สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
วิธีการเปรียบเทียบเพื่อหาความสัมพันธ์ของแบบจำลองข้อมูล 2 ชุด

การรวมแบบจำลองข้อมูลที่นำเสนอในวิทยานิพนธ์ฉบับนี้จะเป็นการรวมแบบจำลองข้อมูลเชิงวัตถุ (Object-Oriented Data Model) 2 ชุด ซึ่งสามารถแบ่งเป็น 2 ขั้นตอน คือขั้นตอนแรกการเปรียบเทียบเทียบแบบจำลองข้อมูลทั้งสอง ซึ่งความสัมพันธ์ที่ได้จากการเปรียบเทียบจะใช้พิจารณาในขั้นตอนต่อไปคือขั้นตอนการรวมซึ่งจะทำการรวมเป็นแบบจำลองข้อมูลรวม โดยในบทนี้จะกล่าวถึงรูปแบบของแบบจำลองข้อมูลที่จะนำมาเปรียบเทียบ ตามด้วยขั้นตอนการเปรียบเทียบ สำหรับขั้นตอนวิธีการรวมแบบจำลองข้อมูลจะกล่าวถึงในบทต่อไป

3.1 รูปแบบของแบบจำลองข้อมูลสำหรับการรวม

รูปที่ 3.1 เป็นตัวอย่างแบบจำลองข้อมูลที่เกี่ยวกับนักศึกษาจากมหาวิทยาลัยสองแห่ง ซึ่งแสดงอยู่ในรูปแบบที่กำหนดด้วยสัญกรณ์บีเอ็นเอฟ (BNF Notation) ดังในรูปที่ 3.2 รูปแบบของแบบจำลองข้อมูลดังกล่าวนี้จะมีส่วนเพิ่มเติมจากแบบจำลองข้อมูลเชิงวัตถุโดยทั่วไปที่แสดงเฉพาะความสัมพันธ์ของคลาสต่างๆ และแอตทริบิวท์ต่างๆ ในแต่ละคลาสเท่านั้น โดยการเพิ่มเติมจะเป็นไปใน 2 ลักษณะดังนี้คือ

1) การเพิ่มความหมาย (Semantics) ให้กับแบบจำลองข้อมูล
2) การระบุคำที่มีความหมายเหมือนกัน (Synonym) หรือเป็นส่วนประกอบกัน (Compound)

3.1.1 การเพิ่มความหมายให้กับแบบจำลองข้อมูล

แบบจำลองข้อมูลทั่วไปประกอบไปด้วยคลาสต่างๆ ที่มีความสัมพันธ์กัน โดยคลาสหนึ่งๆ อาจได้รับการออกแบบให้มีความหมายในลักษณะหนึ่ง เช่นเดียวกับแอตทริบิวท์ที่เป็นส่วนประกอบของคลาสก็อาจถูกออกแบบให้มีความหมายต่างกันไปได้ ฉะนั้นการระบุความหมายที่ชัดเจนสำหรับคลาสหรือแอตทริบิวท์จึงมีความสำคัญอย่างยิ่งสำหรับขั้นตอนการเปรียบเทียบแบบจำลองข้อมูลซึ่งพิจารณาความเหมือนหรือความแตกต่างกันของแบบจำลอง โดยผลของการเปรียบเทียบซึ่งจะแสดงถึงความสัมพันธ์ระหว่างคลาสต่างๆ ในแบบจำลองข้อมูลทั้งสองจะถูกนำมาใช้ในขั้นตอนการรวมแบบจำลองต่อไป

ความหมายของคลาส (Class Semantics) หรือความหมายของแอตทริบิวท์ (Attribute Semantics) ที่จะเพิ่มในแบบจำลองข้อมูลจะอยู่ในรูปของชื่อของความหมายและค่าของความหมาย ซึ่งสามารถมีได้หลายค่า เช่น จากรูปที่ 3.1

ความหมายของคลาส Student ของ Schema1 คือ

Faculty = { 'science' , 'engineer' }

Major = { 'computer' , 'math' }

ความหมายของแอตทริบิวท์ height ของคลาส Student ของ Schema1 คือ

unit-of-height = { 'centimetre' }

Schema1	Schema2
Class Student Class semantics : Faculty={'science','engineer'} Major={'computer','math'} Integrity constraint : sex={'F','M'} Attribute : name : string(40) sex : string(1) height : real <unit-of-height={'centimetre'}> address : string(60) faculty : Faculty_Detail gpa : real Class Faculty_Detail Attribute : name : string(20) dean : string(40) address : string(25)	Class Person Integrity constraint : gender={'Female','Male'} Attribute : name : string(40) gender : string(10) height : real <unit-of-height={'metre'}> home_no : string(10) home_city : string(20) Class GradStudent ISA Person Class semantics : Faculty={'science','account','engineer'} Major={'computer','stat','math','electronics'} Attribute : faculty_name : string(20) gpa : real thesis : string(60) lab_name : string(15)
Synonym-list : Schema1.Student.sex = Schema2.GradStudent.gender Schema1.Student.faculty.name = Schema2.GradStudent.faculty_name Compound-list : Schema1.Student.address = Schema2.GradStudent.home_no + Schema2.GradStudent.home_city	

รูปที่ 3.1 ตัวอย่างของแบบจำลองข้อมูล 2 ชุด

```
<schema>            ::= <schema-name> <class-list>
<schema-name>       ::= <schema-name1> | <schema-name2>
<schema-name1>      ::= string              <schema-name2>    ::= string
<class-list>        ::= <class> | <class> <class-list>
<class>             ::= <classDefinition> ['Attribute :' <attribute-list>]
<classDefinition>   ::= 'Class' <class-name>['ISA' <class-name-list>][<class-semantic-list>]
                    [<integrity-constraint>]
<class-name-list>   ::= <class-name> | <class-name><class-name-list>
<class-name>        ::= string              <attribute-name>  ::= string
<integrity-constraint> ::= <attribute-name>+ '=' + '{' + <string-list>|<numeric-list>+'}' |
                    <attribute-name> + '=' + <numeric-string>+'<->'+ <attribute-name>+ '=' + <numeric-string>
                    [<integrity-constraint>]
<string-list>       ::= string | string + ',' + <string-list>
<numeric-list>      ::= numeric | numeric + ',' + <numeric-list>
<numeric-string>    ::= numeric | string
<class-semantic-list> ::= <meaning> | <meaning> <class-semantic-list>
<attribute-list>    ::= <attribute> | <attribute> <attribute-list>
<attribute> ::= <attribute-name>+ ':' +<attribute-type>['<'+<attribute-semantic-list>+'>']['<isSet>']
<attribute-type>    ::= <basicType> | <class-name>
<basicType>         ::= 'integer' | 'real' | 'string' + '(' + <length> + ')'
<length>            ::= numeric
<attribute-semantic-list> ::= <meaning> | <meaning><attribute-semantic-list>
<meaning>           ::= <meaning-name> + '=' + '{' + <meaning-value-list> + '}'
<meaning-value-list> ::= <string-list>|<numeric-list>
<meaning-name>      ::= string
```

(ก)

```
<synonym-list>      ::= <synonym-item>|<synonym-item><synonym-list>
<synonym-item>      ::=<schema-class-path1> '=' <schema-class-path2>|<schema-attribute-path1>
                    '='<schema-attribute-path2>
<compound-list>     ::= <compound-statement>|<compound-statement><compound-list>
<compound-statement> ::=<schema-attribute-path1>'='<schema-attribute-path2>'+'<compound-item2>|
                    <schema-attribute-path2>'='<schema-attribute-path1>'+'<compound-item1>
<compound-item1>    ::=<schema-attribute-path1>|<schema-attribute-path1>'+'<compound-item1>
<compound-item2>    ::=<schema-attribute-path2>|<schema-attribute-path2>'+'<compound-item2>
<schema-attribute-path1> ::= <schema-name1> + '.' + <class-name> + '.' + <attribute-path>
<schema-attribute-path2> ::= <schema-name2> + '.' + <class-name> + '.' + <attribute-path>
<schema-class-path1>  ::= <schema-name1> + '.' + <class-name>
<schema-class-path2>  ::= <schema-name2> + '.' + <class-name>
<attribute-path>      ::= <attribute-name> | <attribute-name> + '.' + <attribute-path>
```

(ข)

รูปที่ 3.2 (ก) ไวยากรณ์สำหรับรูปแบบของแบบจำลองข้อมูล 1 ชุด (ข) ไวยากรณ์สำหรับการระบุคำเหมือนหรือการเป็นส่วนประกอบกันของแบบจำลองข้อมูล 2 ชุด

3.1.2 การระบุคำที่มีความหมายเหมือนกันหรือคำที่เป็นส่วนประกอบกัน.

การออกแบบแบบจำลองข้อมูลโดยผู้ออกแบบคนละคนนั้น ผู้ออกแบบมีโอกาสที่จะตั้งชื่อของคลาสหรือแอตทริบิวท์ที่ต่างกัน แต่สื่อถึงสิ่งเดียวกันได้ ซึ่งหมายถึงคำที่มีความหมายเหมือนกัน และมีโอกาสที่แอตทริบิวท์หนึ่งแอตทริบิวท์ในคลาสของแบบจำลองข้อมูลหนึ่งสามารถแทนได้ด้วยการรวมกันจากหลายแอตทริบิวท์ของคลาสในอีกแบบจำลองข้อมูลหนึ่ง ซึ่งหมายถึงคำที่เป็นส่วนประกอบกัน ฉะนั้นเพื่อแก้ปัญหาในการบ่งชี้ว่าคำใดมีความหมายเหมือนกันหรือมีลักษณะเป็นส่วนประกอบกัน ผู้ทำการรวมแบบจำลองจึงต้องเป็นผู้ระบุข้อมูลดังกล่าวด้วย โดยการระบุคำที่มีความหมายเหมือนกัน ได้แก่ การระบุว่าชื่อคลาสในแบบจำลองที่ 1 มีความหมายเหมือนกับชื่อคลาสในแบบจำลองข้อมูลที่ 2 หรือการระบุว่าชื่อแอตทริบิวท์ของคลาสหนึ่งในแบบจำลองที่ 1 มีความหมายเหมือนกับชื่อแอตทริบิวท์ของคลาสหนึ่งในแบบจำลองข้อมูลที่ 2 ส่วนการระบุคำที่เป็นส่วนประกอบกัน ได้แก่ การระบุว่าแอตทริบิวท์จำนวนหนึ่งในคลาสหนึ่งของแบบจำลองที่ 1 สามารถประกอบเป็นแอตทริบิวท์ 1 ตัวในคลาสหนึ่งของแบบจำลองข้อมูลที่ 2 ได้ นอกจากนี้การระบุคำที่มีความหมายเหมือนกันหรือคำที่เป็นส่วนประกอบกันสามารถระบุผ่านความสัมพันธ์ประเภทแอกกรีเกชันได้ด้วย โดยมีหลักการว่าสิ่งนั้นต้องมีความหมายเดียวกัน เช่น จากรูปที่ 3.1

การระบุคำที่มีความหมายเหมือนกัน ได้แก่

Schema1.Student.sex = Schema2.GradStudent.gender

Schema1.Student.faculty.name = Schema2.GradStudent.faculty_name

การระบุคำที่เป็นส่วนประกอบกัน ได้แก่

Schema1.Student.address = Schema2.GradStudent.home_no
+ Schema2.GradStudent.home_city

การเพิ่มความหมายและการระบุคำที่มีความหมายเหมือนกันหรือคำที่เป็นส่วนประกอบกันสำหรับแบบจำลองข้อมูล 2 ชุด เป็นสิ่งสำคัญที่จะช่วยในการพิจารณาหาความเหมือนหรือความแตกต่างสำหรับขั้นตอนการเปรียบเทียบที่จะกล่าวถึงในบทถัดไป ดังนั้นผู้ทำการรวมแบบจำลองข้อมูลควรมีหลักในการระบุดังนี้

- ถ้าสิ่ง 2 สิ่งมีชื่อต่างกัน แต่สื่อด้วยความหมายเดียวกัน ควรระบุว่าเป็นคำที่มีความหมายเหมือนกัน
- ถ้าสิ่ง 2 สิ่งมีชื่อเหมือนกัน แต่ความหมายต่างกัน ควรระบุความหมายเพิ่มให้ต่างกัน (มีชื่อของความหมายที่ต่างกัน)

- ถ้าสิ่ง 2 สิ่งมีชื่อเหมือนกัน และสื่อด้วยความหมายเดียวกัน ควรระบุความหมายที่ใกล้เคียงกัน (มีชื่อของความหมายเหมือนกัน แต่ค่าของความหมายอาจต่างกันได้)
- ถ้าสิ่งหนึ่งประกอบด้วยสิ่งหลายสิ่ง ควรระบุว่าเป็นคำที่เป็นส่วนประกอบกัน

นอกจากนี้การตั้งชื่อคลาสและแอตทริบิวท์ควรตั้งชื่อให้มีความหมายและอยู่ในรูปเอกพจน์ [13] รวมทั้งควรตั้งชื่อให้แตกต่างกันในขอบเขตหนึ่ง เช่น ชื่อคลาสภายในแบบจำลองข้อมูลหรือสกีมาหนึ่งควรตั้งชื่อให้แตกต่างกันและชื่อแอตทริบิวท์ภายในคลาสหนึ่งควรจะตั้งชื่อให้แตกต่างกัน [14]

3.2 วิธีการเปรียบเทียบแบบจำลองข้อมูล

วิธีการเปรียบเทียบเพื่อหาความสัมพันธ์ของแบบจำลองข้อมูล 2 ชุดจะใช้ส่วนประกอบต่างๆ ที่มีในแบบจำลองมาเป็นตัวพิจารณา ซึ่งวิธีการเปรียบเทียบเพื่อหาความสัมพันธ์นี้สามารถแบ่งได้เป็น 3 ส่วน ดังนี้คือ

1) การเปรียบเทียบความหมาย (Semantic Comparison) เป็นการเปรียบเทียบความหมายที่ผู้ทำการรวมแบบจำลองได้ระบุเพิ่มเติมให้กับคลาสหรือแอตทริบิวท์
2) การเปรียบเทียบแอตทริบิวท์ (Attribute Comparison) เป็นการเปรียบเทียบแอตทริบิวท์ต่างๆ ภายในคลาสที่นำมาเปรียบเทียบกัน โดยพิจารณาความหมายของแอตทริบิวท์ที่มีการระบุเพิ่มเติมด้วย (ดังข้อ 1)
3) การเปรียบเทียบคลาส (Class Comparison) เป็นการเปรียบเทียบเพื่อหาความสัมพันธ์ระหว่างคลาส โดยพิจารณาความหมายของคลาส (ดังข้อ 1) รวมทั้งพิจารณาความสัมพันธ์ของแอตทริบิวท์ (ดังข้อ 2)

3.2.1 การเปรียบเทียบความหมาย

วิทยานิพนธ์นี้จะแสดงวิธีการนำค่าวิทยาการศึกษาสำนึกมาใช้ โดยพัฒนาเพิ่มเติมจาก [7] ซึ่งค่าวิทยาการศึกษาสำนึกที่ได้จะระบุความสัมพันธ์ระหว่างแบบจำลองข้อมูลที่นำมาเปรียบเทียบ โดยวิธีการเปรียบเทียบความหมายมีดังนี้ คือ

ให้ T1 และ T2 เป็นเทอม (Term) ที่จะพิจารณาเปรียบเทียบความหมาย ซึ่งเทอมนี้อาจหมายถึงคลาสหรือแอตทริบิวท์ ดังนั้น ความหมายที่จะเปรียบเทียบอาจหมายถึงความหมายของคลาสหรือความหมายของแอตทริบิวท์ตามลำดับ

และให้ α เป็นจำนวนความหมายของ T1

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

การเปรียบเทียบแต่ละคู่ของความหมายจะทำการเปรียบเทียบชื่อและค่าของความหมาย ความสัมพันธ์ที่ได้จากการเปรียบเทียบจะแทนได้ด้วยค่า r ดังแสดงในตารางที่ 3.1

การกำหนดค่า r จากการเปรียบเทียบ 1 คู่ความหมายซึ่งประกอบไปด้วยชื่อของความหมายและค่าของความหมาย จะพิจารณาจากอัตราส่วนระหว่างความเหมือนของชื่อของความหมายกับความเหมือนของค่าของความหมายเป็น 1 : 1 โดยที่ ค่าความเหมือนเป็น 1 แสดงความเหมือนกัน ค่าความเหมือนเป็น 0 แสดงความต่างกัน และค่าความเหมือนระหว่าง 0 ถึง 1 แสดงการเป็นสับเซตหรือมีบางส่วนที่เหมือนกัน เมื่อให้ค่า r = ผลรวมของอัตราส่วนการเปรียบเทียบชื่อและค่าของความหมายหารด้วย 2 ค่า r จะอยู่ในช่วง 0-1 ซึ่งถ้าค่าใกล้ 1 แสดงว่าความหมายทั้งคู่เหมือนกันมาก ถ้าค่าเป็น 0 แสดงว่าทั้งคู่มีความหมายต่างกัน ดังอธิบายได้ดังนี้

- ถ้าชื่อของความหมายเหมือนกันและค่าของความหมายเหมือนกัน จะได้อัตราส่วนเป็น 1 : 1 จะได้ r = 1
- ถ้าชื่อของความหมายเหมือนกันและค่าของความหมายเป็นสับเซตกัน จะได้อัตราส่วนเป็น 1 : 0.1 - 0.9 จะได้ r อยู่ในช่วง 0.55-0.95 ซึ่งจากผลการทดลองโดยใช้ค่า r ต่างๆ กัน (อธิบายในบทที่ 5) จะได้ค่าที่เหมาะสมสำหรับค่า r นี้คือ 0.9
- ถ้าชื่อของความหมายเหมือนกันและค่าของความหมายมีส่วนเหมือนกันบางส่วน จะได้อัตราส่วนเป็น 1 : 0.1 - 0.9 จะได้ r อยู่ในช่วง 0.55-0.95 ซึ่งจากผลการทดลองโดยใช้ค่า r ต่างๆ กัน (อธิบายในบทที่ 5) จะได้ค่าที่เหมาะสมสำหรับค่า r นี้คือ 0.6
- ถ้าชื่อของความหมายเหมือนกันแต่ค่าของความหมายต่างกัน จะได้อัตราส่วนเป็น 1 : 0 จะได้ r = 0.5
- ถ้าชื่อของความหมายต่างกันและค่าของความหมายเป็นค่าใดๆ กรณีนี้เราจะถือว่าความหมายที่ทำการเปรียบเทียบต่างกัน เพราะชื่อต่างกันแสดงถึงความหมายที่ต่างกันแม้ว่าจะมีค่าที่เหมือนกัน นั่นคือ r = 0

ผลของการเปรียบเทียบแต่ละคู่ของความหมายของเทอม T1 และ T2 จะเก็บไว้ในเวกเตอร์ (Vector) $R = [r_1, r_2, \ldots, r_\alpha]$ และ $Ms = \sum r_i$ โดยที่ $r_i \in R$ และ $i = 1, \ldots, \alpha$ ค่า Ms ที่ได้จะแสดงระดับความเหมือนและความสัมพันธ์ระหว่าง T1 และ T2 ดังแสดงในตารางที่ 3.2 (สำหรับค่า s ในตารางจะได้กล่าวถึงในหัวข้อ 3.2.2)

ตารางที่ 3.1 ค่าความสัมพันธ์จากการเปรียบเทียบแต่ละคู่ของความหมาย

ชื่อของความหมาย	ค่าของความหมาย	r
เหมือนกัน	เหมือนกัน	1
เหมือนกัน	เป็นสับเซตกัน	0.8
เหมือนกัน	มีส่วนเหมือนกันบางส่วน (Overlap)	0.8
เหมือนกัน	ไม่เหมือนกันเลย	0.5
ต่างกัน	เป็นค่าใดๆ	0

ตารางที่ 3.2 ความสัมพันธ์จากการเปรียบเทียบชุดของความหมายของเทอม T1 และ T2

Ms	ระดับความเหมือน	ความสัมพันธ์	s
α (α=β)	Equivalence	Equivalence	1
α (α<β)	Inclusion	Superclass*	1
$\geq$ ¾α (α=β)	Equivalence	Equivalence	1
$\geq$ ¾α (α<β)	Inclusion	Superclass*	1
$\geq$ ½α	Tight-Intersection	Sibling	1
< ½α	Loose-Intersection	Disjoint	0
0	Disjoint	Disjoint	0

*T1 is Superclass of T2 = T2 is Subclass of T1

ทั้งนี้ความสัมพันธ์ที่ได้จากตารางที่ 3.2 มีความหมายดังนี้

- ความสัมพันธ์แบบเท่ากัน (Equivalence) หมายถึงเทอม T1 และ T2 มีความหมายส่วนใหญ่เหมือนกัน (กรณี Ms $\geq$ 3/4α) หรือความหมายเหมือนกันหมด (กรณี Ms = α)
- ความสัมพันธ์แบบซูเปอร์คลาส (Superclass) หมายถึงเทอม T1 มีความหมายน้อยกว่า T2 และ T2 มีความหมายส่วนหนึ่งเหมือนกับความหมายทั้งหมดของ T1
- ความสัมพันธ์แบบพี่น้อง (Sibling) หมายถึงความหมายส่วนหนึ่งในชุดความหมายของ T1 เหมือนกับความหมายส่วนหนึ่งในชุดความหมายของ T2 หรือ T1 และ T2 มีชื่อของความหมายเหมือนกันหมด แต่ค่าของความหมายต่างกัน
- ความสัมพันธ์แบบแตกต่างกัน (Disjoint) หมายถึงเทอม T1 และ T2 มีความหมายส่วนใหญ่ไม่เหมือนกัน

รูปที่ 3.3 เป็นตัวอย่างคลาส Student ใน Schema1 และคลาส GradStudent ใน Schema2 จากรูปที่ 3.1 ซึ่งแสดงแอตทริบิวท์ทั้งหมดรวมทั้งที่สืบทอดจากคลาสอื่นโดยมีส่วนของการระบุคำที่มีความหมายเหมือนกันและคำที่เป็นส่วนประกอบไว้ด้วย การเปรียบเทียบความหมายของคลาส Student ใน Schema1 และคลาส GradStudent ใน Schema2 นั้นจะได้ค่าดังนี้

ให้ T1 = คลาส Student

T2 = คลาส GradStudent

โดยที่ $\alpha = 2$ $\beta = 2$

จากการเปรียบเทียบความหมาย Major จะได้ชื่อของความหมายเหมือนกัน แต่ค่าของความหมายเป็นสับเซตกัน ดังนั้น จะได้ค่า r = 0.8 (จากตารางที่ 3.1)

จากการเปรียบเทียบความหมาย Faculty จะได้ชื่อของความหมายเหมือนกัน แต่ค่าของความหมายเป็นสับเซตกัน ดังนั้น จะได้ค่า r = 0.8

ซึ่งจะได้ R = [0.8,0.8] Ms = 1.8 เมื่อ $\alpha = 2$

ดังนั้น จากตารางที่ 3.2 จะได้ความสัมพันธ์ระหว่าง T1 และ T2 เป็นความสัมพันธ์แบบเท่ากัน

นั่นคือ คลาส Student และคลาส GradStudent มีความหมายที่สัมพันธ์แบบเท่ากัน

ผลที่ได้จากการเปรียบเทียบความหมายของคลาสหรือแอตทริบิวท์จะแสดงถึงระดับความสัมพันธ์ของคลาสหรือแอตทริบิวท์ที่แตกต่างกันออกไป ซึ่งจะมีส่วนช่วยในการตัดสินใจในขั้นตอนการรวมต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Schema1	Schema2
Class Student Class semantics : Faculty={'science','engineer'} Major={'computer','math'} Integrity constraint : sex='F','M' Attribute : name : string(40) sex : string(1) height : real <unit-of-height={'centimetre'}> address : string(60) faculty : Faculty_Detail gpa : real	Class GradStudent Class semantics : Faculty={'science','account','engineer'} Major={'computer','stat','math','electronics'} Integrity constraint : gender='Female','Male' Attribute : name : string(40) gender : string(10) height : real <unit-of-height={'metre'}> home_no : string(10 home_city : string(20) faculty_name : string(20) gpa : real thesis : string(60) lab_name : string(15)
Synonym-list : Schema1.Student.sex = Schema2.GradStudent.gender Schema1.Student.faculty.name = Schema2.GradStudent.faculty_name Compound-list:Schema1.Student.address=Schema2.GradStudent.home_no+Schema2.GradStudent.home_city	

รูปที่ 3.3 ตัวอย่างคลาสบางส่วนจากรูปที่ 3.1

3.2.2 การเปรียบเทียบแอตทริบิวท์

ส่วนประกอบของแอตทริบิวท์ที่ใช้ในการเปรียบเทียบมีดังนี้

1) ชื่อของแอตทริบิวท์
2) ความหมายของแอตทริบิวท์
3) ชนิดข้อมูล
4) ข้อกำหนดบูรณภาพของแอตทริบิวท์ (Integrity Constraint)

ส่วนที่จะใช้พิจารณาระดับความเหมือนหรือความแตกต่างในการเปรียบเทียบกับแอตทริบิวท์อื่นคือ ชื่อแอตทริบิวท์และความหมายของแอตทริบิวท์ สำหรับชนิดข้อมูลและข้อกำหนดบูรณภาพจะใช้พิจารณาในขั้นตอนการรวมกับแอตทริบิวท์อื่นต่อไป

3.2.2.1 การหาความสัมพันธ์จากการเปรียบเทียบแต่ละคู่ของแอตทริบิวท์

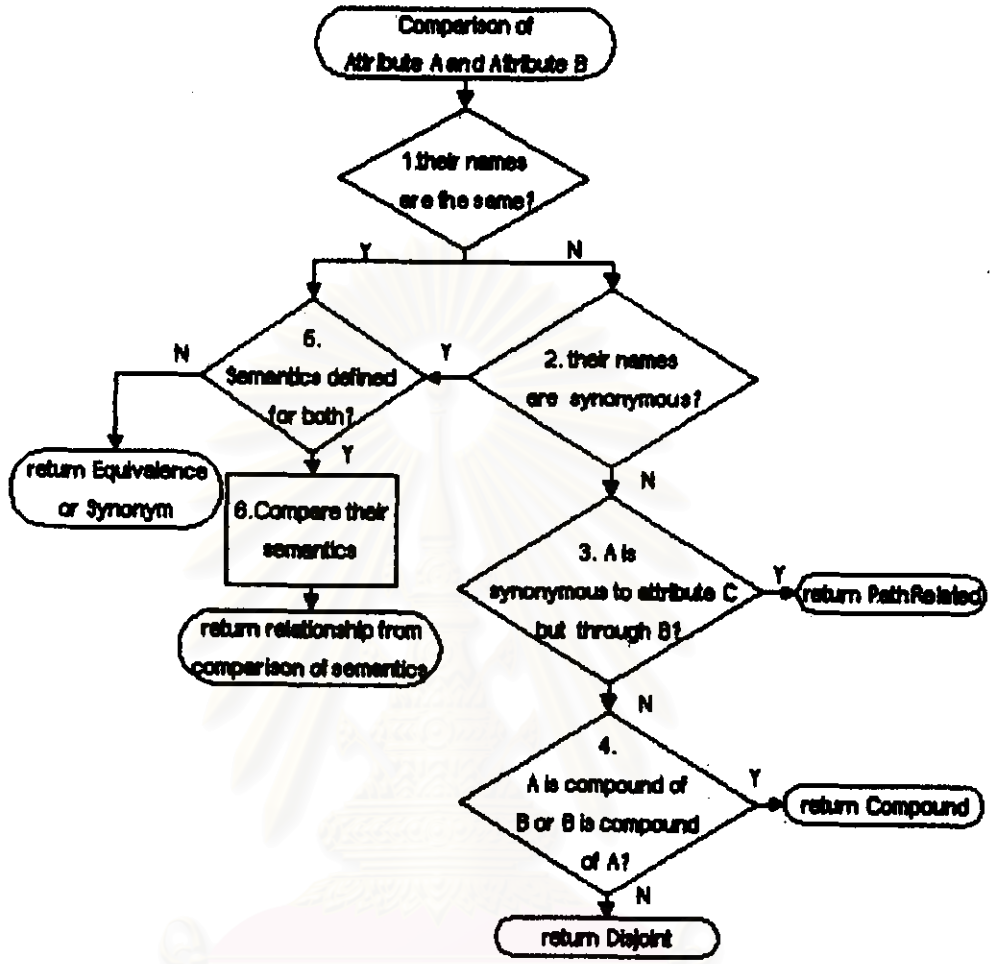
การเปรียบเทียบแอตทริบิวท์ A และ B ของคลาสใดๆ จากแบบจำลองข้อมูล 2 ชุดว่ามีความสัมพันธ์กันหรือไม่ มีขั้นตอนดังแสดงในรูปที่ 3.4 และผลที่ได้จากการเปรียบเทียบ

พจนานุกรม สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แอตทริบิวท์ตามขั้นตอนดังกล่าวจะคืนค่า s ซึ่งเป็นค่าวิทยาการศึกษาสำนึกซึ่งแสดงว่าสามารถทำการรวมกันได้หรือไม่ พร้อมทั้งได้ค่าความสัมพันธ์ดังแสดงในตารางที่ 3.3 โดยค่า s สามารถมีค่าเป็น 0 หรือ 1 ซึ่งถ้าค่าเป็น 1 แสดงว่าแอตทริบิวท์ทั้งคู่มีความสัมพันธ์กันซึ่งสามารถรวมกันได้เป็นแอตทริบิวท์เดียวกัน และค่าเป็น 0 แสดงว่าแอตทริบิวท์ทั้งคู่ไม่มีความสัมพันธ์กันเพียงพอที่จะรวมกันได้ ซึ่งอธิบายได้ดังนี้

- ความสัมพันธ์แบบเท่ากัน (Equivalence) แสดงว่าแอตทริบิวท์ทั้งสองเหมือนกัน ซึ่งเกิดจากการเปรียบเทียบชื่อของแอตทริบิวท์หรือความหมายของแอตทริบิวท์แล้ว มีระดับความสัมพันธ์ที่เหมือนกันมากจนสามารถรวมกันได้เป็นแอตทริบิวท์เดียวกัน จะได้ค่า s = 1
- ความสัมพันธ์แบบซูเปอร์คลาส (Superclass) / ซับคลาส (Subclass)แสดงว่าแอตทริบิวท์ทั้งสองมีการเปรียบเทียบความหมายของแอตทริบิวท์แล้ว แอตทริบิวท์หนึ่งมีความหมายส่วนหนึ่งครอบคลุมความหมายทั้งหมดของอีกแอตทริบิวท์หนึ่ง ซึ่งสามารถรวมกันเป็นหนึ่งแอตทริบิวท์ เพื่อให้ครอบคลุมแอตทริบิวท์ทั้งสองได้ จะได้ค่า s = 1
- ความสัมพันธ์แบบพี่น้อง (Sibling) แสดงว่าแอตทริบิวท์ทั้งสองมีการเปรียบเทียบความหมายของแอตทริบิวท์แล้ว ความหมายส่วนหนึ่งของแอตทริบิวท์หนึ่งเหมือนกับความหมายส่วนหนึ่งของอีกแอตทริบิวท์หนึ่ง ซึ่งในการรวมกันของแอตทริบิวท์ทั้งสองสามารถสร้างแอตทริบิวท์ใหม่ เพื่อแสดงความหมายทั่วไปของแอตทริบิวท์ทั้งสองได้ จะได้ค่า s = 1
- ความสัมพันธ์แบบความหมายเหมือนกัน (Synonym) เป็นการระบุจากผู้ทำการรวมว่าแอตทริบิวท์ทั้งสองมีความหมายเดียวกัน จึงสามารถรวมกันเป็นแอตทริบิวท์เดียวกันได้ จะได้ค่า s = 1
- ความสัมพันธ์แบบเป็นส่วนประกอบกัน (Compound) เป็นการระบุจากผู้ทำการรวมว่าแอตทริบิวท์หนึ่งเป็นส่วนประกอบของอีกแอตทริบิวท์ จึงสามารถรวมกันเป็นแอตทริบิวท์เดียวกันได้ในลักษณะของการเป็นส่วนประกอบ จะได้ค่า s = 1
- ความสัมพันธ์แบบการมีเส้นทางเกี่ยวข้องกัน (PathRelated) สำหรับแอตทริบิวท์ A และ B เป็นความสัมพันธ์จากการระบุจากผู้ทำการรวมว่าแอตทริบิวท์ A มีความสัมพันธ์กับแอตทริบิวท์อื่น โดยผ่านความสัมพันธ์กับแอตทริบิวท์ B ด้วยความสัมพันธ์แบบแอกกรีเกชัน ซึ่งกรณีนี้จะทำการรวมในลักษณะของแอตทริบิวท์ที่มีความหมายเหมือนกัน (ดังจะกล่าวไว้ในบทที่ 4.3.3) จะได้ค่า s = 1

- ความสัมพันธ์แบบแตกต่างกัน (Disjoint) แสดงว่าแอตทริบิวท์ทั้งคู่ไม่มีความสัมพันธ์กัน ซึ่งไม่สามารถรวมกันได้ จะได้ค่า s = 0



รูปที่ 3.4 ขั้นตอนการเปรียบเทียบแต่ละคู่ของแอตทริบิวท์เพื่อหาค่าความสัมพันธ์

ตารางที่ 3.3 ค่าความสัมพันธ์จากการเปรียบเทียบแอตทริบิวท์

การเปรียบเทียบแอตทริบิวท์ A และ B	s	ความสัมพันธ์
ชื่อเหมือนกันและทั้งคู่มีความหมายของแอตทริบิวท์	ได้ค่า s จากตารางที่ 3.2 จากการเปรียบเทียบความหมาย	ได้ค่าความสัมพันธ์จากตารางที่ 3.2 จากการเปรียบเทียบความหมาย
ชื่อต่างกันแต่เป็นคำที่มีความหมายเหมือนกันและทั้งคู่มีความหมายของแอตทริบิวท์	ได้ค่า s จากตารางที่ 3.2 จากการเปรียบเทียบความหมาย	ได้ค่าความสัมพันธ์จากตารางที่ 3.2 จากการเปรียบเทียบความหมาย
ชื่อเหมือนกันและไม่มีความหมายของแอตทริบิวท์	1	Equivalence
ชื่อต่างกันแต่เป็นคำที่มีความหมายเหมือนกันและไม่มีความหมายของแอตทริบิวท์	1	Synonym
แอตทริบิวท์ A มีเป็นคำที่มีความหมายเหมือนกับแอตทริบิวท์ที่ผ่านโดยผ่านทางความสัมพันธ์กับแอตทริบิวท์ B	1	PathRelated
มีความสัมพันธ์ในลักษณะของคำที่เป็นส่วนประกอบกัน	1	Compound
ไม่มีความสัมพันธ์กัน	0	Disjoint

ลักษณะของการเปรียบเทียบแอตทริบิวท์ A และ B สามารถแบ่งได้เป็นหลายกรณีดังนี้ คือ

<u>กรณีที่ 1</u> ชื่อเหมือนกัน (ขั้นตอนที่ 1 จากรูปที่ 3.4)

เมื่อแอตทริบิวท์ที่ทำการเปรียบเทียบมีชื่อเหมือนกัน จะทำการเปรียบเทียบความหมายของแอตทริบิวท์ทั้งสอง กรณีที่อย่างน้อยแอตทริบิวท์หนึ่งไม่มีการระบุความหมายเพิ่มเติมจะสรุปว่า ทั้งสองมีความสัมพันธ์แบบเท่ากันซึ่งจะคืนค่า s=1 แต่ถ้าทั้งสองมีการระบุความหมายเพิ่มเติม ค่าความสัมพันธ์และค่า s จะได้จากการเปรียบเทียบความหมายของแอตทริบิวท์ดังตารางที่ 3.2

ตัวอย่างการเปรียบเทียบแอตทริบิวท์ที่ชื่อเหมือนกัน จากแอตทริบิวท์ height ของคลาส Student และแอตทริบิวท์ height ของคลาส GradStudent ในรูปที่ 3.3 มีดังนี้

height : real //class Student (schema1)

<unit-of-height={'centimetre'}>

height : real //class GradStudent (schema2)

<unit-of-height={'metre'}>

เนื่องจากทั้งสองแอตทริบิวท์มีชื่อเหมือนกันและมีการกำหนดความหมายเพิ่มเติมทั้งคู่ จึงพิจารณาความสัมพันธ์ของแอตทริบิวท์ทั้งสองจากความหมายที่กำหนด

∴ จากตารางที่ 3.2 จะได้ R=[0.5] Ms=0.5

ระดับความเหมือน = มีส่วนเหมือนกันมาก (Tight-intersection)

s=0.1 ความสัมพันธ์=แบบพี่น้อง (Sibling)

นั่นคือ แอตทริบิวท์ทั้งสองมีความสัมพันธ์แบบพี่น้อง

<u>กรณีที่ 2</u> ชื่อต่างกันแต่มีความหมายเหมือนกัน (ขั้นตอนที่ 2 จากรูปที่ 3.4)

เมื่อแอตทริบิวท์ที่ทำการเปรียบเทียบมีชื่อต่างกัน จะทำการตรวจสอบว่าทั้งสองเป็นคำที่มีความหมายเหมือนกันตามที่ผู้ทำการรวมได้ระบุไว้หรือไม่ ถ้าทั้งสองเป็นคำที่มีความหมายเหมือนกัน จะทำการเปรียบเทียบความหมายของแอตทริบิวท์ในลักษณะเดียวกับกรณีที่ 1

ตัวอย่างการเปรียบเทียบแอตทริบิวท์ที่ชื่อต่างกัน แต่ความหมายเหมือนกัน จากรูปที่ 3.3

sex : string(1) //Class Student (Schema1)

gender : string(1) //Class GradStudent (Schema2)

<Synonym-list : Schema1.Student.sex = Schema2.GradStudent.gender>

∴ จากตารางที่ 3.3 จะได้

s=1 ความสัมพันธ์=ความหมายเหมือนกัน (Synonym)

นั่นคือ แอตทริบิวท์ทั้งสองมีความสัมพันธ์ในลักษณะความหมายเหมือนกัน

กรณีที่ 3 แอตทริบิวท์ A เป็นคำที่มีความหมายเหมือนกับแอตทริบิวท์อื่น โดยผ่านทางความสัมพันธ์กับแอตทริบิวท์ B (ขั้นตอนที่ 3 จากรูปที่ 3.4)

กรณีนี้เกิดจากการระบุคำที่มีความหมายเหมือนกันผ่านความสัมพันธ์ในลักษณะแอกกรีเกชันไปยังแอตทริบิวท์อื่น เช่น จากรูปที่ 3.3 ส่วนของการระบุคำที่มีความหมายเหมือนกัน ได้แก่

Schema1.Student.faculty.name = Schema2.GradStudent.faculty_name

สังเกตได้ว่าแอตทริบิวท์ faculty ของคลาส Student มีความสัมพันธ์แบบแอกกรีเกชันกับคลาส Faculty_Detail ในแบบจำลองข้อมูลเดียวกัน โดยแอตทริบิวท์ที่มีความหมายเหมือนกันจริงๆ คือแอตทริบิวท์ name ในคลาส Faculty_Detail (ผ่านความสัมพันธ์ของแอตทริบิวท์ faculty ของคลาส Student) กับแอตทริบิวท์ faculty_name ของคลาส GradStudent

การเปรียบเทียบแอตทริบิวท์ faculty จากคลาส Student และแอตทริบิวท์ faculty_name จากคลาส GradStudent จะได้ค่าดังนี้

faculty : Faculty_Detail //Class Student (Schema1)

faculty_name : string(35) //Class GradStudent (Schema2)

<Synonym-list :

Schema1.Student.faculty.name = Schema2.GradStudent.faculty_name>

∴ จากตารางที่ 3.3 จะได้

s=1 ความสัมพันธ์=มีเส้นทางเกี่ยวข้องกัน (PathRelated)

นั่นคือ แอตทริบิวท์ทั้งสองมีความสัมพันธ์ในลักษณะการมีเส้นทางเกี่ยวข้องกัน

กรณีที่ 4 มีความสัมพันธ์ในลักษณะคำที่เป็นส่วนประกอบกัน (ขั้นตอนที่ 4 จากรูปที่ 3.4)

กรณีนี้เกิดจากการที่แอตทริบิวท์ A เป็นส่วนประกอบหนึ่งของแอตทริบิวท์ B หรือแอตทริบิวท์ B เป็นส่วนประกอบหนึ่งของแอตทริบิวท์ A ทั้งนี้ชุดแอตทริบิวท์ที่เป็นส่วนประกอบของแอตทริบิวท์หนึ่งในอีกแบบจำลองข้อมูลนั้น จะต้องอยู่ภายในคลาสเดียวกัน เช่น ถ้าแอตทริบิวท์ A และ C เป็นส่วนประกอบของแอตทริบิวท์ B แล้ว แอตทริบิวท์ A และ C จะต้องอยู่ในคลาสเดียวกัน ตัวอย่างในการเปรียบเทียบแอตทริบิวท์ address ของ

คลาส Student และแอตทริบิวท์ home_no ของคลาส GradStudent ในรูปที่ 3.3 มีการระบุคำที่เป็นส่วนประกอบกัน คือ

Schema1.Student.address = Schema2.GradStudent.home_no
+ Schema2.GradStudent.home_city

นั่นคือ แอตทริบิวท์ home_no เป็นส่วนประกอบหนึ่งของแอตทริบิวท์ address เมื่อเปรียบเทียบกันโดยตารางที่ 3.3 จะได้ค่า s=1 และความสัมพันธ์=เป็นส่วนประกอบ (Compound) นั่นคือแอตทริบิวท์ทั้งสองมีความสัมพันธ์ในลักษณะการเป็นส่วนประกอบกัน

<u>กรณีที่ 5</u> ไม่มีความสัมพันธ์กัน

นอกเหนือจากกรณีที่ 1-4 จะได้ว่าแอตทริบิวท์ทั้งสองไม่มีความสัมพันธ์กัน นั่นคือจากตารางที่ 3.3 จะได้ s=0 ความสัมพันธ์=แตกต่างกัน (Disjoint)

3.2.2.2 การเปรียบเทียบชนิดข้อมูลและค่าที่เป็นไปได้สำหรับแอตทริบิวท์

เมื่อทำการเปรียบเทียบแอตทริบิวท์และได้ว่าแอตทริบิวท์ทั้งสองมีความสัมพันธ์กัน (ความสัมพันธ์ใดๆ ที่ไม่ใช่ความสัมพันธ์แบบแตกต่างกัน) จะทำการเปรียบเทียบชนิดข้อมูลและค่าที่เป็นไปได้ของแอตทริบิวท์ ดังนี้

1) ในการเปรียบเทียบชนิดข้อมูล ถ้าชนิดข้อมูลไม่เหมือนกัน จะทำการบันทึกว่าถ้าแอตทริบิวท์ทั้งสองถูกรวมกันในขั้นตอนการรวมแบบจำลอง ผู้ทำการรวมจะต้องระบุฟังก์ชันการแปลงเพิ่มเติมสำหรับการแปลงชนิดข้อมูลให้เหมือนกัน
2) ในการเปรียบเทียบข้อกำหนดบูรณภาพของแอตทริบิวท์ จะพิจารณาใน 2 ลักษณะคือ ค่าที่เป็นไปได้ของแอตทริบิวท์กับการขึ้นกับค่าของแอตทริบิวท์อื่น ในการพิจารณาค่าที่เป็นไปได้นั้น ถ้าค่าที่เป็นไปได้ของแอตทริบิวท์หนึ่งไม่อยู่ในขอบเขตของค่าที่เป็นไปได้ในอีกแอตทริบิวท์หนึ่ง จะทำการบันทึกว่าถ้าแอตทริบิวท์ทั้งสองถูกรวมกันในขั้นตอนการรวมแบบจำลอง ผู้ทำการรวมจะต้องระบุฟังก์ชันการแปลงเพิ่มเติมสำหรับการแปลงค่าของแอตทริบิวท์ให้เหมือนกัน เช่น ในการเปรียบเทียบค่าที่เป็นไปได้ของแอตทริบิวท์ sex ซึ่งมีค่าที่เป็นไปได้คือ {'F','M'} กับแอตทริบิวท์ gender ซึ่งมีค่าที่เป็นไปได้คือ {'Female','Male'} เนื่องจากค่าที่เป็นไปได้ของทั้งสองแอตทริบิวท์ไม่เหมือนกันเลย จึงจะมีการบันทึกไว้ว่า ในการรวมจะต้องมีฟังก์ชันในการแปลงค่าของแอตทริบิวท์ทั้งสองให้เหมือนกัน นอกจากนี้จะพิจารณาการขึ้นอยู่กับค่าของแอตทริบิวท์อื่นด้วย (จากการระบุ *<attribute-name>+'='+<numeric-string>+'←→'+<attribute-*

name>+'='+<*numeric-string*> ในสัญกรณ์บีเอ็นเอฟจากรูปที่ 3.2 โดยที่ '↔' เป็นการแสดงว่าค่าของแอตทริบิวท์หนึ่งขึ้นอยู่กับค่าของอีกแอตทริบิวท์หนึ่ง) ถ้ากรณีที่แอตทริบิวท์ A มีการบันทึกว่าต้องมีการแปลงค่าของแอตทริบิวท์ด้วยแล้ว แอตทริบิวท์อื่นที่มีการขึ้นอยู่กับค่าของแอตทริบิวท์ A ก็จะต้องทำการบันทึกด้วยว่าต้องมีฟังก์ชันในการแปลงค่าด้วย

3.2.2.3 การหาความสัมพันธ์จากการเปรียบเทียบชุดแอตทริบิวท์ของคลาส

เมื่อได้ความสัมพันธ์จากการเปรียบเทียบในแต่ละคู่แอตทริบิวท์จากขั้นตอนข้างต้น ในส่วนนี้จะเป็นวิธีการหาความสัมพันธ์จากชุดแอตทริบิวท์ทั้งหมดจาก 2 คลาสที่จะพิจารณาทำการเปรียบเทียบกัน

ให้ C1 และ C2 เป็นคลาสที่จะพิจารณาเปรียบเทียบชุดแอตทริบิวท์

โดยมี θ_{C1} เป็นจำนวนแอตทริบิวท์ที่กำหนดใน C1

θ_{C2} เป็นจำนวนแอตทริบิวท์ที่กำหนดใน C2

α เป็นจำนวนแอตทริบิวท์ทั้งหมดของ C1

β เป็นจำนวนแอตทริบิวท์ทั้งหมดของ C2 โดยที่ $\alpha \leq \beta$

ค่า α และ β หาได้จากขั้นตอนการคำนวณดังนี้

1. $\alpha = \theta_{C1}$,$\beta = \theta_{C2}$
2. ถ้าแอตทริบิวท์หนึ่งใน C1 มีกลุ่มของแอตทริบิวท์ใน C2 เป็นส่วนประกอบ (Compound) แล้ว $\alpha = \alpha - 1 +$ จำนวนแอตทริบิวท์ใน C2 ที่เป็นส่วนประกอบ ดำเนินการซ้ำสำหรับแอตทริบิวท์อื่นใน C1 ที่มีลักษณะเช่นนี้
3. ถ้าแอตทริบิวท์หนึ่งใน C2 มีกลุ่มของแอตทริบิวท์ใน C1 เป็นส่วนประกอบ แล้ว $\beta = \beta - 1 +$ จำนวนแอตทริบิวท์ใน C1 ที่เป็นส่วนประกอบ ดำเนินการซ้ำสำหรับแอตทริบิวท์อื่นใน C2 ที่มีลักษณะเช่นนี้

ผลของการเปรียบเทียบแต่ละแอตทริบิวท์ของ C1 และ C2 จะเก็บไว้ในเวกเตอร์ $R_{all\text{-}attributes} = [s_1, s_2, ..., s_\alpha]$ โดยค่า s_i จะได้จากการเปรียบเทียบเพื่อหาความสัมพันธ์ระหว่างแต่ละคู่ของแอตทริบิวท์ซึ่งได้อธิบายไว้ในข้อ 3.2.2.1 และจะได้ว่า $Ms_{all\text{-}attributes} = \sum s_i$ โดยที่ $s_i \in R_{all\text{-}attributes}$ และ $i = 1,...,\alpha$ ค่า $Ms_{all\text{-}attributes}$ ที่ได้นี้จะแสดงความสัมพันธ์ระหว่างคลาส C1 และ C2 เมื่อพิจารณาเฉพาะชุดแอตทริบิวท์ทั้งหมด ดังแสดงในตารางที่ 3.4

ตารางที่ 3.4 ความสัมพันธ์ที่ได้จากการเปรียบเทียบชุดของแอตทริบิวท์ของคลาส C1 และ C2

$Ms_{all\text{-}attributes}$	ระดับความเหมือน	ความสัมพันธ์
α (α=β)	Equivalence	Equivalence
α (α<β)	Inclusion	Superclass*
$\geq$ ¾α (α=β)	Equivalence	Equivalence
$\geq$ ¾α (α<β)	Inclusion	Superclass*
$\geq$ ½α	Tight-intersection	Sibling
< ½α	Loose-intersection	Disjoint
0	Disjoint	Disjoint

*C1 is Superclass of C2 = C2 is Subclass of C1

จากรูปที่ 3.5 จะสามารถหาความสัมพันธ์จากชุดแอตทริบิวท์ของคลาส Student และคลาส GradStudent ได้ดังนี้

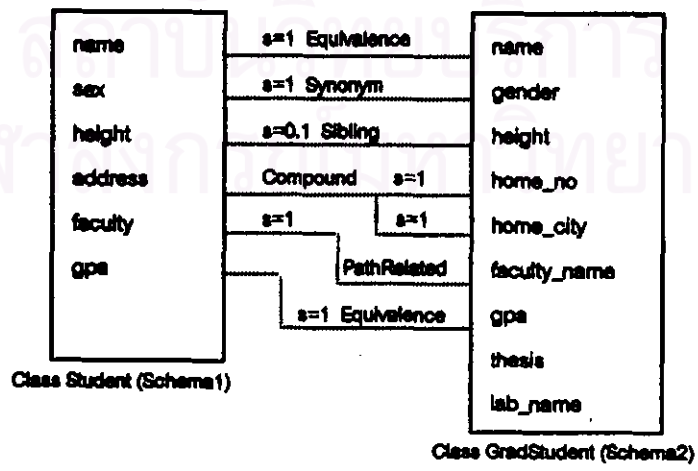
C1 = คลาส Student ใน Schema1, C2 = คลาส GradStudent ใน Schema2

$\theta_{C1} = 6$ $\theta_{C2} = 9$ $\alpha = 7$ $\beta = 9$

$R_{all\text{-}attributes}$ =[1, 1, 0.1, 1, 1, 1, 1]

$Ms_{all\text{-}attributes} = 6.1$

∴ จากตาราง 3.4 จะได้ว่าจากการพิจารณาแอตทริบิวท์ทั้งหมดของคลาส Student และคลาส GradStudent คลาสทั้งสองมีระดับความเหมือนคือเป็นส่วนภายใน (Inclusion) ดังนั้น คลาส Student มีความสัมพันธ์ในลักษณะซูเปอร์คลาสกับคลาส GradStudent



รูปที่ 3.5 ตัวอย่างการแสดงความสัมพันธ์จากการเปรียบเทียบแต่ละคู่แอตทริบิวท์จากคลาส Student และคลาส GradStudent ในรูปที่ 3.3

3.2.3 การเปรียบเทียบคลาส

วิธีการเปรียบเทียบคลาส C1 และ C2 เพื่อหาความสัมพันธ์ของคลาสทั้งสองมีขั้นตอนดังนี้คือ

1) การหาความสัมพันธ์จากการเปรียบเทียบความหมายของคลาส (จากวิธีในหัวข้อ 3.2.1)
2) การหาความสัมพันธ์จากการเปรียบเทียบชุดแอตทริบิวท์ของคลาส (จากวิธีในหัวข้อ 3.2.2.3)
3) การพิจารณาความสัมพันธ์จาก 1) และ 2) จะได้เป็นความสัมพันธ์ของคลาสดังแสดงในตารางที่ 3.5

รูปที่ 3.6 จะเป็นการสรุปขั้นตอนการเปรียบเทียบคลาส Student และคลาส GradStudent จากรูปที่ 3.3 ตามขั้นตอนดังกล่าวข้างต้น ซึ่งความสัมพันธ์ที่ได้จากการเปรียบเทียบความหมายของคลาสทั้งสองคือความสัมพันธ์แบบเท่ากัน (ดังตัวอย่างที่ใช้อธิบายในข้อ 3.2.1) และความสัมพันธ์ที่ได้จากการเปรียบเทียบชุดแอตทริบิวท์คือซูเปอร์คลาส (ดังตัวอย่างที่ใช้อธิบายในข้อ 3.2.2.3) จากนั้นทำการพิจารณาจากตารางที่ 3.5 จะสรุปได้ว่าคลาส Student มีความสัมพันธ์แบบซูเปอร์คลาสกับคลาส GradStudent

การหาความสัมพันธ์ของคลาสอื่นๆ ก็จะทำในลักษณะเดียวกัน และในตารางที่ 3.6 จะเป็นความสัมพันธ์ที่ได้จากการเปรียบเทียบคลาสต่างๆ ใน Schema1 และ Schema2 ในรูปที่ 3.1

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

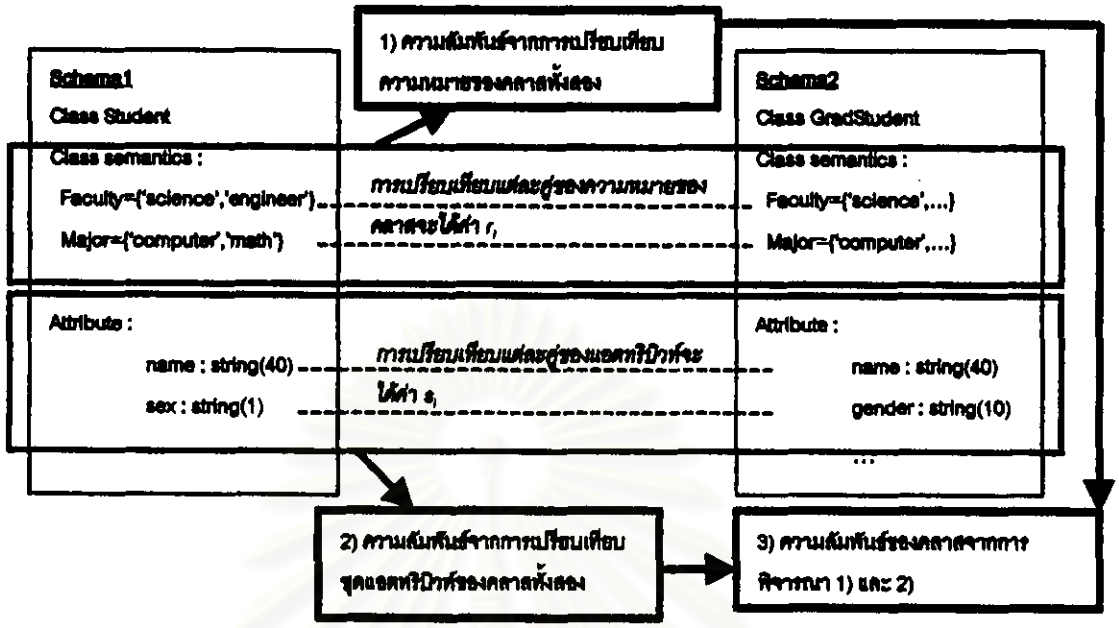
ตารางที่ 3.5 ความสัมพันธ์ของคลาส C1 และ C2 จากการพิจารณาความสัมพันธ์ของความหมายของคลาสและความสัมพันธ์ของชุดแอตทริบิวท์

ความสัมพันธ์จากการเปรียบเทียบความหมายของคลาส	ความสัมพันธ์จากการเปรียบเทียบชุดแอตทริบิวท์ของคลาส	ความสัมพันธ์ของคลาส
Equivalence	Equivalence	Equivalence
Equivalence	Superclass	Superclass
Equivalence	Subclass	Subclass
Equivalence	Sibling	Sibling
Equivalence	Disjoint	Disjoint
Superclass	Equivalence	Equivalence
Superclass	Superclass	Superclass
Superclass	Subclass	Subclass
Superclass	Sibling	Sibling
Superclass	Disjoint	Disjoint
Subclass	Equivalence	Equivalence
Subclass	Superclass	Superclass
Subclass	Subclass	Subclass
Subclass	Sibling	Sibling
Subclass	Disjoint	Disjoint
Sibling	Equivalence	Equivalence
Sibling	Superclass	Sibling
Sibling	Subclass	Sibling
Sibling	Sibling	Sibling
Sibling	Disjoint	Disjoint
Disjoint	Equivalence	Disjoint
Disjoint	Superclass	Disjoint
Disjoint	Subclass	Disjoint
Disjoint	Sibling	Sibling
Disjoint	Disjoint	Disjoint
Somenone*	Equivalence	Equivalence
Somenone	Superclass	Superclass
Somenone	Subclass	Subclass
Somenone	Sibling	Sibling
Somenone	Disjoint	Disjoint

*Either or none of the two classes have class semantics.

ตารางที่ 3.6 ความสัมพันธ์จากการเปรียบเทียบคลาสต่างๆ ในรูปที่ 3.1

Schema2 / Schema1	Person	GradStudent
Student	Subclass	Superclass
Faculty_Detail	Disjoint	Disjoint



รูปที่ 3.6 ขั้นตอนการเปรียบเทียบคลาส

ในบทนี้ได้กล่าวถึงรูปแบบของแบบจำลองข้อมูลที่จะนำมารวมและขั้นตอนวิธีการเปรียบเทียบเพื่อหาความสัมพันธ์จากแบบจำลองข้อมูล 2 ชุด โดยได้อธิบายถึงวิธีการเปรียบเทียบความหมาย ซึ่งรวมถึงความหมายของคลาสและความหมายของแอตทริบิวท์ จากนั้นได้อธิบายวิธีการเปรียบเทียบแอตทริบิวท์ ซึ่งมีการอธิบายการเปรียบเทียบแต่ละคู่ของแอตทริบิวท์ การเปรียบเทียบชุดของแอตทริบิวท์ และส่วนสุดท้ายคือวิธีการเปรียบเทียบคลาส ซึ่งจะนำความสัมพันธ์ที่ได้จากการเปรียบเทียบความหมายและการเปรียบเทียบแอตทริบิวท์มาเป็นตัวพิจารณาหาความสัมพันธ์ ซึ่งความสัมพันธ์นี้จะเป็นส่วนที่จะใช้ในการพิจารณาในขั้นตอนวิธีการรวม ซึ่งรายละเอียดของขั้นตอนวิธีการรวมจะกล่าวถึงในบทต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย