บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยต่างๆ ที่เกี่ยวข้อง เพื่อใช้ในการออกแบบและพัฒนาขั้นตอนวิธีการออกแบบในการรวมแบบจำลองข้อมูลเชิงวัตถุ โดยส่วนที่จะกล่าวถึงมีดังนี้

1) แนวคิดการรวมแบบจำลองข้อมูล [1]
2) แนวคิดเชิงวัตถุ [10]
3) งานวิจัย ."A Methodology for Integration of Heterogeneous Databases" [3]
4) งานวิจัย "Description Logics and Correspondence Refinement Process in Databases Interoperability's Service" [7]

2.1 แนวคิดการรวมแบบจำลองข้อมูล [1]

ขั้นตอนโดยทั่วไปของการรวมแบบจำลองข้อมูลมีดังนี้ คือ

1. ขั้นตอนการเปรียบเทียบ (Comparison Step) เป็นขั้นตอนของการวิเคราะห์ และเปรียบเทียบข้อมูลในแบบจำลองข้อมูล ซึ่งสามารถชี้ปัญหาที่เกิดจากความขัดแย้งของชื่อ ความขัดแย้งของโครงสร้าง เป็นต้น และสามารถบอกถึงความสัมพันธ์ของวัตถุในระหว่างแบบจำลอง
2. ขั้นตอนการการรวม (Merging Step) เป็นขั้นตอนที่ทำการรวมวัตถุต่างๆ ในระหว่างแบบจำลองข้อมูล โดยวัตถุที่เหมือนกัน อาจจะรวมเป็นวัตถุเดียวกันในแบบจำลองข้อมูลใหม่
3. ขั้นตอนการเชื่อมโยง (Mapping Step) จะทำการบอกว่าวัตถุในแบบจำลองข้อมูลใหม่ที่ได้จากการรวม จะเชื่อมไปที่วัตถุใดของแบบจำลองข้อมูลเดิม

2.2 แนวคิดเชิงวัตถุ [10]

แนวคิดเชิงวัตถุสำหรับแบบจำลองข้อมูลเชิงวัตถุทั่วไป มีรายละเอียดพอสังเขป ดังนี้

2.2.1 โครงสร้างพื้นฐานของวัตถุ

วัตถุ (Object) เป็นเอนทิทีที่มีทั้งแอตทริบิวท์ซึ่งอธิบายคุณลักษณะของวัตถุและเมธอดซึ่งเป็นการกระทำที่เกี่ยวกับวัตถุนั้น แต่ก็มีวัตถุหลายชนิดที่มีคุณสมบัติคล้ายกัน จึงเกิดแนวคิดของคลาสซึ่งเป็นต้นแบบของวัตถุหลายๆ ชนิดที่มีลักษณะเหมือนกันบางอย่างทำให้สามารถจัดให้

อยู่ในจำพวกเดียวกันได้ ยกตัวอย่างเช่น คลาสของ Teacher หมายถึงวัตถุใดๆที่มีชื่อ ตำแหน่ง เงินเดือน เป็นต้น

องค์ประกอบพื้นฐานที่สำคัญของคลาสนั้นแบ่งออกได้เป็น 2 ชนิด คือแอตทริบิวท์และเมธทอด

- แอตทริบิวท์ เป็นคุณลักษณะของคลาส ซึ่งเป็นสิ่งที่เป็นลักษณะเฉพาะตัวที่จะแสดงให้เห็นความแตกต่างกันของวัตถุแต่ละชนิด
- เมธทอด เป็นคุณสมบัติของคลาสและเป็นส่วนที่แสดงให้เห็นว่าคลาสนั้นๆ สามารถกระทำอะไรได้บ้าง ซึ่งมีคุณลักษณะคล้ายกับฟังก์ชัน (Function) หรือโพรซีเจอร์ (Procedure) ในโปรแกรมภาษาอื่นๆ

2.2.2 ความสัมพันธ์ระหว่างคลาส

เป็นโครงสร้างที่แสดงความสัมพันธ์ของคลาส ซึ่งจะพิจารณา 3 ประเภทด้วยกันคือ

- ความสัมพันธ์แบบสืบทอดคุณลักษณะหรืออิสอะ (ISA)

เป็นความสัมพันธ์ระหว่างซูเปอร์คลาส (Superclass) และสับคลาส (Subclass) โดยที่ สับคลาสจะสืบทอดทั้งแอตทริบิวท์และเมธทอดของซูเปอร์คลาส หากคลาส A มีชุดแอตทริบิวท์และเมธทอดเป็น {Ai} และคลาส B มีชุดแอตทริบิวท์และเมธทอดเป็น {Bi} แล้ว A อิสอะ B ก็ต่อเมื่อ {Bi} $\subseteq$ {Ai} [11] รวมทั้ง ทุกอินสแตนซ์ (Instance) ของสับคลาสจะเป็นอินสแตนซ์ของซูเปอร์คลาสด้วย

จากรูปที่ 2.1 Teacher อิสอะ Person หมายถึง

แอตทริบิวท์และเมธทอดของ Person $\subseteq$ แอตทริบิวท์และเมธทอดของ Teacher

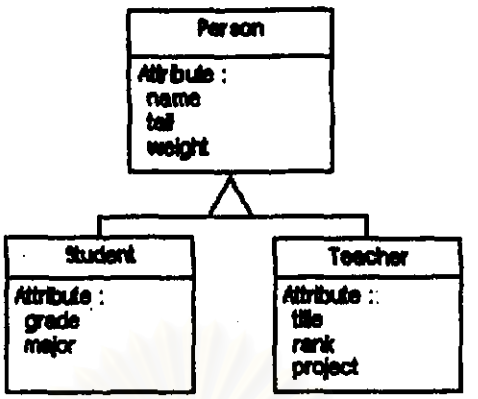
และ อินสแตนซ์ของ Teacher $\in$ อินสแตนซ์ของ Person

และ Student อิสอะ Person หมายถึง

แอตทริบิวท์และเมธทอดของ Person $\subseteq$ แอตทริบิวท์และเมธทอดของ Student

และ อินสแตนซ์ของ Student $\in$ อินสแตนซ์ของ Person

จากรูปที่ 2.1 คลาส Teacher และคลาส Student เป็นสับคลาสของคลาส Person นั่นหมายความว่า คลาส Teacher และคลาส Student สืบทอดคุณลักษณะจากคลาส Person คืออย่างน้อยจะมีแอตทริบิวท์และเมธทอดเช่นเดียวกับในคลาส Person และทุกอินสแตนซ์ของคลาส Teacher และคลาส Student จะเป็นอินสแตนซ์ของคลาส Person

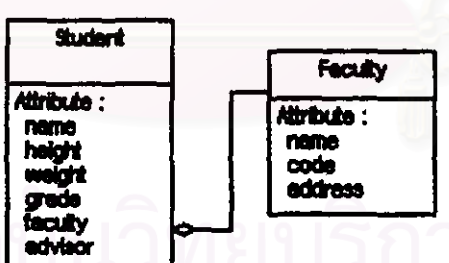


รูปที่ 2.1 ความสัมพันธ์แบบสืบทอดคุณลักษณะและแบบพี่น้อง

- ความสัมพันธ์แบบพี่น้อง

  เป็นความสัมพันธ์ระหว่างสับคลาสใดๆ ที่อยู่ในระดับ (level) เดียวกันและมีซูเปอร์คลาสเดียวกัน [12] จากรูปที่ 2.1 คลาส Teacher และคลาส Student มีความสัมพันธ์แบบพี่น้องกัน

- ความสัมพันธ์แบบแอกกรีเกชัน เป็นความสัมพันธ์ระหว่างคลาสที่มีคลาสอื่นเป็นส่วนประกอบ ดังรูปที่ 2.2 คลาส Student มีความสัมพันธ์แบบแอกกรีเกชันกับคลาส Faculty โดยผ่านทางแอตทริบิวท์ faculty



รูปที่ 2.2 ความสัมพันธ์แบบแอกกรีชัน

2.3 งานวิจัย "A Methodology for Integration of Heterogeneous Databases" [3]

เป็นงานวิจัยที่ทำการรวมสกีมาในเชิงวัตถุ โดยผู้ทำการระบุจะเป็นผู้ระบุว่าคลาสใดจาก 2 สกีมาที่มีความเหมือนกันหรือมีความสัมพันธ์แบบสืบทอดคุณลักษณะจากนั้นก็จะทำการรวมความสัมพันธ์ที่ผู้ทำการรวมได้ระบุนี้เป็นสกีมาใหม่ ซึ่งขั้นตอนการรวมจะเป็นไปอย่างอัตโนมัติ นอกจากนี้ในงานวิจัยนี้ยังได้กล่าวถึงการใส่ความหมายให้แก่แอตทริบิวท์ ซึ่งอธิบายว่ากรณีที่สองแอตทริบิวท์ใดที่ผู้ทำการรวมระบุว่าเหมือนกัน แต่มีค่าของความหมายต่างกัน จะต้องมีฟังก์ชันการแปลงค่าของความหมายที่ต่างกันให้เหมือนกัน แต่ไม่ได้มีการนำความหมายมาช่วยในขั้นตอนการ

เปรียบเทียบว่าแอตทริบิวท์ทั้งสองเหมือนกันหรือไม่ นอกจากนี้ ในงานวิจัยนี้ยังไม่ได้กล่าวถึงความสัมพันธ์แบบแอกกรีเกชันและลักษณะของแอตทริบิวท์ที่มีลักษณะเป็นส่วนประกอบกัน ดังนั้นในวิทยานิพนธ์นี้จึงจะแก้ปัญหาดังกล่าว รวมถึงการนำความหมายมาช่วยในขั้นตอนการเปรียบเทียบด้วย

2.4 งานวิจัย "Description Logics and Correspondence Refinement Process In Databases Interoperability's Service" [7]

วิธีการนี้จะทำการรวมสกีมาโดยเลือกใช้แบบจำลองข้อมูลที่มีรูปแบบเป็นตรรกศาสตร์เทอร์มินอลอจิคับ (Terminological Logics) โดยมีการอธิบายรายละเอียด (Description) และความหมายให้กับแต่ละเทอม (Term) เช่น เทอมของ Price ประกอบไปด้วยความหมาย เช่น Currency : "Rupee"และมีการใช้ค่าวิทยาการศึกษาสำนึกเพื่อช่วยในการหาความสัมพันธ์ระหว่างเทอม โดยพิจารณาจากความหมายต่างๆ ในแต่ละเทอม ซึ่งประกอบไปด้วยชื่อและค่าของมัน ดังนี้

ให้ T1, T2 เป็นเทอมที่จะพิจารณาความสัมพันธ์กัน

α เป็นจำนวนความหมายของ T1

β เป็นจำนวนความหมายของ T2 ซึ่ง $\alpha \leq \beta$

จากนั้น ทำการพิจารณาแต่ละความหมายซึ่งผลที่ได้จะแทนด้วยค่า r_i โดยที่

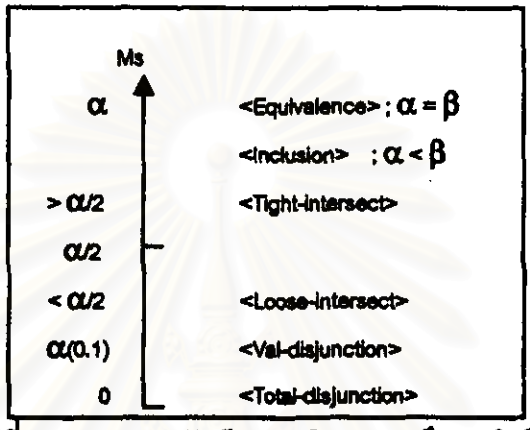
- $r_i = 1$ ถ้าชื่อและค่าของความหมายเหมือนกัน
- $r_i = 0.1$ ถ้าชื่อเหมือนกันแต่ค่าของความหมายต่างกัน
- $r_i = 0$ ถ้าชื่อต่างกันและค่าของความหมายต่างกัน

ผลที่ได้จากการเปรียบเทียบแต่ละความหมายจะเก็บในเวกเตอร์ (Vector) $R = [r_1, r_2, ..., r_\alpha]$ โดยที่ $Ms = \sum r_i$ เมื่อ $r_i \in R$ ค่า Ms นี้จะบ่งบอกระดับความสัมพันธ์ดังแสดงในรูปที่ 2.3 ซึ่งอธิบายได้ดังนี้

- ถ้าค่า $Ms = \alpha$ จะมีความสัมพันธ์เป็นแบบเท่ากัน (Equivalence เมื่อ $\alpha = \beta$)หรือแบบเป็นส่วนภายใน (Inclusion เมื่อ $\alpha < \beta$)
- ถ้าค่า $\alpha > Ms \geq \alpha/2$ จะมีความสัมพันธ์แบบมีส่วนเหมือนกันมาก (Tight-intersect)
- ถ้าค่า $\alpha/2 > Ms > 0$ จะมีความสัมพันธ์แบบมีส่วนเหมือนกันบ้าง (Loose-intersect)
- ถ้าค่า $Ms = 0.1 * \alpha$ จะมีความสัมพันธ์แบบมีค่าต่างกัน (Val-disjunction)

- ถ้าค่า Ms = 0 จะมีความสัมพันธ์แบบแตกต่างกัน (Total-disjunction)

งานวิจัยนี้ได้สรุปกฎการรวมเทอมสำหรับรูปแบบความสัมพันธ์จากรูปที่ 2.3 ไว้เพียงบางรูปแบบ ดังแสดงในตารางที่ 2.1



รูปที่ 2.3 ความสัมพันธ์จากค่าวิทยาการศึกษาสำนึก

ตารางที่ 2.1 ผลที่ได้จากค่าความสัมพันธ์

Relationship	Result
Equivalence	T1 = T2
Inclusion	T2 อิสอะ T1
Val-disjunction	สร้างเทอมทั่วไป (General Term) สำหรับ T1 และ T2
Total-disjunction	T1 แตกต่าง (Disjoint) กับ T2

ผลที่ได้จากตารางที่ 2.1 ได้แก่

- T1 = T2 หมายความว่า T1 สามารถรวมเข้ากับ T2 ให้เป็นเทอมเดียวกันได้
- T2 อิสอะ T1 หมายความว่า T2 เป็นรูปแบบหนึ่งของ T1
- สร้างเทอมทั่วไปสำหรับ T1 และ T2 หมายความว่า จะต้องสร้างเทอมที่เป็นลักษณะทั่วไประหว่าง T1 และ T2 เพื่อครอบคลุมความหมายของเทอมทั้งสอง
- T1 แตกต่างกับ T2 หมายความว่า T1 ไม่มีความสัมพันธ์กับ T2

ตัวอย่างในการเปรียบเทียบ 2 เทอม แสดงได้ดังนี้

Salary (Currency: "Dollar", Period: "month")

Wage (Currency: "Rupee", Period: "week")

ทั้งสองมีความหมายแรกคือ Currency ชื่อเหมือนกันแต่ค่าต่างกัน และความหมายที่สองคือ Period ก็มีชื่อเหมือนกันแต่ค่าต่างกัน

ฉะนั้น เวกเตอร์ R = [0.1, 0.1] และ Ms = 0.2 โดย $\alpha = 2$

ดังนั้น Salary และ Wage มีความสัมพันธ์แบบมีค่าต่างกัน (Val-disjunction) ตามรูปที่ 2.3

จากความสัมพันธ์ที่ได้แบบมีค่าต่างกัน ตารางที่ 2.1 ระบุว่าให้ทำการสร้าง เทอมทั่วไประหว่าง Salary และ Wage ซึ่งอาจให้เทอมทั่วไปคือ Salary ซึ่งมีความหมายเป็น Currency : "Dollar", Period : "month" ในการเข้าถึงเทอมทั่วไปนี้ก็จะเข้าถึงทั้ง Salary และ Wage แต่ในการเข้าถึง Wage จะต้องแปลง Wage ให้มีความหมายเดียวกับ Salary ด้วยคือแปลงเป็นหน่วยดอลลาร์และคิดช่วงเวลาเป็นเดือน

ในส่วนของวิทยานิพนธ์นี้จะนำเสนอการนำแนวคิดข้างต้นมาปรับใช้กับแนวคิดเชิงวัตถุ โดยให้ถือว่าแต่ละวัตถุหรือแต่ละคลาสจะมีทั้งความหมายของคลาสและชุดแอตทริบิวท์ของคลาส ซึ่งเป็นเสมือนคุณลักษณะของคลาส และแต่ละแอตทริบิวท์ก็จะมีความหมายของมันเอง ซึ่งจะสามารถนำค่าวิทยาการศึกษาลำนึกของงานวิจัยนี้มาปรับใช้ได้ นอกจากนี้จะเห็นได้ว่างานวิจัย [7] นี้ได้พิจารณาค่าของความหมายที่มีค่าได้ค่าเดียว ซึ่งถ้ามีการให้ค่าของความหมายได้หลายค่าน่าจะทำให้การเปรียบเทียบถูกต้องได้มากยิ่งขึ้น รวมทั้งงานวิจัย [7] ก็ยังไม่ได้กล่าวถึงความสัมพันธ์ในลักษณะที่เทอมหนึ่งสามารถเป็นส่วนประกอบของอีกเทอมหนึ่ง (Compound) รวมไปถึงความสัมพันธ์แบบแอกกรีเกชันด้วย ดังนั้นวิทยานิพนธ์นี้จึงได้นำแนวคิดของการใช้ค่าวิทยาการศึกษาลำนึกของงานวิจัย [7] มาพัฒนาเพื่อใช้ในการรวมแบบจำลองข้อมูลเชิงวัตถุในลักษณะที่แตกต่างออกไปโดยพิจารณาความสัมพันธ์ระหว่างคลาสในรูปแบบต่างๆ ตามที่ได้กล่าวมาแล้วร่วมด้วย ซึ่งผลที่ได้สำหรับวิทยานิพนธ์นี้จะนำมาเปรียบเทียบกับผลที่ได้จากแนวคิดของงานวิจัย [7] ด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย