



บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

แนวคิดการรวมแบบจำลองข้อมูล (Data Model Integration) มีที่มาจากการรวมสกีมา¹ (Schema Integration) [1] ซึ่งเป็นการรวมกันของสกีมาจาก 2 ฐานข้อมูลที่มีความเกี่ยวข้องกัน หรือเป็นเรื่องเดียวกัน แต่มีความแตกต่างกันเนื่องจากถูกออกแบบโดยนักออกแบบฐานข้อมูลคนละคนกัน ผลลัพธ์ของการรวมจะได้เป็นสกีมารวม (Integrated Schema) ซึ่งเปรียบเสมือนส่วนต่อประสาน (Interface) ให้ผู้ใช้สามารถเข้าถึงข้อมูลจากฐานข้อมูลทั้งสองตัวที่แตกต่างกันได้พร้อมกัน ในการรวมสกีมา สกีมาทั้งสองจะอยู่ในรูปของแบบจำลองข้อมูลประเภทเดียวกัน² เช่น อยู่ในแบบจำลองข้อมูลเชิงวัตถุ (Object-Oriented Data Model) เหมือนกัน แต่อย่างไรก็ตามแบบจำลองข้อมูลทั้งสองยังมีความแตกต่างกันทางความหมาย (Semantic Heterogeneities) อยู่ ซึ่งเกิดจากการที่นักออกแบบฐานข้อมูลของแต่ละระบบออกแบบฐานข้อมูลในเรื่องเดียวกันแตกต่างกันไป อันส่งผลให้แบบจำลองข้อมูลของแต่ละระบบฐานข้อมูลมีความแตกต่างกัน

ความแตกต่างทางความหมายนั้นมีหลายรูปแบบ [3] ได้แก่

- ความขัดแย้งของชื่อ (Name Conflicts) ซึ่งเกิดได้ 2 ลักษณะคือชื่อต่างกันแต่ความหมายเดียวกัน (Synonyms) เช่น คำว่า 'Customer' กับ 'Buyers' และชื่อเหมือนกันแต่ความหมายต่างกัน (Homonyms) เช่น 'Address' อาจเป็นที่อยู่ที่ทำงานหรือที่อยู่ที่บ้าน เป็นต้น
- ความขัดแย้งของโครงสร้าง (Structure Conflicts) เช่น ในแบบจำลองเชิงสัมพันธ์กับเอนทิตี ระบบหนึ่งอาจออกแบบวัตถุหนึ่งเป็นเอนทิตี แต่อีกระบบอาจออกแบบเป็นแอตทริบิวต์
- ความขัดแย้งของหน่วยที่ใช้ (Scaling Conflicts) เช่น หน่วยของเงิน ระบบหนึ่งอาจใช้หน่วยเงินเป็นดอลลาร์ แต่อีกระบบอาจใช้หน่วยเงินเป็นบาท

¹ สกีมา (Schema) เป็นส่วนหนึ่งของระบบฐานข้อมูลที่อธิบายข้อมูล ซึ่งจะมีรายละเอียดแตกต่างกันตามแบบจำลองข้อมูลที่ไว้ เช่น ในแบบจำลองเชิงสัมพันธ์กับเอนทิตี (Entity-Relationship Model) [2] สกีมาจะบอกคำจำกัดความของเอนทิตี (Entity Definition) ซึ่งประกอบไปด้วยแอตทริบิวต์ (Attribute) อะไร ข้อมูลที่ไว้กับเป็นประเภทใดและระดับความสัมพันธ์ระหว่างเอนทิตี

² หากสกีมาทั้งสองยังอยู่ในแบบจำลองข้อมูลคนละประเภท จะต้องทำการแปลงให้อยู่ในประเภทเดียวกันก่อน เรียกว่าการทำการแปลงสกีมา (Schema Translation)

ในการรวมแบบจำลองข้อมูลนั้น จะต้องพิจารณาความสัมพันธ์ระหว่างแบบจำลองข้อมูล ทั้งในแง่ความเหมือนกันและแตกต่างกันในแง่มุมต่างๆ จากนั้นจึงจะทำการรวมให้เป็นแบบจำลองข้อมูลโดยรวม (Global Data Model) ได้

จากงานวิจัยที่ผ่านมาด้านการรวมสกีมาของฐานข้อมูล มีการใช้วิธีต่างๆ เพื่อช่วยในการ รวมความสัมพันธ์ระหว่างสกีมา เช่น งานวิจัย [4] ได้ทำการสร้างโครงสร้างของแนวคิด (Concept Hierarchy) เพื่อแสดงความสัมพันธ์ระหว่างแนวคิด (Concept) ซึ่งคือ ความสัมพันธ์ลักษณะทั่วไป กับลักษณะเฉพาะ (Generalization/Specialization) โดยจะมีการพิจารณาร่วมคลา (Class) หรือ เอนทิตีที่มีลักษณะต่างๆ ในแต่ละสกีมาตรงกับแนวคิดใด ผลที่ได้ก็จะเป็นผลของความสัมพันธ์ ระหว่างแนวคิดซึ่งใช้แทนสกีมารวม ในงานวิจัย [5] ได้นำหลักการใช้ภาษามาช่วยในการเปรียบเทียบ เพื่อหาความสัมพันธ์ภายในสกีมาและระหว่างสกีมา โดยมีข้อสันนิษฐานว่าความสัมพันธ์ ระหว่างคลาหรือเอนทิตีที่อยู่ในรูปของความสัมพันธ์ระหว่างคำในประโยค เช่น คลา Teacher Student และ Subject มีความสัมพันธ์กันเพราะสามารถอยู่ในรูปประโยค Teacher teach Student และ Teacher teach Subject โดยจะต้องใช้พจนานุกรม (Dictionary) มาเป็นตัวช่วยในการพัฒนาด้วย ซึ่งแนวคิดในการนำฐานความรู้ (Knowledge Base) มาใช้ในลักษณะนี้ไม่สามารถครอบคลุมการรวมสกีมาของทุกระบบ เพราะขึ้นกับความหลากหลายของข้อมูลในฐานความรู้ แต่อย่างไรก็ดีก็มีส่วนช่วยให้การรวมสกีมาเป็นไปอย่างอัตโนมัติ

งานวิจัย [6,7] มีแนวคิดในการนำคำวิทยาการศึกษาสำนึก (Heuristics) มาช่วยในการ วิเคราะห์เปรียบเทียบเพื่อหาความสัมพันธ์ระหว่างแบบจำลองข้อมูล และ [3,8,9] มีแนวคิดในการ รวมสกีมาของแบบจำลองเชิงวัตถุโดยมีการพิจารณาเฉพาะการรวมกันของโครงสร้างที่แสดง การสืบทอดคุณลักษณะ (Inheritance Hierarchies) ในทั้ง 2 สกีมาเท่านั้น ซึ่งผู้วิจัยมีความเห็นว่า นอกจากความสัมพันธ์ในลักษณะการสืบทอดคุณลักษณะ (Inheritance) แล้ว ในการรวมแบบ จำลองเชิงวัตถุควรที่จะมีการพิจารณาความสัมพันธ์แบบเอกกริภขน์ (Aggregation) และความสัมพันธ์แบบพี่น้อง (Sibling) เพิ่มเติมด้วย

การรวมสกีมานั้นจะทำได้ก็ต่อเมื่อแต่ละแบบจำลองข้อมูลมีความหมาย (Semantics) ที่ ชัดเจน แต่ในความเป็นจริงผู้ออกแบบเท่านั้นที่จะสามารถเข้าใจความหมายของสกีมาได้ชัดเจน ดังนั้นงานวิจัยทั้งหลายที่พยายามจะทำการรวมสกีมาอย่างอัตโนมัติจึงยังคงต้องการให้ผู้รู้ระบบ ความหมายเพิ่มเติมอันจะมีส่วนช่วยให้ขั้นตอนของการรวมสกีมาทำได้ถูกต้อง

ในส่วนของวิทยานิพนธ์นี้ ได้นำเสนอการออกแบบวิธีการรวมแบบจำลองข้อมูลเชิงวัตถุ โดยปรับปรุงและเพิ่มเติมจากงานวิจัย [3,7] โดยจะมีการพิจารณาความสัมพันธ์การสืบทอดคุณ ลักษณะและความสัมพันธ์แบบเอกกริภขน์ รวมทั้งจะพิจารณาความสัมพันธ์แบบพี่น้องร่วมด้วย พร้อมทั้งจะนำวิทยาการศึกษาสำนึกดังใน [7] มาปรับใช้เพื่อพิจารณาระดับความเหมือนกันของ

คลาสจากทั้ง 2 แบบจำลองข้อมูลด้วย ในวิทยานิพนธ์นี้จะกำหนดให้ผู้ทำการรวมแบบจำลองข้อมูลสามารถระบุความหมายเพิ่มเติมให้กับข้อมูลเพื่อให้การรวมได้ผลถูกต้องมากขึ้นและผู้วิจัยจะทำการสร้างต้นแบบสำหรับการรวมแบบจำลองข้อมูลเพื่อทดสอบวิธีการที่ได้ออกแบบไปด้วย

1.2 วัตถุประสงค์ของการวิจัย

เพื่อออกแบบวิธีการรวมแบบจำลองข้อมูลเชิงวัตถุ โดยจะพิจารณาปัญหาความแตกต่างทางความหมายในเรื่องของชื่อ โครงสร้าง และหน่วยที่ใช้ ดังที่กล่าวมาแล้วข้างต้น โดยใช้วิทยาการศึกษาสำนึกและการวิเคราะห์ความสัมพันธ์ของคลาสแบบสืบทอดคุณลักษณะ แบบเอกกรีเกชันและแบบพีน้อง เพื่อช่วยในการรวม ในการนี้จะมีการพัฒนาต้นแบบสำหรับรวมแบบจำลองตามวิธีที่ออกแบบเพื่อทดสอบด้วยว่าวิธีการที่ได้ออกแบบไว้นั้นมีประสิทธิภาพในระดับที่ยอมรับได้

1.3 ขอบเขตของการวิจัย

1. วิธีการที่ออกแบบจะใช้สำหรับการรวมแบบจำลองข้อมูลเชิงวัตถุ 2 ชุดเข้าด้วยกัน โดยไม่พิจารณาเมธอด (Method) ของวัตถุ
2. ผู้ที่จะใช้วิธีการที่ออกแบบเพื่อทำการรวมแบบจำลองข้อมูล คือผู้ที่ทำการออกแบบแบบจำลองข้อมูลทั้งสองหรือผู้ที่รู้รายละเอียดของแบบจำลองข้อมูลทั้งสองซึ่งอาจมาจากการศึกษาเอกสารของแต่ละระบบ ทั้งนี้เพราะผู้รวมจะต้องรู้ถึงความหมายต่างๆ ที่แฝงอยู่ในแบบจำลอง และจะต้องระบุออกมาให้ถูกต้องชัดเจนเพื่อประโยชน์ในการรวม
3. ในการเลือกแบบจำลองข้อมูลใดๆ เพื่อทำการรวมกันเป็นแบบจำลองข้อมูลรวมอันใหม่ นั้นผู้ทำการรวมจะเป็นผู้พิจารณาว่าแบบจำลองข้อมูลทั้งสองมีความเกี่ยวข้องกันหรือเป็นเรื่องเดียวกัน และเหมาะสมที่จะนำมารวมกันได้
4. ข้อมูลเข้าและผลลัพธ์สำหรับการทดสอบการรวมโดยต้นแบบเป็นข้อมูลในแผนภาพคลาสของแบบจำลองซึ่งอาจอยู่ในรูปแบบของแฟ้มข้อความ (Text File) ซึ่งเก็บรายละเอียดของคลาส อันได้แก่ ชื่อคลาส แอตทริบิวท์ของคลาส ความสัมพันธ์แบบสืบทอดคุณลักษณะ เอกกรีเกชัน และแบบพีน้อง และมีรายละเอียดของความหมายเพิ่มเติมอยู่ด้วย เป็นต้น
5. ในการวัดผลจะใช้ต้นแบบเพื่อทำการสร้างแบบจำลองข้อมูลรวมจากแบบจำลองข้อมูลเข้า ซึ่งมีลักษณะตามหลักการของแบบจำลองข้อมูลเชิงวัตถุที่ดี อย่างน้อย 2 ตัวอย่าง ซึ่งจะต้องครอบคลุมทุกลักษณะของความสัมพันธ์ที่พิจารณา คือความสัมพันธ์แบบสืบทอดคุณลักษณะ แบบเอกกรีเกชันและแบบพีน้อง และมีความหลากหลายของระดับความสัมพันธ์ด้วย คือ แบบเท่าเทียมกัน (Equivalence), แบบเป็นส่วนหนึ่ง (Inclusion), แบบมีส่วนเหมือนกันมาก (Tight-intersect), แบบมีส่วนเหมือนกันบ้าง (Loose-intersect) และ

แบบแยกต่างหาก (Disjoint) โดยในแต่ละตัวอย่างจะมีการสร้างสกีมา 2 จุดบนฐานข้อมูลจริงตามแบบจำลองข้อมูลเข้าทั้งสอง ส่วนแบบจำลองรวมที่ได้จากต้นแบบจะนำไปสร้างเป็นสกีมารวม การทดสอบจะทำโดยการสอบถามข้อมูล (Query) จากสกีมารวมเพื่อดูว่าผลที่ได้ครบถ้วนตามที่ควรจะได้จากการสอบถามข้อมูลไปยังสกีมาดั้งเดิม 2 จุดโดยตรงหรือไม่ นอกจากนั้นจะทำการพิจารณาว่าแบบจำลองรวมที่ได้ มีลักษณะตรงตามหลักการออกแบบแบบจำลองเชิงวัตถุที่ดีด้วยหรือไม่

1.4 ขั้นตอนและวิธีดำเนินการวิจัย

1. วิเคราะห์และออกแบบวิธีการในการรวมแบบจำลองข้อมูล
2. วิเคราะห์ออกแบบและพัฒนาต้นแบบ
3. ทดสอบและปรับปรุงวิธีการรวมแบบจำลองข้อมูลโดยอาศัยต้นแบบและวัดผล
4. สรุปผลการวิจัยและข้อเสนอแนะ
5. จัดทำรายงานวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้วิธีการใหม่สำหรับการรวมแบบจำลองข้อมูลเชิงวัตถุ ซึ่งวิธีการนี้จะเป็นเพียงอีกทางเลือกหนึ่งสำหรับการรวม
2. สามารถนำวิธีการรวมที่ได้ไปปรับใช้กับการรวมสกีมาของฐานข้อมูลหรือการรวม 2 ระบบงานเชิงวัตถุ
3. ได้ต้นแบบสำหรับการรวมแบบจำลองข้อมูล ซึ่งสามารถนำไปพัฒนาให้สมบูรณ์ขึ้นได้
4. ผลพลอยได้จากการวิเคราะห์ความสัมพันธ์ระหว่างแบบจำลองข้อมูล คือข้อเสนอแนะซึ่งจะบอกผู้ทำการรวมว่า แบบจำลองข้อมูลเชิงวัตถุ 2 ชุดสามารถรวมกันได้หรือไม่ และมากน้อยเพียงใด

1.6 บทความที่ตีพิมพ์จากการวิจัย

ส่วนหนึ่งของผลงานจากการวิจัยนี้ได้รับการตีพิมพ์และนำเสนอในงานประชุมวิชาการวิทยาศาสตร์คอมพิวเตอร์และวิศวกรรมคอมพิวเตอร์แห่งชาติ ปี 2542 (The National Computer Science and Engineering Conference: NCSEC'99) โดยมหาวิทยาลัยอีสต์ลันซิงก์ เมื่อวันที่ 15 - 17 ธันวาคม พ.ศ.2542 ในบทความเรื่อง Integrating OODB Schemas Using Heuristics and Richer Semantics โดย นางสาวสุพัตรา สวัสดิ์รักเกียรติ และ อ.ดร.ทวีชัย เสนิงค์ ณ อัญญา

1.7 เนื้อหาของวิทยานิพนธ์

เนื้อหาของวิทยานิพนธ์ฉบับนี้จะแบ่งออกเป็น 7 บทและภาคผนวกอีก 5 บท โดยบทที่ 1 ที่ได้กล่าวมาแล้วเป็นบทนำ บทที่ 2 จะกล่าวถึงทฤษฎีและงานวิจัยต่างๆ ที่เกี่ยวข้องกับวิทยานิพนธ์นี้ บทที่ 3 จะกล่าวถึงรูปแบบของแบบจำลองข้อมูลที่นำมาใช้ในการทดสอบเพื่อทำการรวมและขั้นตอนวิธีการเปรียบเทียบ จากนั้นในบทที่ 4 จะกล่าวถึงขั้นตอนวิธีการรวมแบบจำลองข้อมูลอย่างละเอียดตามลำดับ บทที่ 5 จะกล่าวถึงต้นแบบสำหรับการรวมแบบจำลองข้อมูลที่ได้พัฒนาขึ้นจากวิธีในการรวมแบบจำลองข้อมูลตามที่ได้เสนอไว้สำหรับวิทยานิพนธ์นี้ ในบทที่ 6 เป็นบททดสอบ จะแสดงที่มาของค่าวิทยาการศึกษาสำนึกซึ่งใช้ในงานวิจัยนี้ โดยการพิจารณาเลือกค่าวิทยาการศึกษาสำนึกนั้น จะเลือกค่าซึ่งทำให้แบบจำลองรวมที่ได้มีลักษณะตรงตามหลักการออกแบบแบบจำลองเชิงวัตถุที่ดี รวมทั้งจะทำการเปรียบเทียบกับแบบจำลองข้อมูลรวมที่จะได้จากงานวิจัย [7] นอกจากนี้จะนำแบบจำลองข้อมูลรวมไปสร้างเป็นตึกมารวม เพื่อทดสอบการเข้าถึงข้อมูลว่าผลที่ได้จะครบถ้วนตามที่ควรจะได้จากการสอบถามข้อมูลไปยังตึกมาดั้งเดิม 2 ชุดหรือไม่ ในบทที่ 7 จะเป็นบทสรุปผลการวิจัย รวมทั้งข้อเสนอนะซึ่งควรนำไปปรับปรุงให้ดียิ่งขึ้น ในส่วนของภาคผนวกนั้น ภาคผนวก ก จะแสดงตัวอย่างแบบจำลองข้อมูล 10 คู่ที่ใช้สำหรับการทดสอบในบทที่ 6 ภาคผนวก ข จะแสดงแบบจำลองข้อมูลรวมที่ได้จากการรวมตัวอย่างแบบจำลองข้อมูล 10 คู่ในภาคผนวก ก โดยผู้เชี่ยวชาญในเชิงวัตถุ ภาคผนวก ค เป็นรายละเอียดของแบบจำลองข้อมูลรวมที่ได้จากการทดสอบโดยใช้ค่าวิทยาการศึกษาสำนึกที่ได้นำเสนอในงานวิจัยนี้ ส่วนภาคผนวก ง จะแสดงตัวอย่างการนำแบบจำลองข้อมูลรวมมาสร้างบนฐานข้อมูลจริง และส่วนสุดท้ายคือภาคผนวก จ จะเป็นบทความจากงานวิจัยนี้ที่ได้รับการตีพิมพ์แล้ว