

การเข้าถึงฐานข้อมูลบนพื้นฐานของออนโทโลยีด้วยวิธีการปรับเปลี่ยนคำสั่งการสืบค้น

Ontology-Based Database Access Using Query Reformulation

สมมติ ลูชะวงษ์ งามนิจ อาจอินทร์

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

อำเภอเมือง จังหวัดขอนแก่น 40002

Email:sommany@windowsslive.com, ngamnij@kku.ac.th

บทคัดย่อ

เว็บแบบสื่อความหมายได้มีบทบาทสำคัญในการสนับสนุนงานวิจัยหลายๆด้าน โดยเฉพาะการบูรณาการองค์ความรู้จากฐานข้อมูลหลายแหล่งที่แตกต่างกัน องค์ประกอบที่สำคัญของเว็บแบบสื่อความหมายได้แก่ออนโทโลยีที่ใช้สำหรับเก็บคำศัพท์ที่ใช้งานร่วมกัน และการแก้ไขความขัดแย้งเชิงความหมายของโดเมนหนึ่งๆ บทความนี้ได้แสดงเทคนิคการปรับเปลี่ยนคำสั่งการสืบค้นสำหรับการเข้าถึงฐานข้อมูลโดยใช้วิธีการออนโทโลยีเข้ามาช่วย จุดประสงค์หลักของงานวิจัยคือการแก้ปัญหาการเข้าถึงฐานข้อมูลที่ใช้ภาษา SQL ที่เป็นเหมือนกับการสืบค้นด้วยคำเฉพาะ ซึ่งผู้ใช้จะต้องรู้จักชื่อตารางและคอลัมน์ของฐานข้อมูลนั้นๆ และเพื่อที่จะให้ได้ผลลัพธ์ของการสืบค้นมีความสอดคล้อง และตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น บทความนี้ได้นำเสนอโดเมนออนโทโลยีและกฎการปรับเปลี่ยนคำสั่งที่สามารถใช้ในการแปลงคำสั่งการสืบค้นที่มีความหมายซึ่งแสดงอยู่ในรูปแบบของภาษา SPARQLให้เป็นภาษาสืบค้นสำหรับฐานข้อมูลได้แก่ภาษา SQL โดยทำให้ผลลัพธ์ที่สืบค้นได้นี้จะตรงกับความต้องการของผู้ใช้อย่างแท้จริง

คำสำคัญ: การปรับเปลี่ยนคำสั่งการสืบค้น, ออนโทโลยี, เว็บแบบสื่อความหมาย, ฐานข้อมูล

Abstract

The Semantic Web plays an important role in several research areas, especially; the semantic knowledge integration from heterogeneous databases. A core component of Semantic Web is an ontology that is used for providing a shared vocabulary and solving the semantic conflicts of a domain. This research presents a query reformulation technique for database access using ontology-based approach. The main objective of this research is to solve the database access that uses SQL as a keyword-based query, such that, users need to know the physical table names and column names of a database. In order to provide more meaningful result satisfying the user's intention, this research proposes a domain ontology and set of rewriting rules that can be used to formulate a semantic query expressed in SPARQL into relational database queries expressed in SQL. As a consequence, the query results given in a semantically meaningful way can thus satisfy the user's intention.

Keywords: Query Reformulation, Ontology, Semantic Web, Database

1. บทนำ

ในปัจจุบันองค์กร และหน่วยงานต่างๆ ได้ให้ความสำคัญกับสารสนเทศที่มีอยู่ในองค์กร โดยหลายหน่วยงานพยายามสร้างฐานความรู้ของตนเองขึ้นมาเป็นแหล่งรวบรวมข้อมูล โดยมีจุดประสงค์เพื่อใช้ในการบูรณาการและการสืบค้นข้อมูลสารสนเทศ และเป็นฐานข้อมูลเชิง

สัมพันธ์เป็นส่วนมาก แต่เนื่องจากการสืบค้นฐานข้อมูลโดยใช้คำสำคัญ (Keyword-Based) ผู้ใช้งานต้องรู้คำศัพท์ที่ใช้ในฐานข้อมูลอย่างละเอียด เพื่อระบุถึงชื่อตารางหรือคอลัมน์ให้ตรงเพื่อสืบค้นไม่สามารถสืบค้นข้อมูลด้วยวิธีการสื่อความหมายของคำที่ใช้เช่น ข้อมูลเหมือนกันแต่ใช้คำศัพท์ที่ต่างกัน หรือใช้คำศัพท์ที่เหมือนกัน แต่ข้อมูลต่างกัน จึงทำให้ผลของการสืบค้นไม่ตรงต่อความต้องการและไม่ครอบคลุมเพียงพอต่อผู้ใช้งาน

จากเหตุผลดังกล่าว ผู้วิจัยจึงมีแนวความคิดที่จะนำเทคโนโลยีของเว็บแบบสื่อความหมายมาประยุกต์ใช้ในงานวิจัย การเข้าถึงฐานข้อมูลบนพื้นฐานของออนโทโลยีด้วยวิธีการปรับเปลี่ยนคำสั่งการสืบค้น เพื่อมุ่งเน้นการเข้าถึงฐานข้อมูลแบบสื่อความหมาย โดยใช้วิธีการแปลงจากฐานข้อมูลเชิงสัมพันธ์ไปเป็นออนโทโลยี (Ontology) [10] ที่อธิบายด้วยภาษา Ontology Web Language (OWL) [12] ซึ่งเป็นภาษาสำหรับใช้บรรยายข้อมูลเชิงความหมายและมีการสนับสนุนการแสดงผลข้อมูลได้หลายรูปแบบ [1] พร้อมทั้งสามารถทำการเชื่อมโยงอธิบายความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ได้ ซึ่งง่ายต่อการสืบค้นอ้างอิง

ดังนั้นเทคนิคการปรับเปลี่ยนคำสั่งการสืบค้นบนพื้นฐานของออนโทโลยี ถูกนำมาใช้ในการสืบค้นฐานข้อมูล ในลักษณะคำถามเชิงความหมาย ด้วยการสืบค้นข้อมูลโดยใช้ภาษา SPARQL จากนั้นระบบจะทำการปรับเปลี่ยนคำสั่งการสืบค้น (Query Reformulation) เป็น SQL เพื่อปรับคำสั่งสืบค้นให้เหมาะสมกับโครงสร้างการจัดเก็บข้อมูลในฐานข้อมูลแต่ละแหล่ง เพียงเท่านี้ระบบที่ใช้ออนโทโลยีเข้ามาช่วยในการสืบค้นฐานข้อมูลเชิงสัมพันธ์ ก็จะเป็นระบบ “intelligent answer” [3] ได้ ข้อมูลที่สื่อความหมายมากขึ้นกว่าการสืบค้นจากฐานข้อมูลธรรมดา

2. ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

2.1 ความรู้เบื้องต้นเกี่ยวกับออนโทโลยี

ออนโทโลยี เป็นการอธิบายถึงสิ่งต่าง ๆ ที่เราสนใจในโดเมนหนึ่ง ๆ อย่างชัดเจนไม่มีความกำกวม เพื่อทำให้

มนุษย์ และคอมพิวเตอร์หรือซอฟต์แวร์เอเจนต์ สามารถเข้าใจความหมายของสิ่งนั้นได้ร่วมกัน และยังช่วยขจัดปัญหาความขัดแย้งกันในเชิงความหมายของข้อมูล ทำให้การสืบค้นข้อมูลโดยคอมพิวเตอร์สามารถสืบค้นคำที่มีความหมายเหมือนกันออกมาได้ แม้จะมีการใช้คำที่ต่างกัน ในปัจจุบันออนโทโลยีได้ถูกนำไปใช้งานวิจัยหลาย ๆ ด้าน เช่น ด้านปัญญาประดิษฐ์ (AI), วิศวกรรมความรู้ (Knowledge Engineering) และประยุกต์ใช้ในการจัดการภาษาธรรมชาติ (Natural Language Processing: NLP)

2.2 ภาษา Ontology Web Language (OWL)

ภาษา OWL เป็นภาษาใช้แทนออนโทโลยี เพื่อให้คอมพิวเตอร์หรือซอฟต์แวร์เอเจนต์สามารถเข้าใจความหมายข้อมูลร่วมกัน และพัฒนาขึ้นบนพื้นฐานของภาษา RDF (Resource Description Framework) [11] ในภาษา OWL จึงมีการบรรยายข้อมูลแบบผสมผสานกันระหว่างการใช่ RDF, RDFS, XML Syntax ซึ่งแบ่งตามประเภทของการใช้งานและรายละเอียดของข้อมูลต่างๆ ไฟล์นามสกุลที่ใช้ในการสร้างเอกสารเป็นไฟล์นามสกุล .rdf หรือ .owl เป็นต้น

2.3 ภาษา SPARQL

SPARQL เป็นภาษาสำหรับการสืบค้นข้อมูลจากออนโทโลยีอย่างสื่อความหมาย ซึ่งจะมีการเข้าถึงข้อมูลโดยอาศัยโครงสร้าง Triple (Subject, Predicate, Object) แนวความคิดของหลักการสืบค้นด้วยภาษา SPARQL ที่สามารถเปรียบเทียบกันได้กับคำสั่งภาษา SQL โดยแบ่งเป็นสองส่วนหลักๆ คือ ส่วน SELECT และ WHERE ซึ่ง SELECT จะไปอธิบายตัวแปรที่ใช้ prefix “ ? ” ซึ่งเป็นตัวแปรที่เก็บค่าผลลัพธ์ และ WHERE จะเป็นเงื่อนไขการสืบค้นข้อมูล

นิยาม: การสืบค้นออนโทโลยีด้วยภาษา SPARQL

Sparql → **SELECT** varlist **WHERE** (b_gp)

โดยที่ varlist = (v₁, v₂, ..., v_n) varlist ⊆ var(b_gp)

b_gp คือ Basic Graph Pattern

2.4 การแปลง SPARQL เป็น SQL

เนื่องจากออนโทโลยีที่ Represent ด้วยภาษา OWL DL

ที่จัดเก็บแบบ Semantic Repository นั้นสามารถสืบค้นด้วยภาษา SPARQL บนพื้นฐานของ Basic Graph Pattern (BGP) หรือรู้จักในรูปแบบ RDF Triple ที่ประกอบด้วย Subject, Predicate และ Object แต่ในขณะเดียวกันในฐานข้อมูลเชิงสัมพันธ์ การเข้าถึงโครงสร้างของข้อมูลในระดับตรรกะ (Logical data model) ผ่าน SQL อยู่บนพื้นฐานของ Relational Algebra (RA) [8] คือพื้นฐานกลุ่มของ Operators ซึ่งใช้ในการประมวลผลตาราง ดังนั้นจึงนำเอาหลักการดังกล่าวมาประยุกต์ใช้เป็นแบบจำลองของการแปลงจาก SPARQL Operators ไปเป็น SPARQL Algebra ที่มีหลักการการทำงานคล้ายกับ Relational Algebra เพื่อปรับเปลี่ยนคำสั่งเป็น SQL ที่สามารถสืบค้นข้อมูลในฐานข้อมูล

Translation of SPARQL Operators into Relational Algebra
1. SELECT ?X1...?Xn WHERE {P}:
Mapped to a projection operation $\pi_{?X1...?Xn}([P]_D)$
2. P1 AND P2 (also denoted as "P1.P2"):
Mapped to and algebraic join operation $[P1]_D \bowtie [P2]_D$
3. P1 OPTIONAL P2:
Mapped to a left outer join $[P1]_D \ltimes [P2]_D$
4. P FILTER R:
Mapped to and algebraic selection $\sigma_R([P]_D)$
5. P1 UNION P2:
Mapped to a set-theoretic union $[P1]_D \cup [P2]_D$

ภาพที่ 1 แบบจำลองของการแปลงจาก SPARQL Operators ไปเป็น Relational Algebra

จากภาพที่ 1 แสดงแบบจำลองการปรับเปลี่ยนคำสั่งสืบค้น จากภาษา SPARQL เป็น SQL [4,7] ซึ่งเป็นการแปลง SPARQL Operator ได้แก่ Triple pattern, AND, OPTIONAL, FILTER และ UNION ไปเป็น Relational Algebra อาทิเช่น π คือ Projection, \bowtie คือ Join, \ltimes คือ Left Outer Join, σ คือ Select, \cup คือ Union ตามลำดับ

2.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับบทความนี้ พบว่ามีงานวิจัยที่เกี่ยวข้องกับการนำเอาออนโทโลยีมาช่วยในการสืบค้นฐานข้อมูลเชิงสัมพันธ์คือ [3] ซึ่งนำเอาความสัมพันธ์ภายในออนโทโลยี เช่น "ISA" "partOf" และ "SynOf" เป็น

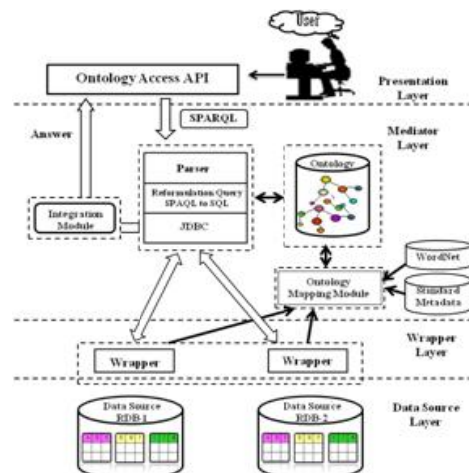
ตัวกำหนดเงื่อนไขการเข้าถึงฐานข้อมูล และใช้ Set ของ Logic Operator \vee และ \wedge เป็นกฎการปรับเปลี่ยนคำสั่งสืบค้นเป็นรูปแบบใหม่ ในการสืบค้นฐานข้อมูล ส่วนงานวิจัยที่คล้ายคลึงกับบทความนี้ได้แก่ [7] ซึ่งบทความดังกล่าวได้ศึกษาวิธีการเขียนคำสั่งสืบค้นจาก SPARQL เป็น SQL โดยใช้วิธีการแปลงโครงสร้างของฐานข้อมูลเชิงสัมพันธ์สองแหล่งไปเป็นออนโทโลยี แล้วอธิบายแหล่งจัดเก็บบนพื้นฐาน SPARQL คิวรี เพื่อนำมาปรับใช้เป็นหลักการเขียนคำสั่งการสืบค้นใหม่

จากงานวิจัยที่เกี่ยวข้องพบว่ายังไม่มียานวิจัยใดที่นำเอาหลักการแปลงจาก SPARQL Query เป็น SPARQL Algebra ที่มีหลักการการทำงานคล้ายกับ Relational Algebra เพื่อปรับเปลี่ยนเป็นคำสั่ง SQL ในการสืบค้นฐานข้อมูลเชิงสัมพันธ์สองแหล่ง เหมือนกับบทความนี้

3. การดำเนินงานวิจัย

3.1 การออกแบบสถาปัตยกรรมของระบบ

สถาปัตยกรรมการเข้าถึงฐานข้อมูลบนพื้นฐานของออนโทโลยีด้วยวิธีการปรับเปลี่ยนคำสั่งการสืบค้นออกแบบมาเพื่อแก้ไขปัญหาการสืบค้นฐานข้อมูลด้วยคำสั่งสำคัญ ที่ต้องระบุคำศัพท์ที่จัดเก็บในตารางเท่านั้น การทำงานของระบบประกอบด้วย 4 ชั้น ได้แก่ Presentation Layer, Mediator Layer, Wrapper Layer และ Data Source Layer ซึ่งได้แสดงภาพรวมของสถาปัตยกรรม ดังภาพที่ 2



ภาพที่ 2 ภาพรวมสถาปัตยกรรมการทำงานของระบบการปรับเปลี่ยนคำสั่งการสืบค้น

Presentation Layer: ประกอบด้วย Ontology Access API ซึ่งทำหน้าที่อำนวยความสะดวกให้กับผู้ใช้งานที่ต้องการสืบค้นข้อมูลในระบบ โดยมีหน้าที่รับคำสั่งการสืบค้นโดยภาษา SPARQL จากผู้ใช้งานแล้วส่งไปที่ชั้น Mediator Layer เพื่อทำการประมวลผล และแสดงผลลัพธ์ของการสืบค้น

Mediator Layer: ประกอบด้วย Query Processor, Integration Module, Ontology, Ontology Mapping Module และ Standard Metadata ซึ่งมีรายละเอียดดังนี้

Query Processor นับได้ว่าเป็นส่วนสำคัญของระบบ ได้แก่ Parser, Reformulation Query และ JDBC โดยที่ Parser เป็นส่วนที่รับข้อมูลจาก Ontology Access API แล้วนำไปตรวจสอบว่าถูกต้องตามรูปแบบของ SPARQL Syntax หรือไม่ ก่อนที่จะส่งคำสั่งไปสืบค้นในออนโทโลยีกลาง (Ontology) ด้วยการหาเหตุผล หรืออนุมานหาผลลัพธ์ เพื่อนำไปใช้สร้างรูปแบบการปรับเปลี่ยนคำสั่งการสืบค้น (Reformulation Query) เป็นคำสั่งสืบค้นเหมือนกับ SQL Syntax เพื่อคิวรีข้อมูลจากฐานข้อมูลตามแต่ละแหล่งผ่านส่วนติดต่อกับฐานข้อมูล (JDBC) ในขั้นต่อไป

Integration Module เป็นส่วนที่ทำหน้าที่บูรณาการผลลัพธ์ก่อนที่จะส่งไปแสดงผลต่อผู้ใช้งาน

Ontology หรือ ออนโทโลยีกลางเป็นส่วนที่เก็บข้อมูลการเชื่อมโยงเชิงความหมายของออนโทโลยีทั้งสองแหล่งไว้ด้วยกัน โดยใช้การบรรยายด้วยภาษา OWL DL ที่เปรียบเสมือนกับฐานองค์ความรู้ที่บรรยายออนโทโลยีตามขอบเขตที่เราสนใจ ซึ่งทำหน้าที่จัดเก็บคำศัพท์ที่ใช้งานร่วมกัน

Ontology Mapping Module เป็นส่วนที่ทำหน้าที่ในการเชื่อมโยงออนโทโลยีที่ได้จากการแปลงฐานข้อมูลเชิงสัมพันธ์ จากชั้น Data Layer เข้าไว้ด้วยกันโดยได้มีการนำเอาฐานข้อมูล WordNet มาช่วยในการค้นหาความคล้ายคลึงเชิงความหมายระหว่างสองคอนเซปต์ เพื่อนำมาเพิ่มเติมคุณสมบัติของ Class และ Property อาทิ เช่น

equivalentClass, equivalentProperty, Functional Property, Transitive Property และ Symmetric Property เป็นต้น

Standard Metadata เป็นคำอธิบายข้อมูลที่ออกแบบตามมาตรฐาน Dublin Core (DC) [9] และ Learning Object Metadata (LOM) [6] เพื่อเป็นมาตรฐานกลางของเมตาดาต้าที่ใช้อธิบายข้อมูล จากแหล่งข้อมูลการเรียนรู้

Wrapper Layer: ทำหน้าที่เป็นส่วนกลางในการแปลงโครงสร้างฐานข้อมูลให้อยู่ในรูปแบบเอกสาร OWL และยังเป็นส่วนที่ส่งคำสั่งคิวรีไปยังฐานข้อมูลแต่ละแหล่ง

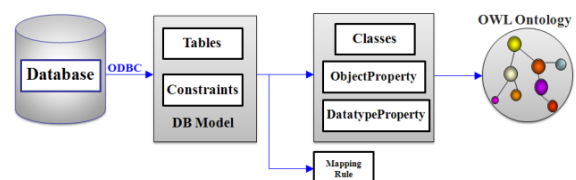
Data Source Layer: เป็นส่วนจัดเก็บข้อมูลในรูปแบบของฐานข้อมูลเชิงสัมพันธ์จากแหล่งข้อมูลต่าง ๆ

3.2 ขั้นตอนการดำเนินงานวิจัย

การดำเนินงานวิจัยประกอบด้วย 3 ขั้นตอนดังนี้

3.2.1 ขั้นตอนการแปลงจากฐานข้อมูลเชิงสัมพันธ์ไปเป็นออนโทโลยี

การแปลงฐานข้อมูลเชิงสัมพันธ์เป็นออนโทโลยีใช้หลักการพื้นฐานที่เรียกว่า Mapping Rule [5] คือ การแปลงตาราง คอลัมน์ ไปเป็นองค์ประกอบที่สอดคล้องของออนโทโลยี เช่น คลาส พร็อพเพอร์ตี้ ในบทความนี้ใช้โปรแกรมโปรเทจ 3.3 ซึ่งใช้ ODBC Driver เป็น Bridge ในการเชื่อมต่อฐานข้อมูลเข้าไปยังโปรแกรม จากนั้นสามารถเลือกแปลงองค์ประกอบที่สอดคล้องของฐานข้อมูลไปเป็นองค์ประกอบที่สอดคล้องของออนโทโลยีตามกฎการแปลงที่โปรแกรมกำหนดไว้ และสามารถนำเข้าข้อมูลทั้งหมดในฐานข้อมูลหรือเลือกแปลงเฉพาะโครงสร้าง ซึ่งงานวิจัยนี้จะทำการแปลงเฉพาะโครงสร้างไม่ได้จัดเก็บอินสแตนซ์หรือข้อมูลไว้ในออนโทโลยีดังภาพที่ 3



ภาพที่ 3 ภาพรวมการแปลงฐานข้อมูลเชิงสัมพันธ์ไปเป็นออนโทโลยี

จากภาพที่ 3 แสดงขั้นตอนการเชื่อมโยงฐานข้อมูลเข้าสู่โปรแกรมโปรเทจโดยเลือกแปลง ตาราง คอลัมน์ภาพที่ 4 ไปเป็นองค์ประกอบที่สอดคล้องของออนโทโลยี เช่น คลาส และพร็อพเพอร์ตี้ภาพที่ 5 ดังต่อไปนี้

CrsCode	CrsName	Credit	Course(CrsCode,CrsName,Credit)
C5322111	Introduction to Computer	3	CrsCode: Course Code
C5322736	Semantic Web Technology	3	CrsName: Course Name
C5322789	XML Technology and Application	3	Credit: Course Credit
			P_KEY (CrsCode)

ภาพที่ 4 ตัวอย่างตาราง Course ที่จัดเก็บในฐานข้อมูล

```
<owl:Class rdf:about="&db;Course"/>
  <owl:DatatypeProperty rdf:about="&db;Credit">
    <rdf:type rdf:resource="&owl;FunctionalProperty"/>
    <rdfs:domain rdf:resource="&db;Course"/>
    <rdfs:range rdf:resource="&xsd:int"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="&db;CrsCode">
    <rdf:type rdf:resource="&owl;FunctionalProperty"/>
    <rdfs:domain rdf:resource="&db;Course"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="&db;CrsName">
    <rdf:type rdf:resource="&owl;FunctionalProperty"/>
    <rdfs:domain rdf:resource="&db;Course"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>
```

ภาพที่ 5 โค้ดภาษา OWL ที่ได้จากการแปลงตาราง Course เป็นคลาส Course และแปลงจากคอลัมน์เป็นพร็อพเพอร์ตี้

3.2.2 ขั้นตอนการเชื่อมโยงความสัมพันธ์ในเชิงความหมาย

การเชื่อมโยงความสัมพันธ์ในเชิงความหมาย (Semantic Relation) คือการสร้างความสัมพันธ์เก็บในออนโทโลยีที่ได้มาจากขั้นตอนการแปลงฐานข้อมูลสองแหล่ง เพื่อแก้ปัญหาความขัดแย้งกันในเชิงความหมายของข้อมูล การเชื่อมโยงข้อมูลเป็นการเชื่อมโยงในรูปแบบที่เรียกว่าออนโทโลยี Mapping เป็นกระบวนการในการเปรียบเทียบความเหมือนระหว่างคอนเซปต์ หรือความสัมพันธ์ของออนโทโลยีที่แตกต่างกัน โดยการนำเอาคอนเซปต์หรือความสัมพันธ์ของออนโทโลยีหนึ่งมาเทียบเข้ากับอีกออนโทโลยีหนึ่งที่ต้องการโดย WordNet Similarity [13] ซึ่งใช้สมการดังนี้

$$\text{Sim}_{\text{wup}} = \frac{2 \times \text{depth}(\text{LCS})}{\text{depth}(\text{concept1}) + \text{depth}(\text{concept2})} \quad (1)$$

จากสมการที่ (1) เป็นการหาค่าความคล้ายคลึงแบบเมตริก โดยหาจากความลึกของ 2 คอนเซปต์ ในหมวดหมู่ที่ถูกจัดไว้ใน WordNet และความลึกที่น้อยที่สุดของหมวดหมู่เหล่านั้นที่เรียกว่า: The Least Common Subsume (LCS) ซึ่งนำมารวมเป็นค่าวัดความคล้ายคลึงของหมวดหมู่ระหว่าง 2 คำศัพท์ โดยมีความคล้ายคลึงอยู่ระหว่าง 0 ถึง 1 มาค้นหาความสัมพันธ์ในเชิงความหมาย และใช้เครื่องมือ [2] เชื่อมโยงความสัมพันธ์ออนโทโลยีทั้งสองแหล่ง นอกจากนั้นยังได้นำเอามาตรฐานคำอธิบายข้อมูลบนพื้นฐานของ DC อาทิเช่น dc: title, dcq:Name เป็นมาตรฐานกลาง เพิ่มความสามารถในการค้นหาข้อมูลสำหรับสื่อดิจิทัลได้ดียิ่งขึ้น

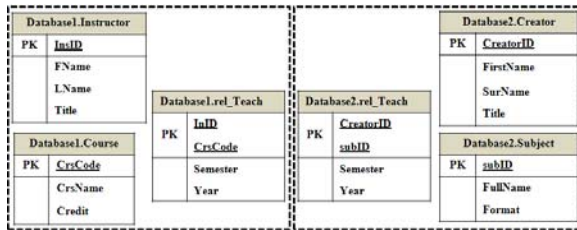
3.2.3 ขั้นตอนการทดสอบสืบค้นข้อมูล

ออนโทโลยีที่บรรยายด้วยภาษา OWL DL ที่จัดเก็บแบบ Semantic Repository นั้นสามารถสืบค้นด้วยภาษา SPARQL บนพื้นฐานของ Basic Graph Pattern (BGP) และใช้งานร่วมกับโปรแกรมประยุกต์ Reasoner ซึ่งมีความยืดหยุ่นสูง สามารถช่วยให้ผู้ใช้งานสืบค้นข้อมูลตรงตามความต้องการ เพราะนำเอาออนโทโลยีมาเป็นตัวกลางในการเข้าถึงฐานข้อมูลแบบสื่อความหมาย ถึงแม้ว่าคำค้นที่ใช้ อาจไม่ตรงตามชื่อ ตาราง และ คอลัมน์ที่จัดเก็บในฐานข้อมูลก็ตาม

4. การทดลองและวิเคราะห์ผล

4.1 การศึกษาโครงสร้างของฐานข้อมูลทั้งสองแหล่ง

บทความนี้จะถูกพัฒนาโดยใช้กรณีศึกษาที่อยู่ในโดเมน E-learning โดยมีแหล่งจัดเก็บข้อมูลสองแหล่งได้แก่ ฐานข้อมูลระบบลงทะเบียน และฐานข้อมูลระบบ E-learning โดยทั้งสองระบบงานเป็นฐานข้อมูลเชิงสัมพันธ์ที่จัดเก็บใน SQL Server 2000 และมีการใช้ชื่อฟิลด์ที่แตกต่างกันออกไป แต่มีการเก็บข้อมูลที่เหมือนกันซึ่งมีรายละเอียดของโครงสร้างในภาพที่ 6 ดังนี้



ภาพที่ 6 ฐานข้อมูลสองแหล่งที่มีความขัดแย้งเชิงความหมาย

จากฐานข้อมูลเชิงสัมพันธ์ทั้งสองแหล่งในภาพที่ 6 ที่มีความขัดแย้งเชิงความหมาย (Name Conflict) โดยแหล่งที่หนึ่งประกอบด้วย ตาราง Instructor ตาราง Course และตาราง Teach แหล่งที่สองประกอบด้วย ตาราง Creator ตาราง Subject และตาราง Teach ซึ่งตารางเหล่านี้มีการจัดเก็บข้อมูล ผู้สอน รายวิชา และรายละเอียดของการสอนที่เหมือนกัน แต่มีการตั้งชื่อที่แตกต่างกัน จึงได้ทำการเลือกคู่คอนเซปต์ เพื่อคำนวณค่าความคล้ายคลึงจากสมการที่ 1 เมื่อ [LCS: (Dept= 11), Dept(Instructor)=12, Dept(Lecturer)=12] ดังตารางที่ 1

ตารางที่ 1 คำนวณค่าความคล้ายคลึงของสองคอนเซปต์ที่ถูกเลือก

Registration	E-learning	Similarity
Course	Subject	0.6667
Instructor	Lecturer	0.9167
Teach	Teach	1

นอกจากนั้นยังได้ทำการเชื่อมโยงโครงสร้างข้อมูลที่มีความหลากหลายนั้น เข้าสู่มาตรฐานเมตาดาต้ากลาง ดังตัวอย่างแสดงในภาพที่ 7

```
<owl:DatatypeProperty rdf:about="&dc:title">
  <owl:equivalentProperty rdf:resource="&eLearn;FullName"/>
  <owl:equivalentProperty rdf:resource="&re;CrsName"/>
</owl:DatatypeProperty>
```

ภาพที่ 7 โค้ดเบื้องต้นการเชื่อมโยงพร็อพเพอร์ตี้กับพร็อพเพอร์ตี้

4.2 การปรับเปลี่ยนคำสั่งการสืบค้นจาก SPARQL เป็น SQL

การปรับเปลี่ยนคำสั่งสืบค้นจาก SPARQL เป็น SQL จะนำเสนอในกรณีของการสืบค้นด้วย Basic Graph Pattern

ซึ่งอยู่บนพื้นฐานของ Triple (Subject, Predicate, Object)

SELECT ?X1...?Xn WHERE {P};
Mapped to a projection operation $\pi_{?X1...?Xn}([P]_D)$

ซึ่งการสืบค้นด้วย Basic Graph Pattern ใน SPARQL นี้จะถูกแปลงเป็นการ Project เฉพาะบาง Attribute ที่กำหนดไว้ในผลลัพธ์หลังคำสั่ง SELECT และ WHERE ที่เป็นเงื่อนไขการสืบค้นข้อมูล เช่นการสืบค้นด้วยภาษา SPARQL ต่อไปนี้ เป็นการสืบค้นรหัส กับรายชื่อนวิชา โดยระบุความสัมพันธ์เป็น dc:title อ้างอิงไปยังชื่อวิชาที่เปิดสอน โปรแกรมประยุกต์ Reasoner จะทำการอนุมานหาผลลัพธ์ เพื่อนำไปใช้สร้างรูปแบบการปรับเปลี่ยนคำสั่งการสืบค้นเหมือนกับ SQL เพื่อดึงข้อมูลจากฐานข้อมูลทั้งสองแหล่ง ตามกฎ และนิยามของคุณสมบัติที่สร้างไว้ในออนโทโลยี ดังต่อไปนี้

SPARQL Query:

PREFIX dc: <http://purl.org/dc/elements/1.1/dc/#>

SELECT ?ID ?Name

WHERE {?ID dc:title ?Name }

จากคำสั่ง SPARQL สามารถเขียนเป็น Relational Algebra ได้ดังนี้

$\pi_{?ID} \leftarrow ?Subject (R)$

$?Name \leftarrow ?Object$

Projection Operator แปลงเป็นการ SELECT คอลัมน์ในภาษา SQL

SELECT R.Subject AS ID, R.Object AS Name
FROM R

เมื่อต้องการสืบค้นข้อมูลจากฐานข้อมูลสามารถปรับเปลี่ยนคำสั่งสืบค้นเป็นภาษา SQL ให้เหมาะสมตามแหล่งจัดเก็บข้อมูลได้ดังต่อไปนี้

SQL Access Database 1

SELECT CrsCode AS [ID], CrsName AS [Name]
FROM Database1.Course

SQL Access Database 2

SELECT subID AS [ID], FullName AS [Name]
FROM Database2.Subject

จากคำสั่ง SQL ที่ใช้สืบค้นฐานข้อมูลทั้งสองแหล่ง ผลลัพธ์เบื้องต้นจะถูกเก็บไว้ใน Integration Module ซึ่งเป็นส่วนที่ใช้ในการบูรณาการคำตอบก่อนส่งออกไปแสดงผลลัพธ์ต่อผู้ใช้งานด้วยคำสั่ง SQL ดังนี้

```
SELECT ID, Name
FROM
  ((SELECT CrsCode AS [ID], CrsName AS [Name]
    FROM Database1.Course
  UNION
    SELECT subID AS [ID], FullName AS [Name]
    FROM Database2.Subject)) AS View
```

4.3 การวัดประสิทธิภาพของการสืบค้น

วิธีการประเมินประสิทธิภาพของการสืบค้น จะใช้ค่าความถูกต้อง (Precision) และค่าความครบถ้วน (Recall) จากการทดลองคิวรีจำนวน 20 ครั้ง โดยใช้ข้อมูลชุดเดียวกัน แต่คำถามการสืบค้นแตกต่างกันไปตามแต่ละกฎที่ใช้ในการแปลงจาก SPARQL เป็น SQL แล้วนำผลการคิวรีข้อมูลมาคำนวณเพื่อวัดประสิทธิภาพด้วยค่าเอฟเมเชอร์ (F-measure) เพื่อทดสอบความถูกต้องดังสมการต่อไปนี้

$$\text{Precision} = \frac{A}{A+C} 100\% \quad (2)$$

$$\text{Recall} = \frac{A}{A+B} 100\% \quad (3)$$

$$F - \text{measure} = \frac{2PR}{P + R} \quad (4)$$

เมื่อ A คือจำนวนข้อมูลได้จากการสืบค้นถูกต้องที่ถูกดึงขึ้นมา B คือจำนวนข้อมูลที่ถูกดึงแต่ไม่ถูกต้องขึ้นมา และ C คือจำนวนข้อมูลที่ได้จากการสืบค้นที่ไม่ถูกต้อง แต่ถูกดึงขึ้นมา จากการทดสอบได้ค่าความถูกต้องในการสืบค้นข้อมูล และตรงตามความต้องการ (Precision) เท่ากับ 95%, ระดับความครบถ้วน (Recall) เท่ากับ 93% และเอฟเมเชอร์ (F-measure) เท่ากับ 93.99%

5. สรุปผลการดำเนินงาน และข้อเสนอแนะ

บทความนี้ศึกษาวิธีการปรับเปลี่ยนคำสั่งสืบค้นจาก SPARQL เป็น SQL ซึ่งช่วยในการสืบค้นฐานข้อมูลเหมาะสมตามโครงสร้างการจัดเก็บข้อมูลในฐานข้อมูลแต่

ละแหล่ง เพื่อแก้ปัญหการสืบค้นฐานข้อมูลด้วยคำสั่งสำคัญ (Keyword-Based) จากการใช้ความสามารถของภาษา OWL มาบรรยายข้อมูลเชิงความหมาย ด้วยวิธีการแปลงจากฐานข้อมูลเชิงสัมพันธ์ไปเป็นออนโทโลยี ในด้านการเก็บรายละเอียดต่างๆ เช่น บทบาท (Role) เงื่อนไข (Restriction) และความสัมพันธ์ ทำให้ฐานข้อมูลทั้ง 2 แหล่งที่แตกต่างกันสามารถทำงานร่วมกันได้ และการใช้เทคนิคการปรับเปลี่ยนคำสั่ง SPARQL Algebra เป็นการสืบค้นข้อมูลบนโครงสร้างข้อมูลของ OWL DL ซึ่งคล้ายกับการดำเนินการของภาษาสอบถามข้อมูล (SQL) ของฐานข้อมูลเชิงสัมพันธ์ และได้ทำการทดลองสืบค้นความสอดคล้องระหว่างโครงสร้างของออนโทโลยี และข้อมูลภายในฐานข้อมูล ได้ค่า F-measure หรือค่าน้ำหนักของการหาค่าเฉลี่ยระหว่าง Precision และ Recall เท่ากับ 93.99 % ซึ่งหมายถึงผลจากการดึงเอกสารที่มีความครบถ้วนและตรงประเด็น จากการนำเอาออนโทโลยีมาช่วยในการเข้าถึงฐานข้อมูลแบบสื่อความหมายมีค่าความเชื่อมั่นเท่ากับร้อยละ 93.99% ดังนั้นการที่นำเอาออนโทโลยีมาช่วยในการสืบค้นข้อมูลนั้นมีประสิทธิภาพค่อนข้างดี ทำให้ได้ผลลัพธ์ตามวัตถุประสงค์ที่กำหนดไว้ และสามารถแก้ปัญหการสืบค้นฐานข้อมูลด้วยคำสั่งสำคัญ ผลลัพธ์จากการสืบค้นได้ข้อมูลข่าวสารที่ถูกต้อง และตรงต่อความต้องการของผู้ใช้งานอย่างแท้จริง

ระบบการปรับเปลี่ยนคำสั่งการสืบค้นนี้ จากกรณีศึกษาใช้ฐานข้อมูลเชิงสัมพันธ์เพียง 2 แหล่งข้อมูลเท่านั้น โดยสามารถนำไปเพิ่มแหล่งข้อมูลเช่น XML, RDF ได้ และพัฒนาเพิ่มในส่วนการสืบค้นโดยใช้ Aggregate Function เช่น count, sum ฯลฯ

เอกสารอ้างอิง

- [1] เจริญพร ศรีตระกูล. (2549). วิธีการในการบูรณาการออนโทโลยี และประเมินความถูกต้องของการบูรณาการ. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น.

- [2] Altova SemanticWorks 2009. Available from <http://www.altova.com/SemanticWorks.html>
- [3] C.B. Necib and J. Freytag. (2004). *Using Ontologies for Database Query Reformulation*. In Proc. on the 18th conference on Advances in Databases and Information Systems.
- [4] R. Cyganiak. (2005). *A relational algebra for SPARQL*. Bristol, UK.
- [5] I. Astrova, N. Korda, & A. Kalja. (2007). *Rule-Based Transformation of SQL Relational Databases to OWL Ontologies*. In Proceedings of the 2nd International Conference on Metadata & Semantics Research.
- [6] IEEE. (2002). *Draft Standard for Learning Object Metadata*. New York, USA.
- [7] J. Wang, Z. Miao, Z. Zhang, & B. Zho. (2009). Querying Heterogeneous Relational Database using SPARQL. *International Conference on Computer and Information Science* (pp. 475-480).
- [8] K.Munir, M.Odeh and R.McClatchey. (2009). Managing the Mapping between Domain Ontologies and Databases Schemas when Formulating Relational Queries. *IDEAS*. Italy: ACM.
- [9] S. Weibel, J. Kunze, C.Lagoze, & M. Wolf. (1998). *Dublin Core Metadata for Resource Discovery*. Retrieved April 17, 2010, from <http://www.faqs.org/rfcs/rfc2413.html>
- [10] S. Zhao, & E. Chang. (2007). *From Database to Semantic Web Ontology: An Overview*. In On the Move to Meaningful Internet System 2007: OTM 2007 Workshops (pp. 1205-1214). Heidelberg: Springer Berlin.
- [11] W3C. (2004). *RDF Vocabulary Description Language 1.0: RDF Schema*. Retrieved October 23, 2009, from W3C Recommendation: <http://www.w3.org/TR/rdf-schema/>
- [12] W3C. (2004). *OWL Web Ontology Language*. Retrieved September 10, 2009, from: <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>
- [13] Z. Wu, & M. Palmer. (2005). *Measuring the semantic similarity of texts*. ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, (pp. 13-18).