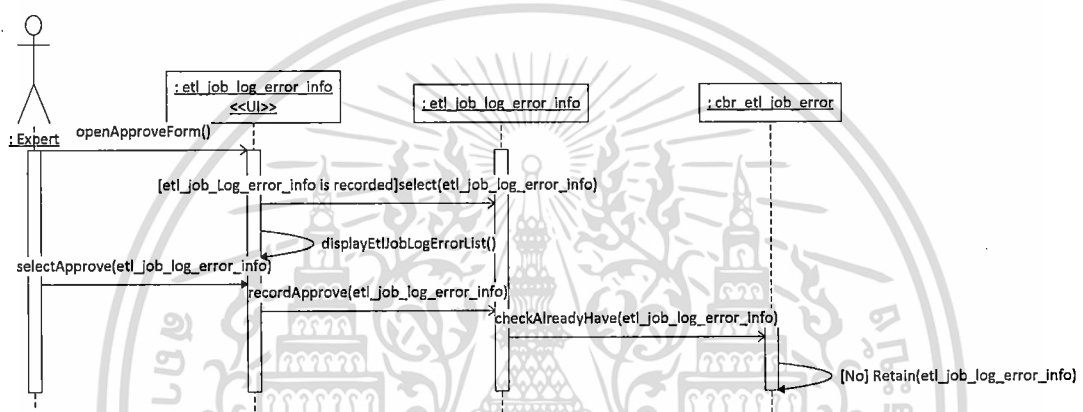
กิจกรรมสำหรับขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด (Ask For Advice Solution) คือ การที่ผู้ใช้กรอกข้อมูลเพื่อขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาดผ่านหน้าจอ จากนั้นระบบจะทำการค้นหาแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาดที่มีความใกล้เคียงมากที่สุดจากคลังกรณีมาแสดงผ่านทางหน้าจอ ดังภาพที่ 4.5



รูปที่ 4.5 ซีควเอนโคอะแกรมสำหรับขอคำแนะนำแนวทางการแก้ไขกระบวนการอีทีแอลที่ผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิจกรรมสำหรับอนุมัติกระบวนการอีทีแอลที่ผิดพลาดและแนวทางการแก้ไขกระบวนการอีทีแอลที่แอกเตอร์ผู้ใช้ทั่วไปและผู้เชี่ยวชาญบันทึกเข้ามาในระบบ (Approve Solution) และกิจกรรมสำหรับการบันทึกกระบวนการอีทีแอลที่ผิดพลาดที่ไม่เคยมีอยู่ในระบบมาก่อนลงคลังกรณี (Retain New Solution) คือ การที่ผู้ใช้ประเภทผู้เชี่ยวชาญ (Expert) เข้าสู่หน้าจอรายการบันทึกกระบวนการอีทีแอลที่ผิดพลาดเข้ามาในระบบ เพื่อทำการอนุมัติรายการที่มีกระบวนการอีทีแอลที่ผิดพลาดและแนวทางการแก้ไขที่ถูกต้อง โดยกดปุ่มอนุมัติผ่านทางหน้าจอ และระบบจะบันทึกข้อมูลการอนุมัติจากนั้นระบบทำการตรวจสอบกระบวนการอีทีแอลที่ผิดพลาดว่าเคยมีอยู่ในระบบแล้วหรือไม่ ถ้าไม่เคยมีอยู่ในระบบ ระบบจะทำการบันทึกลงคลังกรณี ดังภาพที่ 4.6



รูปที่ 4.6 ซีควเอนโคอะแกรมสำหรับอนุมัติและบันทึกกระบวนการอีทีแอลที่ผิดพลาดลงคลังกรณี

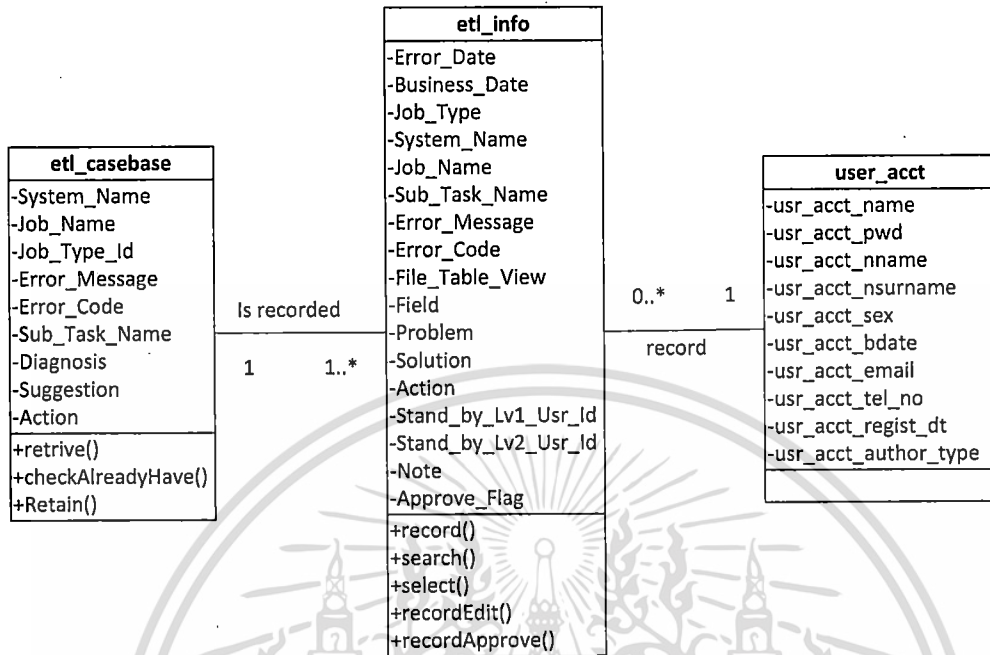
4.3 คลาสโคอะแกรม (Class diagram)

จากขั้นตอนการวิเคราะห์และออกแบบระบบ สามารถที่จะสร้างคลาส (Class) ที่จำเป็นสำหรับการพัฒนาระบบที่ปรึกษาเพื่อแก้ไขปัญหากระบวนการอีทีแอล โดยใช้วิธีการให้เหตุผลเชิงกรณี ได้ดังนี้

- User_acct เป็นคลาสใช้เก็บข้อมูลพนักงานที่เป็นผู้งานใช้ระบบ
- Etl_info เป็นคลาสที่ใช้เก็บประวัติการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดในแต่ละครั้ง ถ้าผู้เชี่ยวชาญอนุมัติรายการกระบวนการอีทีแอลที่ผิดพลาดแล้ว ข้อมูลเหล่านี้จะถูกนำไปใช้ในการตรวจสอบกับคลังกรณีว่าเป็นความรู้ใหม่หรือไม่ ถ้าเป็นความรู้ใหม่ ระบบจะทำการจัดเก็บความรู้นั้น
- Etl_casebase เป็นคลาสที่ใช้จัดเก็บฐานกรณี ทำหน้าที่ค้นคืนแนวทางการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดที่มีความคล้ายคลึงกับกรณีสอบถาม ตรวจสอบความรู้ใหม่ของระบบ และจัดเก็บความรู้ใหม่เข้าคลังกรณี

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากคลาสที่กล่าวมาข้างต้นสามารถนำมาสร้างเป็นคลาสไดอะแกรมได้ดังรูปที่ 4.7



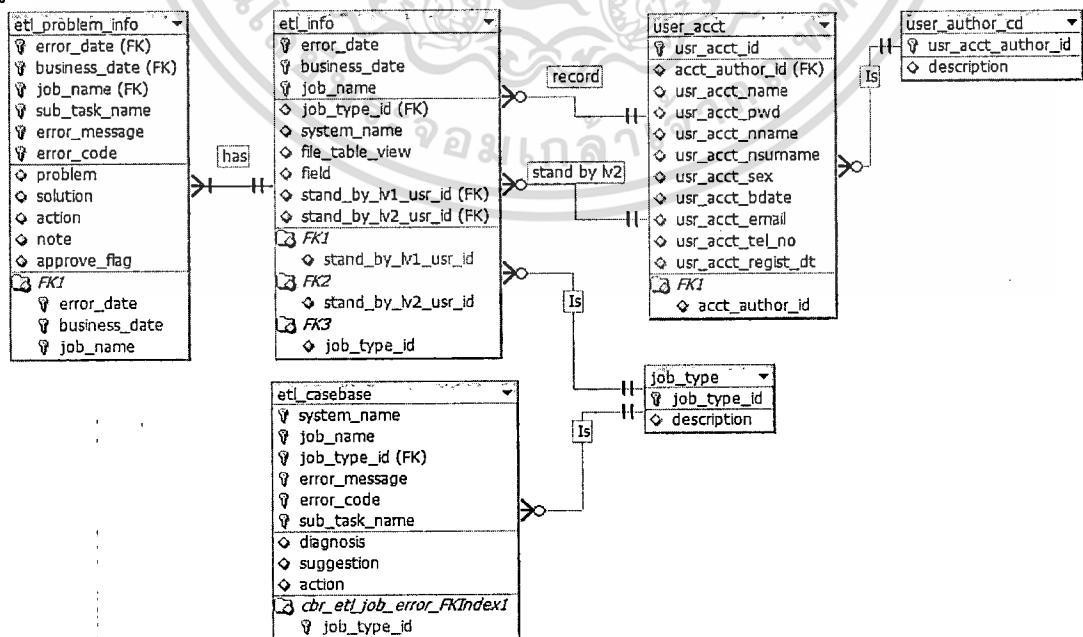
รูปที่ 4.7 คลาสไดอะแกรมระบบที่ปรึกษาเพื่อแก้ไขปัญหากระบวนการอีทีแอล

4.4 การออกแบบฐานข้อมูล

4.4.1 โครงสร้างฐานข้อมูล

จากคลาสไดอะแกรมสามารถออกแบบโครงสร้างและความสัมพันธ์ของฐานข้อมูลได้ดัง

รูป 4.8



รูปที่ 4.8 แบบจำลองความสัมพันธ์ระหว่างเอนทิตี (ER-model)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปสามารถอธิบายความสัมพันธ์ได้ว่าผู้ใช้ (usr_acct) สามารถเป็นประเภทของผู้ใช้ (usr_author_cd) ได้หนึ่งประเภทเช่น ผู้ใช้ทั่วไป หรือ ผู้เชี่ยวชาญ และผู้ใช้สามารถบันทึกข้อมูลกระบวนการอีทีแอลที่ผิดพลาด (etl_info) ได้มากกว่าหนึ่งรายการ อีกทั้งผู้ใช้อังยังเป็นพนักงานแก้ปัญหาในระดับสองที่ถูกบันทึกในข้อมูลกระบวนการอีทีแอลที่ผิดพลาดได้มากกว่าหนึ่งอีกด้วย ส่วนข้อมูลกระบวนการอีทีแอลที่ผิดพลาดเป็นประเภทของงานได้หนึ่งประเภทและคลังข้อมูล (etl_casebase) ก็เป็นประเภทของงานได้หนึ่งประเภทเช่นกัน นอกจากนี้ข้อมูลกระบวนการอีทีแอลที่ผิดพลาดยังมีข้อมูลเกี่ยวกับปัญหาของกระบวนการอีทีแอลที่ผิดพลาด (etl_problem_info) ได้มากกว่า 1 รายการ

4.4.2 พจนานุกรมข้อมูล

ตารางที่ 4.7 ตารางเก็บข้อมูลประเภทของผู้ใช้

ชื่อตาราง : user_author_cd				
ความหมาย : ใช้เก็บข้อมูลประเภทของผู้ใช้				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
usr_acct_author_id	รหัสประเภทผู้ใช้	integer	PK	
description	คำอธิบาย	varchar		

ตารางที่ 4.8 ตารางเก็บข้อมูลประเภทของกระบวนการอีทีแอล

ชื่อตาราง : job_type				
ความหมาย : ใช้เก็บข้อมูลประเภทของกระบวนการอีทีแอล				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
job_type_id	รหัสประเภทกระบวนการอีทีแอล	integer	PK	
description	คำอธิบาย	varchar		

ตารางที่ 4.9 ตารางเก็บข้อมูลผู้ใช้ระบบ

ชื่อตาราง : user_acct				
ความหมาย : ใช้เก็บข้อมูลผู้ใช้ระบบ				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
usr_acct_id	รหัสผู้ใช้ระบบ	integer	PK	
usr_acct_name	ชื่อผู้ล็อกอินผู้ใช้ระบบ	varchar		
usr_acct_pwd	รหัสผ่านผู้ใช้ระบบ	varchar		
usr_acct_nname	ชื่อ	varchar		
usr_acct_nsurname	นามสกุล	varchar		
usr_acct_sex	เพศ	char		
usr_acct_bdate	วันเกิด	date		
usr_acct_email	อีเมล	varchar		
usr_acct_tel_no	เบอร์โทรศัพท์	varchar		
usr_acct_regist_dt	วันลงทะเบียน	datetime		
acct_author_id	รหัสประเภทผู้ใช้	integer	FK	user_author_cd

ตารางที่ 4.10 ตารางเก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาด

ชื่อตาราง : etl_info				
ความหมาย : ใช้เก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาด				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
error_date	วันที่ทำงานผิดพลาด	datetime	PK	
business_date	วันที่ข้อมูลของระบบงาน	date	PK	
job_name	ชื่องาน	varchar	PK	
job_type_id	รหัสประเภทงาน	integer	FK1	job_type
system_name	ชื่อระบบงาน	varchar		
file_table_view	ไฟล์ , ตาราง และวิวฐานข้อมูล ที่ผิดพลาด	char		
field	คอลัมน์ที่ผิดพลาด	date		
stand_by_lv1_usr_id	ชื่อพนักงานแก้ปัญหาระดับหนึ่ง	varchar	FK2	user_acct
stand_by_lv1_usr_id	ชื่อพนักงานแก้ปัญหาระดับสอง	varchar	FK3	user_acct

เอกสารนี้เป็นเอกสารทรัพย์สินทางปัญญาของบริษัทฯ เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 ตารางเก็บข้อมูลรายละเอียดปัญหาของงานกระบวนการอีทีแอลที่ผิดพลาด

ชื่อตาราง : etl_problem_info				
ความหมาย : ใช้เก็บข้อมูลรายละเอียดปัญหาของงานกระบวนการอีทีแอลที่ผิดพลาด				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
error_date	วันที่ที่งานผิดพลาด	datetime	PK,FK	etl_job_log_error_info
business_date	วันที่ข้อมูลระบบ	date	PK,FK	etl_job_log_error_info
job_name	ชื่องาน	varchar	PK,FK	etl_job_log_error_info
sub_task_name	ชื่องานย่อย	varchar	PK	
error_message	ข้อความผิดพลาด	varchar	PK	
error_code	รหัสผิดพลาด	varchar	PK	
problem	ปัญหาที่เกิดขึ้น	varchar		
solution	แนวทางการแก้ไขปัญหา	varchar		
action	สิ่งที่ต้องทำหลังจากพบสาเหตุของปัญหา	varchar		
note	บันทึก	varchar		
approve_flag	0 = อนุมัติแล้ว, 1 = ยังไม่ได้ อนุมัติ	integer		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 ตารางเก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาดที่ระบบมีอยู่ในคลังกรณี

ชื่อตาราง : etl_casebase				
ความหมาย : ใช้เก็บข้อมูลงานกระบวนการอีทีแอลที่ผิดพลาดที่ระบบมีอยู่ในคลังกรณี				
ชื่อเขตข้อมูล	ความหมาย	ชนิดข้อมูล	คีย์	อ้างอิงไปตาราง
system_name	ชื่อระบบงาน	varchar	PK	
job_name	ชื่องาน	varchar	PK	
job_type_id	รหัสประเภทงาน	varchar	PK.FK	job_type
error_message	ข้อความผิดพลาด	varchar	PK	
error_code	รหัสผิดพลาด	varchar	PK	
sub_task_name	ชื่องานย่อย	varchar	PK	
diagnosis	คำวินิจฉัย	varchar		
suggestion	แนวทางการแก้ไขปัญหา	varchar		
action	สิ่งที่ต้องทำหลังจากพบสาเหตุของปัญหา	varchar		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

ผลการทดลองและประเมินประสิทธิผลของระบบ

ในบทนี้จะกล่าวถึง การทดลองขอคำแนะนำการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด โดยใช้ระบบการให้เหตุผลเชิงกรณีที่ได้พัฒนาในบทที่สี่ โดยการวัดประสิทธิผล ซึ่งหมายถึง การวัดค่าความถูกต้องและความแม่นยำในการให้คำแนะนำการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด โดยใช้กรณีทดสอบที่มีส่วนคำอธิบายปัญหา (Problem Description) ที่ไม่เคยอยู่ในระบบมาก่อน มาทำการค้นคืนเพื่อให้ได้ส่วนของคำตอบ (Solution) ที่ระบบให้คำแนะนำ หลังจากนั้นนำส่วนของคำตอบที่ระบบให้คำแนะนำที่ได้ไปให้ผู้เชี่ยวชาญตรวจสอบ โดยในแต่ละกรณีทดสอบ จะแบ่งการวัดความถูกต้องในการให้คำแนะนำการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ออกเป็นสี่ระดับดังนี้

- ค่าความถูกต้องเท่ากับ 1 – คำวินิจฉัยและข้อแนะนำถูกต้อง
- ค่าความถูกต้องเท่ากับ 0.5 – คำวินิจฉัยถูกต้องและข้อแนะนำไม่ถูกต้อง
- ค่าความถูกต้องเท่ากับ 0.3 – คำวินิจฉัยไม่ถูกต้องและข้อแนะนำถูกต้อง
- ค่าความถูกต้องเท่ากับ 0 – คำวินิจฉัยและข้อแนะนำไม่ถูกต้อง

และนำผลที่ได้มาพิจารณาค่าความถูกต้องรวม ค่าความถูกต้องรวมควรมีค่ามากกว่าหรือเท่ากับ 80% จึงจะถือว่าระบบมีประสิทธิผลที่ดี ซึ่งสามารถคำนวณได้จาก

$$\text{ค่าความถูกต้องรวม} = \frac{\text{ผลรวมค่าความถูกต้องของกรณีทั้งหมด}}{\text{จำนวนกรณีที่ทดสอบทั้งหมด}} \times 100$$

ตารางที่ 5.1 แสดงการทดลองการวัดประสิทธิผลของการให้คำแนะนำการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด โดยการทดลองจะใช้ฐานกรณีทั้งหมดหนึ่งร้อยกรณี และกรณีทดสอบสามสิบกรณี ซึ่งแต่ละกรณีในตารางที่ 5.1 จะมีการแสดงรายละเอียดส่วนคำอธิบายปัญหาของกระบวนการอีทีแอลที่ผิดพลาดที่ใช้การค้นคืนส่วนของคำตอบ ส่วนคำวินิจฉัย ข้อแนะนำ และสิ่งที่ควรทำเมื่อพบสาเหตุเป็นข้อแนะนำที่ระบบให้โดยเลือกจากกรณีที่มีความเชื่อมั่นมากที่สุดมาแสดงในตาราง นอกจากนี้ยังมีการแสดงค่าความถูกต้องจากผู้เชี่ยวชาญ 3 ท่าน พร้อมการหาค่ากลางแบบค่าฐานนิยมจากค่าความถูกต้องของผู้เชี่ยวชาญ 3 ท่านเพื่อเป็นค่าความถูกต้องของกรณี อีกทั้งยังมีการแสดงเลขกรณีที่ถูกค้นคืนและค่าความคล้ายคลึง โดยสามารถดูกรณีที่ถูกค้นคืนจากเลขของกรณีได้ในภาคผนวก ก.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 การทดลองการวัดประสิทธิผลของการให้คำแนะนำการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด (อ้างอิง Case no. ดูได้จากภาคผนวก ก)

กรณี ที่	กระบวนการอีทีแอลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
1	Job Type : Transform System Name : LOS Job Name : SCN_T_LOS_TRACK10_PARTASSE_01 Sub Task Name : Insert_flow_I\$_into_table Error Message : Cannot place a null value in a NOT NULL field Error Code : Unknown	Bkey use has not in Bkey Gen	✓	Sample Data and Check All ID FIELD that has value is NOT NULL in Tier2 table compare field in view that has value is null	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ำยคลึง) : 17(87%), 11(65%), 1(43%), 34(43%), 6(42%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิตผล (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)		
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		สิ่งที่ควรทำเมื่อพบ สาเหตุ (Action)	
2	Job Type : Transform System Name : ESL Job Name : SCN_T_ESL_EDW_EVENT_TR Sub Task Name : Assigning_Data Error Message : Null Pointer Error Code : Unknown	ODI Server too many load	✓	Wait a moment and Rerun Job	✓	Call Operation Team to Rerun Job	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 1(88%), 8(52%), 11(50%), 34(44%), 18(43%)							

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ				ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)	สิ่งที่ควรทำเมื่อพบ สาเหตุ (Action)	
3	Job Type : Transform System Name : FCD Job Name : SCN_T_STFC_STACTM10_ACCOPRODHIST_01 Sub Task Name : Load_Consumed Error Message : Login failure Error Code : Unknown	Database Server maybe too many load	✓	If check data in table by where businessdate that have data then force ok this job	✓ Call Operation Team to Force OK Job	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 34(94%), 6(87%), 47(72%), 42(66%), 14(46%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)		
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
4	Job Type : Export System Name : FACT Job Name : SCN_T_FACT_TAXINT1_ACCT_C2J2 Sub Task Name : Is_the_process_already_executed Error Message : already been executed Error Code : Unknown	This Scenario already been executed	✓	Please cleanup S1DCFEDW.CTLFW_Transform for this process on the businessdate and also delete data in table	✓	Call Operation Team to Rerun Job	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 35(99%), 33(65%), 5(50%), 28(42%), 3(42%)							

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิตผล (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อพบ สาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
5	Job Type : Load System Name : TR Job Name : SCN_L_TR_EDW_SYENTITY Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : Unknown	Data File maybe string contains an untranslatable character	✓	Check Error Code In XXDWKEDW.UV_CCAP_ [Ses_no]	✗ -	a ให้ 1 b ให้ 0.5 c ให้ 0.5 ค่าฐานนิยม = 0.5
Retrieved Case No.(ค่าความคล้ายคลึง) : 16(74%), 15(66%), 46(66%), 48(66%), 12(66%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ได้จาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
6	Job Type : Load System Name : EDW Job Name : SCN_L_EDW_EDW_FEATURE Sub Task Name : Read_Control_File_Info_Updat_CSFE Error Message : File not found Error Code : Unknown	File not found - [PATH][FILENAME].CTL	✓	Check File In Server	✓ Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 40(88%), 45(88%), 2(88%), 37(44%), 36(44%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เอลลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
7	Job Type : Export System Name : FACT Job Name : SCN_T_FACT_LEAD_PARTYASSET Sub Task Name : Insert_records Error Message : Invalid DATE Literal Error Code : Unknown	This Scenario already been executed	✘	Please cleanup S1DCFEDW.CTLFW_Transform for this process on the businessdate and also delete data in table [table]	✘ Call Operation Team to Rerun Job	a ให้ 0 b ให้ 0 c ให้ 0 ค่าฐานนิยม = 0
Retrieved Case No.(ค่าความคล้ายคลึง) : 35(50%), 5(50%), 28(42%), 3(42%), 33(42%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนกรที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ				ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)	สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
8	Job Type : BKEY System Name : PLPS Job Name : SCN_T_PLPS_STDRMCD10_BK_PARTYGRP Sub Task Name : CTLFW_AllLoadedFor_Process Error Message : Not all the source files required Error Code : Unknown	Source file load not complete	✓	Check Dependency Job and Register Files wait for the file in view correctly Rerun Job	✓ Wait Load Job complete and Call Operation term for rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 10(87%), 7(79%), 14(73%), 29(34%), 31(34%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิตผล (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
9	Job Type : BMAP System Name : IM Job Name : SCN_T_IMFC_IMTBLM10_M01200_01 Sub Task Name : Insert_new_B-maps Error Message : Login failure Error Code : Unknown	Database Server maybe too many load	✓	If check data in table : [table] by where businessdate that have data then force ok this job	✓ Call Operation Team to Force OK Job	a ให้ 1 b ให้ 1 c ให้ 0.5 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 6(52%), 34(52%), 7(42%), 29(42%), 31(41%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิต (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
10	Job Type : BMAP System Name : PLPS Job Name : SCN_T_PLPS_ICOLLD3_1_0_M22950 Sub Task Name : CTLFW_AllLoadedFor_Process Error Message : Not all the source files required Error Code : Unknown	Source file load not complete	✓	Check Dependency Job and Register Files wait for the file in view correctly	✓ Wait Load Job complete and Call Operation term for rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 7(87%), 10(79%), 14(73%), 29(43%), 31(43%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
11	Job Type : Load System Name : FRC Job Name : SCN_L_FRC_EDW_FRCOSWDT Sub Task Name : ET_UV_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : 6706	The string contains an untranslatable character	✓	1.Sample data record as Field SourceSeq in XXDWKEDW.UV_[System Name]_[Ses_no] 2.Check Field and Sample data that is latin but field in data file is unicode and compare source layout spec	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่านิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 16(88%), 48(66%), 46(66%), 12(66%), 15(66%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนกรรที่เแอลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)		
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
12	Job Type : Load System Name : ALM Job Name : SCN_L_ALM_OCPARMET Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : 2794	Data File duplicate key	✓	Sample Data as Field SourceSeq in XXDWKEDW. UV_[System Name]_[Ses_no]	✓	Inform stand by level 2 and Call source system team for check file as Sample Data	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 48(87%) , 12(87%), 46(65%), 16(65%), 15(65%)							

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)		
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
13	Job Type : Load System Name : BKL Job Name : SCN_L_BKL_EDW_FRAUD Sub Task Name : Run MultiLoad Error Message : Not enough data in vartext record Error Code : Unknown	Data File is not completed or have Enter in content of data.	✓	Check data in DAT file	✓	Open Incident to Source System For Fixed File and Next Call Operation For Rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่านิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 37(66%), 40(44%), 45(44%), 36(44%), 2(43%)							

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เอลลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)		
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
14	Job Type : Bkey System Name : MX Job Name : SCN_T_MX_SECFEATE10_BK_FEATURE_23 Sub Task Name : Unknown Error Message : Not Found Batch Error Code : Unknown	Source file load not complete	✘	Check Source File [filename] in CTLFW_FILE_PROCESS already Load complete? If Not Complete then Rerun [filename] File	✘	Wait Load File Job complete and Call Operation term for rerun	a ให้ 0 b ให้ 0 c ให้ 0 ค่าฐานนิยม = 0
Retrieved Case No.(ค่าความคล้ายคลึง) : 10(56%), 26(51%), 7(35%), 29(34%), 31(34%)							

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง จาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
15	Job Type : Transform System Name : FCD Job Name : SCN_T_STFC_ACTINTFEATURE_01 Sub Task Name : Load_consumed Error Message : Login failure Error Code : Unknown	Database Server maybe too many load	✓	If check data in table : [table] by where businessdate that have data then force ok this job	✓ Call Operation Team to Force OK Job	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 34(72%), 6(64%), 47(50%), 43(43%), 42(42%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิตผล (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		
16	Job Type : Load System Name : HP Job Name : SCN_L_HP_EDW_BALHIPORT Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4015	EOF encountered before end of record	✓	Check Number of record in CTL File must match Number of record in DAT File	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 27(88%), 25(66%), 13(65%), 23(64%), 24(64%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง จาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอ (Suggestion)		
17	Job Type : Transform System Name : HP Job Name : SCN_T_HP_BALANCE10_ACCDEMOG_01 Sub Task Name : Close_changed_records Error Message : Target row updated by multiple source rows Error Code : Unknown	View Duplicate key	✘	Check cause that view duplicate	✓ Inform stand by level 2 to make decison	a ให้ 0.3 b ให้ 0.3 c ให้ 0.3 ค่าฐานนิยม = 0.3
Retrieved Case No.(ค่าความคล้ายคลึง) : 18(87%), 8(86%), 14(43%), 42(43%), 17(43%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนกรอที่เอลลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ				ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)	สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
18	Job Type : Load System Name : INS Job Name : SCN_L_INS_EDW_POLICY3 Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4019	Requested file not found : DAT File not found	✓	Check File In Server	✓ Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 25(89%), 13(66%), 27(66%), 23(65%), 24(65%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
19	Job Type : Load System Name : CC Profile Job Name : SCN_L_CC_EDW_PFTXN Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : 2794	Data File duplicate key	✓	Sample Data as Field SourceSeq in XXDWKEDW.UV_[System Name]_[Ses_no]	✓ Inform stand by level 2 and Call source system team for check file as Sample Data	a ให้ 1 b ให้ 1 c ให้ 1 ค่านิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 48(97%), 12(89%), 46(66%), 16(66%), 15(66%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ				ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)	สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	
20	Job Type : Load System Name : CSM Job Name : SCN_L_CSM_EDW_CUSTOMER_SEGMENT Sub Task Name : Run_MultiLoad Error Message : Not enough data in vartext record Error Code : Unknown	Data File is not completed or have Enter in content of data.	✓	Check data in DAT file	✓ Open Incident to Source System For Fixed File and Next Call Operation For Rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 37(87%), 23(64%), 27(46%), 4(43%), 45(43%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เอลลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง จาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		
21	Job Type : Load System Name : CCAP Job Name : SCN_L_CCAP_EDW_REFHP Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4016	data item too large for field	✓	compare length field in source layout spec , data file and structure table.They are must match	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 13(88%), 25(66%), 27(66%), 23(65%), 24(65%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนกรอที่เผลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูก ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
22	Job Type : Load System Name : IVR Job Name : SCN_L_IVR_EDW_IVRTRANS Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : 2793	Data File duplicate row	✓	Sample data record as Field SourceSeq in XXDWKEDW.UV_CCAP_[Ses_no]	✗ Inform stand by level 2 and Call source system team for check file	a ให้ 0.5 b ให้ 0.5 c ให้ 0.5 ค่านานนิยม = 0.5
Retrieved Case No.(ค่าความคล้ายคลึง) : 15(88%), 16(66%), 46(66%), 48(66%), 12(66%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการอีทีแอลที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
23	Job Type : Load System Name : BI Job Name : SCN_L_BI_EDW_BRANCH_CNSP Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4016	data item too large for field	✓	compare length field in source layout spec , data file and structure table.They are must match	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 13(88%), 25(66%), 27(65%), 23(64%), 24(64%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่ผิดพลาดที่ผลิตผล (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
24	Job Type : Load System Name : RB2C Job Name : SCN_L_RB_EDW_SCBOCCODE Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4015	EOF encountered before end of record	✓	Check Number of record in CTL File must match Number of record in DAT File	✓ Inform stand by level 2 and Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 27(88%), 25(66%), 13(66%), 23(65%), 24(65%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
25	Job Type : Load System Name : BBG Job Name : SCN_L_BBG_EDW_LEAD_HIS Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation Error Message : UV Tolerance Error Code : 6706	The string contains an untranslatable character	✓	1.Sample data record as Field SourceSeq in XXDWKEDW.UV_CCAP_[Ses_no] 2.Check Field and Sample data that is latin but field in data file is unicode and compare source layout spec	✗ Inform stand by level 2 and Call source system team for check file	a ให้ 0.5 b ให้ 0.5 c ให้ 0.5 ค่านานิยม = 0.5
Retrieved Case No.(ค่าความคล้ายคลึง) : 16(88%), 46(66%), 48(66%), 12(66%), 15(66%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		
26	Job Type : Transform System Name : FACT Job Name : SCN_T_FACT_FCD_WITHDRAWDEPOSIT_TX_D Sub Task Name : Transform_Complete_ Process_Done Error Message : NameError: rs Error Code : Unknown	ODI Server maybe too many load	✓	If check data in table : [table] by where businessdate that have data then force ok this job	✓ Call Operation Team to Force OK Job	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 5(60%), 18(42%), 1(42%), 34(42%), 17(42%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อแนะนำ (Suggestion)		
27	Job Type : Bmap System Name : CFE Job Name : SCN_T_CFE_CBBTXN110_M43850 Sub Task Name : CTLFW_AllLoadedFor_Process Error Message : Not all the source files required Error Code : Unknown	Source file load not complete	✓	Check Source File [filename] in CTLFW_FILE_PROCESS already Load complete? If Not Complete then Rerun [filename] File	✓ Wait Load File Job complete and Call Operation term for rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 10(80%), 14(74%), 29(42%), 31(42%), 6(30%)						

ตารางที่ 5.1 (ต่อ)

กรณี ร	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)	ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		
28	Job Type : Load System Name : INV Job Name : SCN_L_INV_EDW_CEFMFUND Sub Task Name : Run_MultiLoad Error Message : Not enough data in vartext record Error Code : Unknown	Data File is not completed or have Enter in content of data.	✓	Check data in DAT file	✓ Open Incident to Source System For Fixed File and Next Call Operation For Rerun	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 37(98%), 23(64%), 36(44%), 27(44%), 13(44%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนการที่เผลอที่ผิดพลาด (Problem)	ระบบให้คำแนะนำ			ค่าความถูกต้อง ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
29	Job Type : Load System Name : TR Job Name : SCN_L_TR_COLLTRAL Sub Task Name : Run_MultiLoad Error Message : Load Error Error Code : UTY4019	Requested file not found : DAT File not found	✓	Check File In Server	✓ Call source system team for check file	a ให้ 1 b ให้ 1 c ให้ 1 ค่าฐานนิยม = 1
Retrieved Case No.(ค่าความคล้ายคลึง) : 25(87%), 13(65%), 23(65%), 24(65%), 27(65%)						

ตารางที่ 5.1 (ต่อ)

กรณี ที่	กระบวนกรอที่เผลดที่ผลิตพลาต (Problem)	ระบบให้คำแนะนำ			ค่าความถูก ต้องจาก ผู้เชี่ยวชาญ 3 ท่าน (a,b,c)	
		คำวินิจฉัย (Diagnosis)		ข้อเสนอแนะ (Suggestion)		สิ่งที่ควรทำเมื่อ พบสาเหตุ (Action)
30	Job Type : BMAP System Name : TR Job Name : SCN_T_TR_SETTLMNT_1_0_M25000 Sub Task Name : CTLFW_AllLoadedFor_Process Error Message : Not all the source files required Error Code : Unknown	Source file load not complete	✓	Check Dependency Job and Register Files wait for the file in view correctly	✗ Wait Load Job complete and Call Operation term for rerun	a ให้ 1 b ให้ 0.5 c ให้ 0.5 ค่าฐานนิยม = 0.5
Retrieved Case No.(ค่าความคล้ายคลึง) : 7(86%), 10(79%), 14(73%), 29(44%), 31(43%)						

จากค่าความถูกต้องของผู้เชี่ยวชาญแต่ละกรณี สามารถอธิบายได้ว่า กรณีที่ได้ผลค่าความถูกต้องเท่ากับศูนย์ เนื่องมาจากปัญหาดังกล่าวเป็นปัญหาที่ระบบไม่เคยเรียนรู้หรือไม่มีอยู่ในระบบมาก่อน ระบบจึงไม่สามารถให้คำแนะนำการแก้ปัญหาที่สามารถเอาไปประยุกต์ใช้ได้ถูกต้อง และมีบางกรณีทดสอบที่ผู้เชี่ยวชาญมีความเห็นของค่าความถูกต้องไม่ตรงกัน เช่นกรณีที่ 5, 9 และ 30 ซึ่งเป็นกรณีที่ระบบให้ข้อเสนอแนะ (suggestion) ยังไม่ครอบคลุมแนวทางการแก้ไข แต่สามารถใช้ข้อเสนอแนะที่ได้บางส่วนไปประยุกต์ใช้ในการแก้ไขได้บ้าง จึงทำให้ผู้เชี่ยวชาญมีความเห็นไม่ตรงกัน

เมื่อระบบมีการจัดเก็บปัญหาและแนวทางแก้ไขปัญหา และมีการปรับปรุงฐานกรณีให้มีความถูกต้องมากขึ้น ระบบก็จะสามารถให้คำแนะนำที่สามารถเอาไปประยุกต์ใช้ในการแก้ปัญหากระบวนการอีทีแอลได้อย่างถูกต้องและแม่นยำมากขึ้น

จากการทดลองการวัดประสิทธิผลของการให้คำแนะนำการแก้ไขปัญหากลุ่มกระบวนการอีทีแอลที่ผิดพลาดสิบกรณี สามารถคำนวณค่าความถูกต้องรวมได้ดังนี้

$$\begin{aligned} \text{ค่าความถูกต้อง} &= \frac{(0.5+1+1+1+1+1+0+1+1+1+1+1+1+0+1+1+0.3+1+1+1+1+1+0.5+1+1+1+1+0.5)}{10} \times 100 \\ &= 84.3\% \end{aligned}$$

จากสมการพบว่าค่าความถูกต้องรวมที่คำนวณได้เท่ากับแปดสิบสี่เปอร์เซ็นต์จึงถือว่าระบบมีประสิทธิผลที่ดีตามที่ตั้งสมมุติฐานไว้

บทที่ 6

บทสรุปและข้อเสนอแนะ

6.1 สรุปผลการศึกษา

โครงการนี้ได้นำเสนอการพัฒนาระบบให้เหตุผลเชิงกรณีเพื่อให้คำแนะนำหรือเป็นแนวทางในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ของแผนงานคลังข้อมูลขององค์กรธนาคารไทยพาณิชย์ ทำให้ลดเวลาในการหาสาเหตุและแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด และยังเป็น การสร้างฐานความรู้ให้กับองค์กรอีกด้วย โดยนำปัญหาและแนวทางแก้ไขของกระบวนการอีทีแอลที่ผิดพลาด จากกรณีที่เกิดขึ้นจริงและเอกสารบันทึกมาประยุกต์ในการให้คำแนะนำตามวิธีการให้เหตุผลเชิงกรณีที่เป็นการใช้ประสบการณ์ในอดีตมาแก้ไขปัญหาใหม่ที่มีลักษณะคล้ายคลึงกับปัญหาเดิม และระบบสามารถเรียนรู้กรณีของการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาดที่ประสบความสำเร็จในการแก้ปัญหา และจัดเก็บเป็นความรู้ใหม่ในคลังกรณีเพื่อใช้เป็นความรู้ต่อไป โดยโครงการนี้ได้ใช้เครื่องมือพัฒนาระบบให้เหตุผลเชิงกรณี โพรเทจ (Protege) และ ปร็อกอินมายซีบีอาร์ (MyCBR) มาช่วยในการสร้างและออกแบบคลังกรณี และใช้ภาษา เจเอสพี (JSP) เชื่อมต่อกับ เอพีไอ (API) ของมายซีบีอาร์ในการพัฒนาเว็บแอปพลิเคชัน

ผลสรุปจากการทดสอบประสิทธิผลของระบบนั้น ระบบมีประสิทธิผลที่ระดับหนึ่ง โดยระบบสามารถให้คำแนะนำปัญหาและแนวทางแก้ไขที่ผู้เชี่ยวชาญประเมินว่าสามารถนำไปประยุกต์ใช้ในการแก้ไขปัญหากระบวนการอีทีแอลที่ผิดพลาด ได้กว่าแปดสิบสี่เปอร์เซ็นต์ แต่ก็มีบางกรณีทดสอบที่ระบบให้คำแนะนำปัญหาหรือแนวทางแก้ไขไม่ถูกต้อง เนื่องจากคลังกรณียังจัดเก็บกรณีเพื่อใช้เป็นความรู้ยังไม่ครอบคลุมปัญหา แต่เมื่อระบบมีการจัดเก็บกรณีใหม่เพื่อใช้เป็นความรู้มากขึ้น ระบบจะสามารถให้คำแนะนำปัญหาและแนวทางแก้ไขได้ถูกต้องมากยิ่งขึ้น

6.2 ข้อเสนอแนะ

1. เพื่อให้ได้แนวทางการแก้ไขปัญหาที่ถูกต้องที่สุด ควรมีการรวบรวมข้อมูลกรณีที่เกี่ยวข้องกับระบบให้ได้ครอบคลุมโดเมนมากที่สุด เพื่อให้ระบบมีประสิทธิผลในการให้คำแนะนำมากที่สุด

2. สูตรของการวัดค่าความคล้ายคลึงในระดับกรณี ไม่จำเป็นจะต้องเป็นลักษณะของค่าถ่วงน้ำหนักคุณค่าความคล้ายคลึงในระดับคุณลักษณะ แล้วนำมาบวกกันและหารด้วยค่าถ่วงน้ำหนักรวม เสมอไป อาจจะมีการปรับปรุงให้เป็นสูตรอื่นที่สามารถหาค่าความคล้ายคลึงในระดับกรณีที่แม่นยำกว่านี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ค่าถ่วงน้ำหนักของแต่ละคุณลักษณะ น่าจะสามารถเรียนรู้(learning) ได้ ถ้าค่าถ่วงน้ำหนักของคุณลักษณะใดมีน้ำหนักที่ให้ความแม่นยำ ระบบควรมีความสามารถเรียนรู้ที่จะเปลี่ยนค่าถ่วงน้ำหนักได้ด้วยตัวเอง

4. เนื่องจากระบบมีการแบ่งประเภทของผู้ใช้เป็น ประเภทผู้ใช้ทั่วไปและผู้เชี่ยวชาญ ในกรณีที่ผู้ใช้งานทั่วไปคนใดมีการแก้ไขปัญหาที่แม่นยำอย่างสม่ำเสมอ ระบบควรมีความสามารถเรียนรู้ที่จะปรับประเภทของผู้ใช้คนนั้นเป็นผู้เชี่ยวชาญได้

5. ส่วนต่อประสานผู้ใช้ (interface) ของระบบควรจะเป็นลักษณะการประมวลผลข้อความ (Text) เมื่อระบบแนะนำแนวทางการแก้ไขปัญหามาแล้ว ระบบควรนำปัญหามากรอกในกล่องข้อความให้อัตโนมัติ

6. นอกจากการใช้การให้เหตุผลเชิงกรณีในการพัฒนาระบบแล้ว ยังสามารถนำเอาทฤษฎีหรือเทคโนโลยีอื่นๆมาประยุกต์ใช้ในวงจรการทำงานของระบบให้เหตุผลเชิงกรณี เช่น นำ Object Database มาใช้เป็นฐานความรู้ หรือนำเจเนติกอัลกอริทึมมาใช้ในการปรับเปลี่ยนแนวทางแก้ไขให้เหมาะสมมากยิ่งขึ้น เพื่อให้ระบบให้เหตุผลเชิงกรณีมีประสิทธิภาพมากยิ่งขึ้น



บรรณานุกรม

- ณัฐพงษ์ วารีประเสริฐและณรงค์ ลำดี. 2552. **ปัญญาประดิษฐ์**. กรุงเทพฯ: เคทีพี.
- พรฤดี เนติโสภาคกุล. 2554. **เทคโนโลยีการจัดการความรู้**. กรุงเทพฯ: ร่มไหมกซ์.
- สมภพ ตริยะประเสริฐพร. 2546. “การประยุกต์ใช้ case-base reasoning ในการเลือกแพ็คเกจทัวร์.”
โครงการพัฒนาระบบงานวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะ
เทคโนโลยีสารสนเทศ, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- อัครพล เอกวงค์อนันต์. 2548. “การระบุคำไทยและคำทับศัพท์ด้วยแบบจำลองเอ็นแกรม.”
วิทยานิพนธ์ ปริญญาอักษรศาสตรมหาบัณฑิต สาขาวิชาภาษาศาสตร์ คณะอักษรศาสตร์,
จุฬาลงกรณ์มหาวิทยาลัย.
- Stahl, A. and Roth-Berghifer, T. R. 2008. **Rapid Prototyping of CBR Application**. [Online]
Available: http://www.dfki.de/lt/publication_show.php?id=3762.
- Zilles L. 2009. **myCBR Tutorial**. [Online] Available: <http://mycbr-project.net/downloads/mycbrTutorialv1.1.pdf>.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

ตัวอย่างกรณีกระบวนการอีทีแอลที่ผิดพลาดในคลังกรณี

- **Case No. : 1 Job Type : Transform System Name : TF**
Job Name : SCN_T_TF_EVENTPARTYPA_01
Sub Task Name : Assigning_Data
Error Message : Null Pointer
Error Code : Unknown
Diagnosis : ODI Server too many load
Suggestion : Wait a moment and rerun job
Action : Call operation team to rerun job

- **Case No. : 2 Job Type : Load System Name : RB2C**
Job Name : SCN_L_RB2C_EDW_RISKMART_LOAN
Sub Task Name : Read_Control_File_Info_Updat_CSFE
Error Message : File not found
Error Code : Unknown
Diagnosis : File not found - [PATH][FILENAME].CTL
Suggestion : Check File In Server
Action : Call source system team for check file

- **Case No. : 3 Job Type : Export System Name : SFC**
Job Name : SCN_T_SFC_ATM_FULL_01
Sub Task Name : Transform_Complete_Process_Done
Error Message : Login failure for Connection
Error Code : Unknown
Diagnosis : Database Server maybe too many load
**Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job**
Action : Call Operation Team to Force OK Job

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **Case No. : 4 Job Type : Load System Name : CCAP**
Job Name : SCN_L_CCAP_EDW_SYSCHE
Sub Task Name : Run_MultiLoad
Error Message : Error limit exceeded
Error Code : UTY0802
Diagnosis : Column edw_record_no in staging table generate Exceeded.
Suggestion : Change no cycle to cycle of [field] then recreate [table]
Action : Inform stand by level 2 and Call DBA Team to recreate table ,Next Call
Operation Team For Rerun

- **Case No. : 5 Job Type : Export System Name : FACT**
Job Name : SCN_T_FACT_FCD_WITHDRAWDEPOSIT_TX_D
Sub Task Name : Transform_Complete_Process_Done
Error Message : NameError: rs
Error Code : Unknown
Diagnosis : ODI Server maybe too many load
Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job
Action : Call Operation Team to Force OK Job

- **Case No. : 6 Job Type : Transform System Name : CCAP**
Job Name : SCN_T_CP_REFHP10_AGMTAGMTCLA_01
Sub Task Name : Load_Consumed
Error Message : Login failure
Error Code : Unknown
Diagnosis : Database Server maybe too many load
Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job
Action : Call Operation Team to Force OK Job

- **Case No. : 7 Job Type : Bmap System Name : CFE**
Job Name : SCN_T_CFE_CBBTXN110_M43850
Sub Task Name : CTLFW_AllLoadedFor_Process
Error Message : Not all the source files required
Error Code : Unknown
Diagnosis : Source file load not complete
Suggestion : Check Source File [filename] in CTLFW_FILE_PROCESS already Load complete? If Not Complete then Rerun [filename] File
Action : Wait Load File Job complete and Call Operation term for rerun

- **Case No. : 8 Job Type : Transform System Name : ESL**
Job Name : SCN_T_ESL_T3_TR_AGREEMENT
Sub Task Name : Close_changed_records
Error Message : Target row updated by multiple source rows
Error Code : Unknown
Diagnosis : Sierra , Murex Data Duplicate
Suggestion : 1.Find Duplicare record

```

SEL Account_num ,Account_Modifier_Num ,COUNT(*)
FROM P1DWKEDW.VESL_T3_TR_AGREEMENT S
GROUP BY 1,2 HAVING COUNT(*) > 1

```

2. Backup Data in Table T3_TREASURY_AGREEMENT
3.Clear ALL Data in Table and Rerun
Action : Call Operation Team to Rerun Job

- **Case No. : 9 Job Type : SetDate System Name : IPP**
Job Name : SCN_IPP_SET_BUSINESSDATE_W
Sub Task Name : Set_BusinessDate
Error Message : Invalid date supplied for CTLFW_BUSINESSDATE.BusinessDate
Error Code : Unknown
Diagnosis : Maybe config Packages of this SCN incorrect
Suggestion : Please check BizDt option type should be “Declear variable”
Action : Regenerate Scenario and Redeploy

- **Case No. : 10 Job Type : Bkey System Name : MX**
Job Name : SCN_T_MX_SECFEATE10_BK_FEATURE_23
Sub Task Name : CTLFW_AllLoadedFor_Process
Error Message : Not all the source files required
Error Code : Unknown
Diagnosis : Source file load not complete
Suggestion : Check Source File [filename] in CTLFW_FILE_PROCESS already Load complete? If Not Complete then Rerun [filename] File
Action : Wait Load File Job complete and Call Operation term for rerun

- **Case No. : 11 Job Type : Transform System Name : ESL**
Job Name : SCN_T_ESL_T3_PARTY_TELEPHONE
Sub Task Name : Insert_flow_I\$_into_table
Error Message : The format or data contains a bad character
Error Code : Unknown
Diagnosis : cannot convert datatype that cast at field in view.
Suggestion : Check Field in view [viewname] maybe has convert numeric-to-character conversion or character-to-numeric.
Action : Inform stand by level 2 to make decision

- **Case No. : 15 Job Type : Load System Name : MX**

Job Name : SCN_L_MX_EDW_SECFEATE

Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation

Error Message : UV Tolerance

Error Code : 2793

Diagnosis : Data File duplicate row

Suggestion : Sample data record as Field SourceSeq in

XXDWKEDW.UV_CCAP_[Ses_no]

Action : Inform stand by level 2 and Call source system team for check file

- **Case No. : 16 Job Type : Load System Name : TR**

Job Name : SCN_L_TR_EDW_LIMITPTY

Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation

Error Message : UV Tolerance

Error Code : 6706

Diagnosis : The string contains an untranslatable character

Suggestion : 1.Sample data record as Field SourceSeq in

XXDWKEDW.UV_[Systemname]_[Ses_no]

2.Check Field and Sample data that is latin but field in data file is unicode
and compare source layout spec

Action : Inform stand by level 2 and Call source system team for check file

- **Case No. : 17 Job Type : Transform System Name : TR**
Job Name : SCN_T_TR_COMMON10_ACCOPRODHIST_01
Sub Task Name : Insert_flow_I\$_into_table
Error Message : Cannot place a null value in a NOT NULL field
Error Code : Unknown
Diagnosis : Bkey use has not in Bkey Gen
Suggestion : Sample Data and Check All ID FIELD that has value is NOT NULL in Tier2 table compare field in view that has value is null
Action : Inform stand by level 2 and Call source system team for check file

- **Case No. : 18 Job Type : Transform System Name : TR**
Job Name : SCN_T_TR_ACCBALSUMDD_02
Sub Task Name : Close_changed_records
Error Message : Target row updated by multiple source rows
Error Code : Unknown
Diagnosis : View Duplicate key
Suggestion : Check cause that view duplicate
Action : Inform stand by level 2 to make decision

- **Case No. : 19 Job Type : Export System Name : BKL**
Job Name : SCN_E_BKL_PERSON_INFO
Sub Task Name : Report_Number
Error Message : TypeError: __add__ nor __radd__
Error Code : Unknown
Diagnosis : Script jython source side not pass parameter to target side
Suggestion : Recompile scenario verion 11g and redeploy on production
Action : Call stand by level 2 to make decision

- **Case No. : 20 Job Type : Export System Name : CRM**
Job Name : SCN_E_CRM_TRANS_INC_EXP_M
Sub Task Name : Run_fastexport_script
Error Message : FastExport Error
Error Code : Unknown
Diagnosis : Export File maybe No more spool space.
Suggestion : See Export Log in EXPORTS/LOGS
Action : Unknown

- **Case No. : 21 Job Type : Export System Name : CRM**
Job Name : SCN_E_CRM_TRANS_INC_EXP_M
Sub Task Name : Run_fastexport_script
Error Message : FastExport Error
Error Code : 2646
Diagnosis : No more spool space
Suggestion : Call DBA team for check and tuning Performance of view
Action : Inform stand by level 2 and Call DBA Team

- **Case No. : 22 Job Type : Load System Name : BI**
Job Name : SCN_L_BI_EDW_EXT_PBG_ACCOUNT
Sub Task Name : Read_Control_File_Info_Updat_CSFE
Error Message : Invalid date supplied for CTLFW_System_File_Extract.ExtractionDate
Error Code : 2666
Diagnosis : Field Date in contorl file incorrent format.
Suggestion : Check Data in Control File
Action : Inform stand by level 2 and Call source system team for check file

- **Case No. : 23 Job Type : Load System Name : ALM**

Job Name : SCN_L_ALM_RCPARMET

Sub Task Name : Run_MultiLoad

Error Message : Load Error

Error Code : Unknown

Diagnosis : Server maybe too many load/unload tasks running

Suggestion : See Load Log file in LOADS\LOGS

Action : Unknown

- **Case No. : 24 Job Type : Load System Name : ALM**

Job Name : SCN_L_ALM_RCPARMET

Sub Task Name : Run_MultiLoad

Error Message : Load Error

Error Code : 2633

Diagnosis : Server maybe too many load/unload tasks running

Suggestion : Wait a moment and rerun job

Action : Call operation term to rerun job.

- **Case No. : 25 Job Type : Load System Name : PMS**

Job Name : SCN_L_PMS_EDW_BMAP

Sub Task Name : Run_MultiLoad

Error Message : Load Error

Error Code : UTY4019

Diagnosis : Requested file not found : DAT File not found

Suggestion : Check File In Server

Action : Call source system team for check file

- **Case No. : 26 Job Type : Transform System Name : AM**
Job Name : SCN_T_AM_POINTLOCATIO_00
Sub Task Name : Unknown
Error Message : Control-M Error But Not show on ODI
Error Code : Unknown
Diagnosis : Server ODI too many load.
Suggestion : Check Ctlfw_transform & Data in Table
Action : Inform level 2 to make decision

- **Case No. : 27 Job Type : Load System Name : ML**
Job Name : SCN_L_ML_EDW_CUSTOMER
Sub Task Name : Run_MultiLoad
Error Message : Load Error
Error Code : UTY4015
Diagnosis : EOF encountered before end of record
Suggestion : Check Number of record in CTL File must match Number of record in DAT File
Action : Inform stand by level 2 and Call source system team for check file

- **Case No. : 28 Job Type : Export System Name : AML**
Job Name : SCN_T_AML_AML_MAIN
Sub Task Name : Deleted_Data_From_Table
Error Message : A login timeout occurred
Error Code : Unknown
Diagnosis : Database Server too many load
Suggestion : Wait a moment and Rerun job.
Action : Call Operation team for rerun

- **Case No. : 29 Job Type : Bmap System Name : TF**
Job Name : SCN_T_TF_IMCOFL_1_0_M09700
Sub Task Name : Insert_Reference_Code
Error Message : NameError: ChkCnt
Error Code : Unknown
Diagnosis : Script jython source side not pass parameter to target side
Suggestion : Rerun Job
Action : Call operation to rerun

- **Case No. : 30 Job Type : SetDate System Name : SL**
Job Name : SCN_SL_SET_BUSINESSDATE_D
Sub Task Name : Set_BusinessDate
Error Message : NameError: s
Error Code : Unknown
Diagnosis : Script jython source side not pass parameter to target side
Suggestion : Rerun Job
Action : Call operation to rerun

- **Case No. : 31 Job Type : Bmap System Name : LOS**
Job Name : SCN_T_LOS_VHCLBDT_1_0_M04700
Sub Task Name : Initialise_for_Control_Framework
Error Message : Login failure for Connection
Error Code : Unknown
Diagnosis : Database Server too many load
Suggestion : Wait a moment and Rerun Job.
Action : Call Operation team to rerun

- **Case No. : 32 Job Type : Export System Name : DMS**
Job Name : SCN_T_DMS_DM_LAR_LOAN_01
Sub Task Name : Insert_Data_to_Table
Error Message : Socket communication failure
Error Code : Unknown
Diagnosis : Database Server too many load
Suggestion : Wait a moment and Rerun Job
Action : Call Operation Team to Rerun Job

- **Case No. : 33 Job Type : Export System Name : CRM**
Job Name : SCN_T_CRM_PRJ_07_ACCT_BAL_WO_OD
Sub Task Name : Is_the_process_already_executed
Error Message : NameError : BusinessDate
Error Code : Unknown
Diagnosis : Script jython source side not pass parameter to target side
Suggestion : Rerun Job
Action : Call Operatoin team to Rerun Job

- **Case No. : 34 Job Type : Transform System Name : FCD**
Job Name : SCN_T_IMFC_EVENT_01
Sub Task Name : Load_Consumed
Error Message : Login failure
Error Code : Unknown
Diagnosis : Database Server maybe too many load
**Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job**
Action : Call Operation Team to Force OK Job

- **Case No. : 35 Job Type : Export System Name : FACT**
Job Name : SCN_T_FACT_TAXINT1_ACCT_J1
Sub Task Name : Is_the_process_already_executed
Error Message : already been executed
Error Code : Unknown
Diagnosis : This Scenario already been executed
**Suggestion : Please cleanup S1DCFEDW.CTLFW_Transform for this process on the
businessdate and also delete data in table**
Action : Call Operation Team to Rerun Job

- **Case No. : 36 Job Type : Load System Name : IX**
Job Name : SCN_L_IX_EDW_IXQDWACN
Sub Task Name : Insert_into_Event_Load_Current
**Error Message : user does not have SELECT WITH GRANT OPTION access to
P1VCFEDW.CTLFW_BUSINESSDATE**
Error Code : Unknown
Diagnosis : System no have authorized for select view
Suggestion : Call DBA Team for grant authorize
Action : Call DBA Team for grant authorize and Call Operation Team to Rerun Job

- **Case No. : 37 Job Type : Load System Name : INV**
Job Name : SCN_L_INV_EDW_MFUND
Sub Task Name : Run_MultiLoad
Error Message : Not enough data in vartext record
Error Code : Unknown
Diagnosis : Data File is not completed or have Enter in content of data.
Suggestion : Check data in DAT file
**Action : Open Incident to Source System For Fixed File and Next Call Operation
For Rerun**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **Case No. : 38 Job Type : Load System Name : BBG**
Job Name : SCN_L_BBG_EDW_WIN
Sub Task Name : ET,_UV,_ACQ_APP_Reconciliation
Error Message : Invalid date
Error Code : 2665
Diagnosis : Data of column that has datatype is date not valid.
Suggestion : Check data in DAT file and Sample data
Action : Call Source system for Fixed File and Next Call Operation For Rerun

- **Case No. : 39 Job Type : Export System Name : PCI**
Job Name : SCN_E_PCI_CARD_NO
Sub Task Name : Run_fastexport_script
Error Message : Database TD_VTPXXX does not exist
Error Code : 3802
Diagnosis : Forgot add topology
Suggestion : Check SP Team,Do SP team add topology?
Action : Wait SP Team add topology and Call Operation Team For Rerun

- **Case No. : 40 Job Type : Load System Name : HRMS**
Job Name : SCN_L_HRMS_EDW_HDC
Sub Task Name : Read_Control_File_Info_Updat_CSFE
Error Message : File not found
Error Code : Unknown
Diagnosis : User ไม่ได้วาง .Dat File ที่ Server
Suggestion : Check File In Server
Action : 1) แจ้ง User ให้วาง File 2) แจ้ง Operation rerun

- **Case No. : 41 Job Type : Transfrom System Name : AM**
Job Name : SCN_T_AM_ACCOCOLLSTAT_CP
Sub Task Name : Transform_Complete_Process_Done
Error Message : The Network Adapter could not establish the connection
Error Code : Unknown
Diagnosis : The Network Adapter could not establish the connection
**Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job**
Action : Call Operation Team to Force OK Job

- **Case No. : 42 Job Type : Transfrom System Name : IM**
Job Name : SCN_T_IMSTPM10_ACCOUNTINSTRU_01
Sub Task Name : Load_Consumed
Error Message : Login failure for Connection
Error Code : Unknown
Diagnosis : Database Server too many load
**Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job**
Action : Call Operation Team to Force OK Job

- **Case No. : 43 Job Type : Transfrom System Name : NROL**
Job Name : SCN_T_NROLCCRD_POINTLOCA_01
Sub Task Name : Insert_new_records
Error Message : Login failure for Connection
Error Code : Unknown
Diagnosis : Database Server too many load
**Suggestion : If check data in table : [table] by where businessdate that have data then
force ok this job else rerun this job**
Action : Call Operation Team to Rerun Job

- **Case No. : 47 Job Type : Transform System Name : FCD**

Job Name : SCN_T_IMFC_ACCTACCTRELA_01

Sub Task Name : Load_Consumed

Error Message : Login failure for Connection

Error Code : Unknown

Diagnosis : Database Server too many load

Suggestion : If check data in table : [table] by where businessdate that have data then

force ok this job

Action : Call Operation Team to Force OK Job

- **Case No. : 48 Job Type : Load System Name : CC**

Job Name : SCN_L_CC_EDW_NEWCARD

Sub Task Name : ET_UV,_ACQ_APP_Reconciliation

Error Message : UV Tolerance

Error Code : 2794

Diagnosis : Data File duplicate key

Suggestion : Sample Data as Field SourceSeq in

XXDWKEDW.UV_[Systemname]_[Ses_no]

Action : Inform stand by level 2 and Call source system team for check file as Sample Data

- **Case No. : 49 Job Type : Export System Name : AML**

Job Name : SCN_E_AML_TF_IMISFL

Sub Task Name : Run_fastexport_script

Error Message : FastExport Error: See [Log]

Error Code : Unknown

Diagnosis : See error log file but not found log file

Suggestion : Check log file [Log] but not found log file

Action : Call Operation Team to Rerun OK Job

- **Case No. :** 50 **Job Type :** Export **System Name :** AML

Job Name : SCN_E_AML_FRC_FRCOBKCD

Sub Task Name : Run_fastexport_script

Error Message : FastExport Error: See [Log]

Error Code : Unknown

Diagnosis : See error log file but not found log file

Suggestion : Check log file [Log] but not found log file

Action : Call Operation Team to Rerun OK Job



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้