หอสมุดกลาง ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

แบบจำลองข้อมูล (Data Model) [1] เป็นแนวคิดที่นำมาใช้อธิบายโครงสร้างของฐานข้อมูล ได้แก่ ชนิด (Type) ความสัมพันธ์ (Relationship) และเงื่อนไขบังคับ (Constraint) ของข้อมูลในการจัดเก็บข้อมูล โดยอาศัยแนวคิดเชิงตรรกะ (Logical Concept) เช่น อ็อบเจกต์ (Object) พรอพเพอร์ตี (Property) และความสัมพันธ์ระหว่างกัน (Interrelationship) ตัวอย่างเช่น แบบจำลองข้อมูลเชิงวัตถุ (Object-Oriented Data Model) ซึ่งแสดงแทนแนวคิดของฐานข้อมูลในรูปของอ็อบเจกต์ ประกอบด้วย คลาส (Class) แอททริบิวต์ (Attribute) เมท็อด (Method) และความสัมพันธ์ระหว่างคลาส (Class Relationship) เป็นต้น

การรวมแบบจำลองข้อมูล (Data Model Integration) มีจุดประสงค์เพื่อการรวมข้อมูล (Data Integration) [2] ที่จัดเก็บไว้ในแหล่งข้อมูลต่างกันสองแหล่ง และจัดเตรียมให้ผู้ใช้สามารถเข้าถึงข้อมูลเหล่านั้นเสมือนว่าข้อมูลทั้งหมดถูกเก็บไว้ในแหล่งเดียวกัน โดยทั่วไปแล้วการรวมแบบจำลองข้อมูลจะต้องประกอบด้วยสามองค์ประกอบสำคัญคือแบบจำลองข้อมูลที่ต้องการรวมอย่างน้อยสองชุด เรียกว่าแบบจำลองข้อมูลโลคัล (Local Data Model) แบบจำลองข้อมูลรวม (Integrated Data Model) เรียกว่าแบบจำลองข้อมูลโกลบอล (Global Data Model) และแมปปิง (Mapping) ระหว่างแบบจำลองข้อมูลที่ต้องการรวมแต่ละชุดกับแบบจำลองข้อมูลรวม ดังนั้นผู้ทำการรวมแบบจำลองข้อมูลจะต้องระบุได้ว่าชื่อข้อมูลฟิลด์ (Field) ต่างๆ ในแต่ละแบบจำลองข้อมูลโลคัลที่ต้องการรวมมีความสัมพันธ์กันอย่างไร (เช่น ฟิลด์ Student ในแบบจำลองข้อมูลแรกเท่ากับฟิลด์ GradStudent ในแบบจำลองข้อมูลที่สอง) และจะแมป (Map) เป็นชื่อฟิลด์ใดในแบบจำลองข้อมูลรวม (เช่น แมปเป็นฟิลด์ Student) ซึ่งสามารถพิจารณาได้จากวากยสัมพันธ์ (Syntax) และความหมาย (Semantic) ของข้อมูลในแบบจำลองข้อมูลทั้งสองชุด

ทั้งนี้ แบบจำลองข้อมูลที่ดีจะต้องแสดงวากยสัมพันธ์และความหมายของข้อมูลที่จัดเก็บให้เพียงพอต่อการจำแนกความขัดแย้ง (Conflict) ระหว่างแบบจำลองข้อมูลแต่ละชุดและการสร้างแมปปิงที่มีความเหมาะสมที่สุด โดยอาจแบ่งประเภทของความขัดแย้งได้เป็น 2 ระดับ [3] คือ

1. ความขัดแย้งระดับข้อมูล (Data Level Conflict) ตัวอย่างเช่น
   - ความขัดแย้งของค่าข้อมูล (Data Value Conflict) เช่น ค่าข้อมูลของฟิลด์ "Faculty" ในแบบจำลองข้อมูลแรกหมายถึง "Science" และ "Engineer" แต่ในแบบจำลองข้อมูลที่สองหมายถึง "Science", "Account" และ "Engineer"

- ความขัดแย้งของหน่วยข้อมูล (Data Unit Conflict) เช่น หน่วยข้อมูลของฟิลด์ "Height" ในแบบจำลองข้อมูลแรกจัดเก็บค่าในหน่วยเซนติเมตร แต่ในแบบจำลองข้อมูลที่สองจัดเก็บค่าในหน่วยเมตร

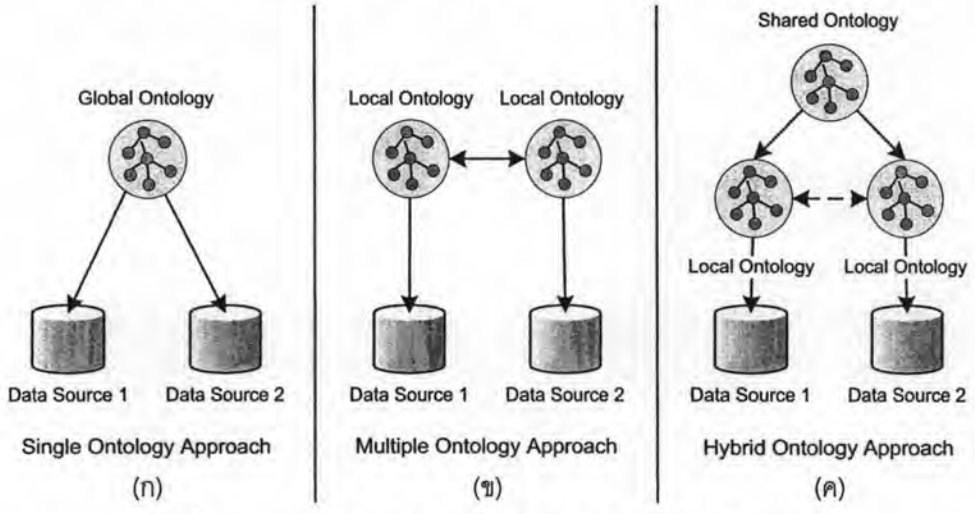
2. ความขัดแย้งระดับสกีมา (Schema Level Conflict) ตัวอย่างเช่น
   - ความขัดแย้งของชื่อ (Naming Conflict) เกิดได้สองลักษณะคือ ชื่อฟิลด์ข้อมูลต่างกันแต่หมายถึงข้อมูลเดียวกัน (Synonyms) เช่น "Student" กับ "GradStudent" หรือชื่อฟิลด์ข้อมูลเดียวกันแต่หมายถึงข้อมูลต่างกัน (Homonyms) เช่น "Address" ในแบบจำลองข้อมูลแรกหมายถึงที่ตั้งของมหาวิทยาลัย แต่ในแบบจำลองข้อมูลที่สองหมายถึงบ้านเลขที่
   - ความขัดแย้งของโครงสร้าง (Structure Conflict) เกิดขึ้นเมื่อข้อมูลเดียวกันถูกเก็บโดยใช้ชนิดของข้อมูลต่างกัน เช่น "Salary" ในแบบจำลองข้อมูลแรกเก็บเป็นจำนวนเต็ม (Integer) แต่ในแบบจำลองข้อมูลที่สองเก็บเป็นจำนวนจริง (Real)
   - ความขัดแย้งของการแอกกรีเกชัน (Aggregation Conflict) เกิดขึ้นเมื่อมีการใช้ความสัมพันธ์แบบแอกกรีเกชันในแบบจำลองข้อมูลหนึ่งแทนกลุ่มของแอททริบิวต์ในอีกแบบจำลองข้อมูล

สำหรับแบบจำลองข้อมูลเชิงวัตถุนั้น แม้ว่าตัวแบบจำลองข้อมูลจะแสดงวากยสัมพันธ์ของข้อมูลอย่างชัดเจน ดังเช่นที่ปรากฏในยูเอ็มแอลคลาสไดอะแกรม (UML Class Diagram) แต่มิได้แสดงถึงความหมายใดๆ ของข้อมูลไว้ภายในเลย ส่งผลให้ไม่สามารถจำแนกความขัดแย้งระดับข้อมูลได้ ดังนั้นการจะรวมแบบจำลองข้อมูลเชิงวัตถุสองชุดเข้าด้วยกันจำเป็นต้องอาศัยวิธีการเพิ่มเติมเพื่อสอดแทรกหรือกำกับความหมายลงบนแบบจำลองข้อมูล

วิธีการหนึ่งที่สามารถนำมาใช้อธิบายความหมายของข้อมูลในแบบจำลองข้อมูลเชิงวัตถุทั้งสองชุดที่ต้องการรวมกันทำได้โดยอาศัยแนวคิดของออนโทโลจี (Ontology) [4] ในการกำหนดคำศัพท์ (Vocabulary) สำหรับอธิบายความหมายของคลาส แอททริบิวต์ เมท็อด และความสัมพันธ์ที่ปรากฏในทั้งสองแบบจำลองข้อมูลร่วมกัน รวมไปถึงความหมายอื่นที่ผู้ทำการรวมแบบจำลองข้อมูลอาจต้องการเพิ่มเติมระหว่างการรวมเพื่อช่วยให้แบบจำลองข้อมูลรวมที่ได้มีความถูกต้องมากยิ่งขึ้น

แนวคิดเรื่องออนโทโลจีถูกนำมาใช้งานอย่างแพร่หลายในงานวิจัยด้านการรวมข้อมูลโดยอาศัยความหมาย (Semantic Data Integration) ซึ่งการใช้ออนโทโลจีจะให้ประโยชน์ทั้งในด้านการรวม (Integrate) และการสอบถาม (Query) ระหว่างแหล่งข้อมูลทั้งสอง ด้วยการทำหน้าที่เป็นตัวกลางติดต่อระหว่างผู้ใช้และแหล่งข้อมูล เพื่อจัดการเรื่องความต่างแบบกัน (Heterogeneity) และการกระจายตัว (Distribution) ของข้อมูลแทนผู้ใช้ ในปัจจุบันงานวิจัยด้านการนำออนโทโลจีมาใช้ในการรวมข้อมูลอาจแบ่งออกได้เป็น 3 แนวคิด [5] ตามวิธีการนำออนโทโลจีมาใช้อธิบายแบบจำลองข้อมูลดังรูปที่ 1.1 ได้แก่

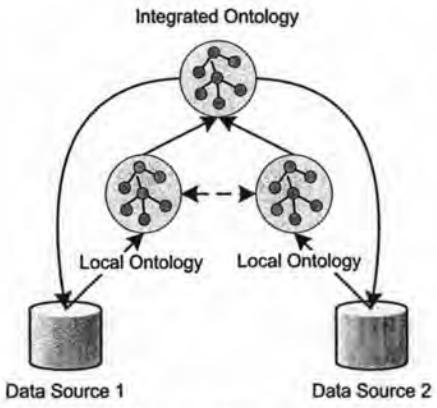
1. แบบออนโทโลจีเดียว (Single Ontology) เป็นแนวคิดที่เสนอให้ใช้ออนโทโลจีหนึ่งชุดกำหนดคำศัพท์เพื่ออธิบายแหล่งข้อมูลทั้งสองร่วมกันเรียกว่า โกลบอลออนโทโลจี (Global Ontology) ตามรูปที่ 1.1(ก) ซึ่งข้อมูลทั้งหมดจากทั้งสองแหล่งข้อมูลจะถูกแมปกับโกลบอลออนโทโลจี จากแนวคิดนี้ ผู้ทำการรวมแบบจำลองข้อมูลอาจสร้างออนโทโลจีชื่อ "Student" ขึ้นมาก่อน แล้วแมปจากฟิลด์ "Student" ในแบบจำลองข้อมูลที่ 1 และฟิลด์ "GradStudent" ในแบบจำลองข้อมูลที่ 2 หากทั้งสองฟิลด์หมายถึงสิ่งเดียวกันไปหาออนโทโลจี "Student" ตัวอย่างเช่น งานวิจัย SIMS [6] และ "Semantic Data Integration in Hierarchical Domains [7]"
2. แบบหลายออนโทโลจี (Multiple Ontologies) เป็นแนวคิดที่เสนอให้อธิบายแบบจำลองข้อมูลแต่ละชุดด้วยออนโทโลจีของตนเองแยกกันเรียกว่า โลคัลออนโทโลจี (Local Ontology) ดังรูปที่ 1.1(ข) แล้วเปรียบเทียบออนโทโลจีที่ได้เพื่อแมประหว่างสองออนโทโลจี จากแนวคิดนี้ผู้ทำการรวมแบบจำลองข้อมูลอาจสร้างออนโทโลจีชื่อ "Student" มาอธิบายฟิลด์ "Student" ในแบบจำลองข้อมูลที่ 1 และสร้างออนโทโลจีชื่อ "GradStudent" มาอธิบายฟิลด์ "GradStudent" ในแบบจำลองข้อมูลที่ 2 แล้วกำหนดว่าออนโทโลจีทั้งสองมีความสัมพันธ์แบบเท่ากัน (Equivalence) ตัวอย่างเช่น งานวิจัย OBSERVER [8]
3. แบบผสม (Hybrid) เป็นแนวคิดที่ผสมผสานระหว่างแนวคิดที่ 1 และ 2 โดยแต่ละแบบจำลองข้อมูลจะถูกอธิบายด้วยโลคัลออนโทโลจีของตนเองเช่นเดียวกับแนวคิดที่ 2 ในขณะเดียวกันจะมีการสร้างแชร์ดออนโทโลจี (Shared Ontology) ขึ้นมาอีกชั้นหนึ่ง เพื่อรองรับการสอบถามจากผู้ใช้และเก็บคำศัพท์ที่ โลคัลออนโทโลจีทั้งสองใช้ร่วมกันสำหรับเปรียบเทียบระหว่างโลคัลออนโทโลจีทำนองเดียวกับโกลบอลออนโทโลจีในแนวคิดที่ 1 แสดงได้ตามรูปที่ 1.1(ค) ตัวอย่างเช่น งานวิจัย COIN [9] และ MECOTA [10]



รูปที่ 1.1 แนวคิดในการนำออนโทโลจีมาใช้ในการรวมข้อมูล [5]

อย่างไรก็ตาม แต่ละแนวคิดต่างมีข้อดีข้อเสียแตกต่างกันออกไป เช่น แนวคิดแบบออนโทโลจีเดียวช่วยให้สอบถามและใช้งานข้อมูลได้สะดวก แต่ขาดความยืดหยุ่นและผู้ออกแบบโกลบอลออนโทโลจีจำเป็นต้องเข้าใจแบบจำลองข้อมูลทั้งสองเป็นอย่างดี ในขณะที่แนวคิดแบบหลายออนโทโลจีมีความยืดหยุ่นมากกว่า แต่มีความยุ่งยากในส่วนของการเปรียบเทียบโลคัลออนโทโลจีและผู้ใช้ไม่สามารถทราบถึงภาพรวมของข้อมูลทั้งหมด ส่วนแนวคิดแบบผสมแม้ว่าจะช่วยลดความยุ่งยากในส่วนการออกแบบออนโทโลจี แต่ในการใช้งานจริงอาจมีปัญหาด้านประสิทธิภาพการทำงาน เนื่องจากทุกครั้งที่มีการเรียกใช้ข้อมูลจะต้องสอบถามข้อมูลสองขั้น คือจากแชร์ดออนโทโลจีไปยังโลคัลออนโทโลจีทุกชุด และจากแต่ละชุดโลคัลออนโทโลจีไปยังข้อมูลที่โลคัลออนโทโลจีนั้นติดต่อ

จากจุดอ่อนที่เกิดขึ้นในการใช้งานออนโทโลจีเพื่อการรวมข้อมูลทั้งสามแนวทางข้างต้น ผู้วิจัยจึงมีแนวคิดที่จะนำข้อดีของแต่ละแนวทางมาประยุกต์ใช้ร่วมกัน เพื่อให้เป็นอีกวิธีการหนึ่งที่สามารถนำไปใช้ในการรวมแบบจำลองข้อมูลโดยอาศัยออนโทโลจีในการอธิบายความหมายของแบบจำลองข้อมูล โดยตั้งชื่อแนวคิดใหม่นี้ว่า "แนวคิดแบบรวมหลายออนโทโลจี (Integrated Multiple Ontologies Approach)" แสดงไว้ตามรูปที่ 1.2 แนวคิดนี้นำข้อดีจากการใช้โลคัลออนโทโลจีในการอธิบายแบบจำลองข้อมูลแต่ละชุด ซึ่งจะช่วยรักษาความหมายที่ปรากฏในแบบจำลองข้อมูลไว้ แล้วนำวิธีการรวมออนโทโลจี (Ontology Integration) [11] มาใช้เปรียบเทียบและรวมโลคัลออนโทโลจีคู่ที่ได้เข้าด้วยกันเรียกว่าออนโทโลจีรวม (Integrated Ontology) ต่อจากนั้นจะนำออนโทโลจีรวมที่ได้ไปแมปกับแบบจำลองข้อมูลทั้งสองชุดในลักษณะเดียวกับการใช้โกลบอลออนโทโลจีในการรวมแบบออนโทโลจีเดียว ด้วยวิธีการนี้จะช่วยให้เจ้าของแหล่งข้อมูลได้รับความสะดวกในการกำหนดโลคัลออนโทโลจีของตนเอง รวมทั้งสะดวกต่อผู้ใช้ในแง่ของการสอบถามข้อมูล เพราะสามารถดำเนินการได้โดยอาศัยการติดต่อกับออนโทโลจีรวมเพียงอย่างเดียว อีกทั้งผู้ใช้ยังเห็นภาพรวมทั้งหมดของข้อมูลผ่านออนโทโลจีรวมด้วย



Integrated Multiple Ontology Approach

รูปที่ 1.2 แนวคิดในการนำออนโทโลจีมาใช้ในการรวมข้อมูลแบบรวมหลายออนโทโลจี

สำหรับขั้นตอนวิธีในการรวมออนโทโลจีนั้น ปัจจุบันมีทั้งงานวิจัยและงานวิจัยเชิงสำรวจ [12-15] ที่เกี่ยวข้องอยู่มากมาย เช่น งานวิจัย OBSERVER [9], PROMPT [16], Chimaera [17] และ FCA-

หอสมุดกลาง ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

MERGE [18] อย่างไรก็ตาม งานวิจัยที่เกี่ยวข้องกับขั้นตอนวิธีการรวมออนโลจีเหล่านี้มุ่งเน้นไปที่การรวมออนโทโลจีขนาดใหญ่เพื่ออธิบายข้อมูลที่หลากหลาย เช่น ข้อมูลบนเว็บไซต์ จึงส่งผลให้ความหมายของข้อมูลเหล่านี้ซึ่งอธิบายด้วยออนโทโลจีมีรูปแบบความสัมพันธ์ที่หลากหลายและซับซ้อนตามไปด้วย และทำให้ขั้นตอนวิธีในการรวมออนโทโลจีจำต้องมีความซับซ้อนมาก เช่น มีการนำการประมวลผลภาษาธรรมชาติ (Natural Language Processing) มาใช้ เป็นต้น

ผู้วิจัยเล็งเห็นว่า ในการรวมออนโทโลจีที่อธิบายแบบจำลองข้อมูลนั้น แม้ว่าฐานข้อมูลที่ต้องการรวมจะมีความซับซ้อน แต่โครงสร้างและความสัมพันธ์ที่ปรากฏมีรูปแบบที่ค่อนข้างชัดเจนตายตัว รูปแบบของความหมายมีจำกัด เนื่องจากถูกตีกรอบด้วยตัวแบบจำลองข้อมูลเองจึงน่าจะลดความซับซ้อนของขั้นตอนวิธีลงได้ [19] ในวิทยานิพนธ์นี้ ผู้วิจัยจึงได้นำวิธีการรวมแบบจำลองข้อมูลเชิงวัตถุโดยใช้วิทยาการศึกษาสำนึก (Heuristics) และการวิเคราะห์ความสัมพันธ์ระหว่างคลาสที่ใช้ในงานวิจัย [20] และแนวทางการแปลงโมเดล (Model) ระหว่างออนโทโลจีและแบบจำลองข้อมูลเชิงวัตถุ [21] มาประยุกต์ใช้ร่วมกัน เพื่อเสนอเป็นวิธีการรวมออนโทโลจีแบบง่ายโดยอาศัยการวิเคราะห์ความสัมพันธ์ระหว่างคลาส เนื่องจากจุดมุ่งหมายของการรวมออนโทโลจีในวิทยานิพนธ์นี้ก็คือการนำมาใช้กับการรวมแบบจำลองข้อมูลเชิงวัตถุนั่นเอง

ดังนั้น หากสามารถรวมโลคัลออนโทโลจีแต่ละชุดเข้าด้วยกันเป็นออนโทโลจีรวมแล้วนำไปใช้กำกับลงบนการรวมแบบจำลองข้อมูลเชิงวัตถุที่แบบจำลองข้อมูลแต่ละชุด จะได้เป็นแนวคิดใหม่ในการนำออนโทโลจีมาใช้ในการรวมข้อมูล ซึ่งจะช่วยลดทั้งความยุ่งยากในการออกแบบออนโทโลจีรวมและความซับซ้อนในการติดต่อระหว่างออนโทโลจีขณะสอบถามข้อมูล โดยอาศัยสมมติฐานว่า หากนำแบบจำลองข้อมูลที่ได้รับการอธิบายและแสดงแทนความหมายในรูปแบบที่เหมาะสมและเพียงพอ มารวมเข้าด้วยกันโดยใช้ขั้นตอนวิธีที่เหมาะสม จะทำให้การรวมแบบจำลองข้อมูลเป็นไปอย่างถูกต้องและครบถ้วน ทั้งนี้เมื่อเปรียบเทียบกับองค์ประกอบทั้งสามในการรวมแบบจำลองข้อมูลแล้ว โลคัลออนโทโลจีและออนโทโลจีรวมจะทำหน้าที่แสดงแทนแบบจำลองข้อมูลโลคัลและแบบจำลองข้อมูลโกลบอล ส่วนการกำกับออนโทโลจีรวมลงบนแบบจำลองข้อมูลเชิงวัตถุก็คือส่วนการสร้างแมปปิงระหว่างแบบจำลองข้อมูลโลคัลและแบบจำลองข้อมูลโกลบอล ตามลำดับ

ในส่วนของวิทยานิพนธ์นี้ ได้นำเสนอถึงการนำแนวคิดนี้ไปประยุกต์ใช้งาน พร้อมทั้งเปรียบเทียบข้อแตกต่างระหว่างแนวคิดนี้กับทั้งสามแนวคิดที่มีอยู่เดิม ภายในวิทยานิพนธ์นี้ผู้วิจัยจะนำเสนอวิธีการรวมแบบจำลองข้อมูลเชิงวัตถุ โดยเลือกใช้ออนโทโลจีในการอธิบายแบบจำลองข้อมูลแต่ละชุด เพื่อแสดงแทนความหมายที่ปรากฏในแบบจำลองข้อมูลและความหมายอื่นที่ผู้ทำการรวมแบบจำลองข้อมูลอาจต้องการเพิ่มระหว่างการรวม ซึ่งขั้นตอนนี้จะเปิดให้เจ้าของฐานข้อมูลเป็นผู้สร้างออนโทโลจีขึ้นมาอธิบายแบบจำลองข้อมูลของตน แล้วนำออนโทโลจีทั้งคู่ที่ได้มารวมเข้าด้วยกันโดยอัตโนมัติ ด้วยวิธีการเปรียบเทียบความแตกต่างเพื่อกำหนดค่าความสัมพันธ์ตามวิธีการรวมออนโทโลจีที่ประยุกต์ขึ้น หลังจาก

นั้นจะนำออนโทโลจีรวมที่ได้มาใช้กำกับลงบนแบบจำลองข้อมูลแต่ละชุด และในอนาคตจะสามารถพัฒนาต่อเป็นเฟรมเวิร์คที่สมบูรณ์เพื่อรองรับการสอบถามข้อมูลได้

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อนำเสนอแนวคิดและต้นแบบในการนำออนโทโลจีมาใช้ในการรวมข้อมูลแบบรวมหลายออนโทโลจี
2. เพื่อประยุกต์ใช้แนวคิดแบบรวมหลายออนโทโลจีกับการรวมแบบจำลองข้อมูลเชิงวัตถุ
3. เพื่อกำหนดแนวทางในการอธิบายแบบจำลองข้อมูลเชิงวัตถุด้วยออนโทโลจี และการกำกับออนโทโลจีลงบนแบบจำลองข้อมูลเชิงวัตถุ
4. เพื่อออกแบบวิธีการรวมออนโทโลจีโดยการสร้างออนโทโลจีรวมจากการวิเคราะห์ความสัมพันธ์ระหว่างคลาส
5. เพื่อแสดงวิธีการนำออนโทโลจีรวมที่ได้ไปใช้ในการรวมและการสอบถามข้อมูลจากฐานข้อมูลเชิงวัตถุที่นำมารวม

1.3 ขอบเขตของการวิจัย

1. เสนอแนวทางในการสร้างออนโทโลจีรวมโดยการวิเคราะห์ความสัมพันธ์ระหว่างคลาสของแบบจำลองข้อมูลเชิงวัตถุ โดยประยุกต์จากงานวิจัย [20] ร่วมกับแนวทางการแปลงโมเดลระหว่างออนโทโลจีและแบบจำลองข้อมูลเชิงวัตถุ [21]
2. เสนอแนวทางในการกำกับออนโทโลจีรวมลงบนแบบจำลองข้อมูลเชิงวัตถุ
3. พัฒนาต้นแบบระบบการรวมแบบจำลองข้อมูลซึ่งสนับสนุนการสร้างออนโทโลจีรวม การกำกับด้วยออนโทโลจีรวม และการสอบถามข้อมูลผ่านออนโทโลจีรวม
4. การเปรียบเทียบแบบจำลองข้อมูลเชิงวัตถุจะพิจารณาเฉพาะส่วนข้อมูล ได้แก่ แอททริบิวต์และคุณลักษณะที่เกี่ยวข้องกับแอททริบิวต์ในคลาสเท่านั้น โดยไม่คำนึงถึงส่วนการดำเนินการ (Opetation) ได้แก่ เมท็อด และคุณลักษณะที่เกี่ยวข้องกับเมท็อดในคลาส
5. ข้อมูลความสัมพันธ์ระหว่างคลาสที่จะใช้พิจารณาในการรวมออนโทโลจีเป็นไปตามที่กำหนดในงานวิจัย [20] ได้แก่ข้อมูลเกี่ยวกับ
   - คลาส
   - แอททริบิวต์
   - ชนิดของแอททริบิวต์
   - ความสัมพันธ์แบบสืบทอดคุณลักษณะ
   - ความสัมพันธ์แบบพี่น้อง
   - ความสัมพันธ์แบบแอกกรีเกชัน
   - ความหมายของคลาส
   - ความหมายของแอททริบิวต์
   - ข้อกำหนดบูรณภาพของแอททริบิวต์
   - คำที่มีความหมายเหมือนกัน
   - คำที่เป็นส่วนประกอบกัน

6. ต้นแบบที่พัฒนาสามารถใช้งานในสภาพแวดล้อมของระบบปฏิบัติการไมโครซอฟท์วินโดวส์ เอ็กซ์พี เป็นอย่างน้อย
7. ฐานข้อมูลที่จะนำมารวมจะต้องอยู่ในรูปของแบบจำลองข้อมูลเชิงวัตถุ หรือได้รับการแปลงให้อยู่ในรูปของแบบจำลองข้อมูลเชิงวัตถุแล้วเท่านั้น
8. ฐานข้อมูลที่จะนำมารวมจะต้องเป็นฐานข้อมูลที่เก็บข้อมูลเรื่องเดียวกันหรือมีความเกี่ยวข้องกัน
9. ออนโทโลจีจะได้รับการอธิบายในรูปแบบอาวล์ดีแอล (OWL DL: Description Logic) เป็นอย่างน้อย
10. ในการอนุมานออนโทโลจี ต้นแบบจะพัฒนาโดยใช้จีนาเอพีไอ รุ่นที่ 2.3 เป็นอย่างน้อย
11. การสอบถามข้อมูลจากระบบการรวมแบบจำลองข้อมูลเชิงวัตถุกระทำผ่านแผ่นแบบ (Template) โดยอ้างอิงจากออนโทโลจีที่กำหนดไว้
12. การทดสอบต้นแบบการรวมและการสอบถามข้อมูลที่พัฒนาขึ้นในขั้นแรกจะทดสอบด้วยแบบจำลองข้อมูลที่ใช้ในงานวิจัย [20] เพื่อตรวจสอบความถูกต้องในการรวม เมื่อได้ผลตรงตามต้องการแล้วจะนำไปทดสอบด้วยแบบจำลองข้อมูลที่ถูกใช้งานจริงซึ่งมีความซับซ้อนขึ้น เพื่อหาข้อจำกัดของระบบ
13. วิเคราะห์เปรียบเทียบคุณสมบัติของระบบการใช้ออนโทโลจีในการรวมข้อมูลที่ได้พัฒนาขึ้นตามแนวคิดนี้กับทั้งสามแนวคิดที่มีอยู่เดิม

1.4 วิธีดำเนินการวิจัย

1. ศึกษาทฤษฎีของการรวมสารสนเทศ
2. ศึกษาทฤษฎีของออนโทโลจี การรวมออนโทโลจี อาร์ดีเอฟ (RDF) และการอนุมาน
3. ศึกษาทฤษฎีการรวมแบบจำลองข้อมูลเชิงวัตถุ
4. ศึกษาทฤษฎีการแปลงระหว่างแบบจำลองข้อมูลเชิงวัตถุและออนโทโลจี
5. ออกแบบระบบการรวมออนโทโลจีและการกำกับออนโทโลจีรวมที่สามารถนำไปใช้ในการรวมแบบจำลองข้อมูลเชิงวัตถุ
6. พัฒนาต้นแบบระบบการรวมออนโทโลจีและการกำกับออนโทโลจีรวมเพื่อนำไปใช้ในการรวมแบบจำลองข้อมูลเชิงวัตถุ
7. ทดสอบและปรับปรุงแก้ไขต้นแบบ
8. วิเคราะห์และสรุปผล พร้อมข้อเสนอแนะ
9. จัดทำรายงานวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แนวทางใหม่ในการนำออนโทโลจีมาใช้ในการรวมข้อมูล โดยใช้การสร้างออนโทโลจีรวมที่ได้จากการรวมโลคัลออนโทโลจีมาใช้เป็นตัวกลางในการรวมและการสอบถามข้อมูลระหว่างแต่ละฐานข้อมูล เรียกว่าแนวคิดแบบรวมหลายออนโทโลจี
2. ได้ขั้นตอนวิธีการรวมออนโทโลจีโดยการสร้างออนโทโลจีรวมจากการวิเคราะห์ความสัมพันธ์ระหว่างคลาส
3. สามารถอธิบายแบบจำลองข้อมูลเชิงวัตถุด้วยออนโทโลจีที่เหมาะสม
4. สามารถนำแนวคิดและขั้นตอนวิธีที่ได้มาพัฒนาเป็นต้นแบบการใช้ออนโทโลจีเพื่อรวมแบบจำลองข้อมูลเชิงวัตถุ โดยใช้การสร้างออนโทโลจีรวมจากการวิเคราะห์ความสัมพันธ์ระหว่างคลาส แล้วนำไปกำกับลงบนแบบจำลองข้อมูลเชิงวัตถุเพื่อใช้ในการสอบถามข้อมูล
5. สามารถเปรียบเทียบข้อดีข้อเสียของแนวทางนี้กับแนวทางที่มีอยู่เดิมได้ ซึ่งแนวทางใหม่นี้สามารถนำไปประยุกต์ใช้กับฐานข้อมูลในองค์กรต่อไปได้

1.6 โครงสร้างวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้จัดแบ่งเนื้อหาออกเป็น 6 ส่วนได้แก่บทที่ 2 ถึงบทที่ 7 ตามลำดับ เริ่มจากในบทที่ 2 เป็นการสรุปถึงทฤษฎีและงานวิจัยต่างๆ ที่เกี่ยวข้องและนำมาประยุกต์ใช้กับส่วนต่างๆ ของวิทยานิพนธ์นี้ บทที่ 3 จะกล่าวถึงแนวคิดและวิธีการสร้างออนโทโลจีเพื่ออธิบายและเพิ่มความหมายให้กับแบบจำลองข้อมูลเชิงวัตถุ ซึ่งแบ่งเป็นสองส่วนคือส่วนการสร้างออนโทโลจีเพื่ออธิบายความหมายที่ปรากฏในแบบจำลองข้อมูลเชิงวัตถุ และส่วนการเพิ่มเติมความหมายของแบบจำลองข้อมูลเชิงวัตถุลงในออนโทโลจี เพื่อให้ได้โลคัลออนโทโลจีที่สามารถอธิบายความหมายของแบบจำลองข้อมูลเชิงวัตถุได้อย่างสมบูรณ์ ส่วนบทที่ 4 จะอธิบายถึงขั้นตอนวิธีในการเปรียบเทียบและการรวมออนโทโลจีอย่างละเอียด ซึ่งเป็นการนำโลคัลออนโทโลจีสองชุดที่ได้จากบทที่ 3 มาเปรียบเทียบหาความสัมพันธ์ระหว่างกัน และนำค่าความสัมพันธ์ที่ได้มาวิเคราะห์เพื่อหารูปแบบออนโทโลจีรวมที่เหมาะสม จากนั้นในบทที่ 5 จะนำเสนอเฟรมเวิร์คสำหรับการรวมแบบจำลองข้อมูลเชิงวัตถุด้วยออนโทโลจีที่ได้พัฒนาขึ้นตามแนวคิดของวิทยานิพนธ์นี้ประกอบด้วย แนวคิดในการออกแบบเฟรมเวิร์ค องค์ประกอบของเฟรมเวิร์ค ได้แก่ ส่วนการสร้างโลคัลออนโทโลจี ส่วนการสร้างออนโทโลจีรวม และส่วนการสร้างแมปปิงระหว่างออนโทโลจีรวมมายังฐานข้อมูลแต่ละชุด รวมถึงแผ่นแบบการสอบถามข้อมูลผ่านการทำงานของเฟรมเวิร์คเพื่อแสดงการนำแนวคิดไปใช้งานจริง ส่วนบทที่ 6 จะแสดงการทดสอบการทำงานเพื่อประเมินความสามารถของเฟรม-เวิร์ค ได้แก่ (1) การทดสอบผลการสร้างออนโทโลจีรวม โดยเลือกนำมาจากผลการทดสอบด้วยแบบจำลองข้อมูลตัวอย่าง 13 คู่ ได้แก่ แบบจำลองข้อมูลเชิงวัตถุที่ใช้ทดสอบในงานวิจัย [20] สิบคู่ และแบบจำลองข้อมูลที่ออกแบบสำหรับใช้งานในโครงการระบบสารสนเทศเพื่อการบริหารของจุฬาลงกรณ์มหาวิทยาลัย (MISCU) จากงานวิจัย [22-25] สามคู่ และ (2) การทดสอบความถูกต้องในการสอบถาม

ข้อมูลผ่านออนโทโลจีรวมที่ได้ไปยังฐานข้อมูลเชิงวัตถุที่นำมารวม ซึ่งผลลัพธ์ที่ได้จะต้องถูกต้องและครบถ้วนเทียบเท่ากับผลรวมจากการสอบถามโดยตรงไปยังฐานข้อมูลแต่ละชุด บทที่ 7 จะเป็นบทสรุปผลการวิจัย การวิเคราะห์เปรียบเทียบคุณสมบัติของแนวคิดนี้กับแนวคิดที่มีอยู่เดิม พร้อมทั้งปัญหาและข้อจำกัดที่พบจากการวิจัย รวมถึงข้อเสนอแนะเพื่อปรับปรุงแก้ไขหรือเป็นแนวทางการทำวิจัยในอนาคต

นอกจากนี้ ในตอนท้ายของวิทยานิพนธ์ยังได้แทรกภาคผนวกไว้ 3 บทเริ่มจากในภาคผนวก ก และ ข จะแสดงตัวอย่างการรวมแบบจำลองข้อมูลสองคู่ด้วยออนโทโลจีที่ได้พัฒนาขึ้นตามแนวคิดของวิทยานิพนธ์นี้ ซึ่งในแต่ละบทจะประกอบด้วยแบบจำลองข้อมูลเชิงวัตถุคู่ที่นำมารวม โลคัลออนโทโลจีสองชุดที่สร้างขึ้นจากแบบจำลองข้อมูลเชิงวัตถุคู่นั้น ออนโทโลจีรวมที่ได้จากการรวมโลคัลออนโทโลจี และตัวอย่างการนำออนโทโลจีรวมมาใช้งานโดยสร้างเป็นวิวกำกับลงบนฐานข้อมูลเชิงวัตถุเพื่อใช้ในการสอบถามข้อมูล ตามลำดับ ส่วนภาคผนวก ง เป็นการรวบรวมบทความที่ได้รับการตีพิมพ์จากวิทยานิพนธ์นี้

1.7 ผลงานที่ตีพิมพ์จากการวิจัย

ส่วนหนึ่งของผลงานจากการวิจัยนี้ได้รับการตีพิมพ์และนำเสนอเป็นบทความทางวิชาการในหัวข้อเรื่อง "An Integration of Data Sources with UML Class Models Based on Ontological Analysis" โดย Manachaya Jamadhvaja และ Twittie Senivongse ในงานประชุมวิชาการ "The ACM Workshop on Interoperability of Heterogeneous Information Systems (IHIS'05)" ซึ่งจัดขึ้นที่เมืองเบรเมน ประเทศสหพันธ์สาธารณรัฐเยอรมนี เมื่อวันที่ 4 พฤศจิกายน 2548

และบางส่วนของผลงานได้รับการตีพิมพ์และนำเสนอเป็นบทความทางวิชาการในหัวข้อเรื่อง "An Approach to Constructing an Integrated Ontology for Integrating Object-Oriented Data Models" โดย Manachaya Jamadhvaja และ Twittie Senivongse ในงานประชุมวิชาการ "The 1st National Conference on Computing and Information Technology (NCCIT 2005)" จัดโดยสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ ระหว่างวันที่ 24-25 พฤษภาคม 2548 ซึ่งบทความนี้ได้รับการตีพิมพ์ซ้ำในวารสารเทคโนโลยีสารสนเทศ ปีที่ 1 ฉบับที่ 2 กรกฎาคม – ธันวาคม 2548 โดยคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ