# การค้นคืนข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับเทคโนโลยีเว็บเชิงความหมาย

***ผุสดี บุญรอด****
*อาจารย์ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ*
***ประกายมาศ ศรีสุขทักษิณ***
*วิศวกร ฝ่ายเทคโนโลยีสารสนเทศเพื่อสนับสนุนธุรกิจ บริษัท กสท โทรคมนาคม จำกัด (มหาชน)*

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0-2555-2000 ต่อ 2733 อีเมล: pudsadee@kmutnb.ac.th
รับเมื่อ 16 ตุลาคม 2557 ตอบรับเมื่อ 12 มกราคม 2558 เผยแพร่ออนไลน์ 18 พฤษภาคม 2558
DOI: 10.14416/j.kmutnb.2015.01.002 © 2015 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

**บทคัดย่อ**

การค้นคืนข้อมูลให้ตรงตามความต้องการของผู้ใช้งานมีความจำเป็นมากในปัจจุบัน เนื่องจากข้อมูลมีปริมาณเพิ่มมากขึ้น ทำให้การประมวลผลล่าช้า และการค้นคืนไม่ตรงตามความต้องการของผู้ใช้งาน วัตถุประสงค์ของงานวิจัยนี้จึงนำฐานข้อมูลไม่สัมพันธ์มาประยุกต์ใช้กับการจัดการข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้าง และโครงสร้างออนโทโลยีที่เป็นเทคโนโลยีเว็บเชิงความหมายมาช่วยค้นคืนข้อมูลให้มีประสิทธิภาพมากขึ้น โดยทำการออกแบบโครงสร้างออนโทโลยีเพื่อจัดเก็บลงใน MongoDB ซึ่งเป็นฐานข้อมูลไม่สัมพันธ์ จากการวิจัยพบว่าผลการทดสอบความเร็วในการอ่านข้อมูลที่จำนวนเธรดเท่ากับ 100 เพื่อเข้าถึงข้อมูลจำนวน 100,000 แถว ใช้เวลาในการประมวลผลมากที่สุด ซึ่งสามารถประมวลผลเสร็จภายในเวลา 36.42 วินาที ส่วนการแก้ไขและการลบข้อมูลพบว่าจำนวนเธรดและจำนวนข้อมูลที่แตกต่างกันส่งผลต่อเวลาในการประมวลผลน้อยมาก ซึ่งสามารถประมวลผลเสร็จภายในเวลาไม่ถึง 1 วินาที

**คำสำคัญ**: การค้นคืนสารสนเทศ ภาษาสอบถามแบบไม่มีโครงสร้าง ออนโทโลยี เว็บเชิงความหมาย ฐานข้อมูลไม่สัมพันธ์

การอ้างอิงบทความ: ผุสดี บุญรอด และ ประกายมาศ ศรีสุขทักษิณ, "การค้นคืนข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับเทคโนโลยีเว็บเชิงความหมาย," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 25, ฉบับที่ 2, หน้า 255 - 264, พ.ค. - ส.ค. 2558. http://dx.doi.org/10.14416/j.kmutnb.2015.01.002

# Retrieving Big Data Using NoSQL and Semantic Web Technology

***Pudsadee Boonrawd****
*Lecturer, Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

***Pragaimas Srisuktaksin***
*Engineer, Information Technology for Business Support Division, CAT Telecom Public Company Limited, Bangkok, Thailand*

* Corresponding Author, Tel. 0-2555-2000 Ext. 2733, E-mail: pudsadee@kmutnb.ac.th
Received 16 October 2014; Accepted 12 January 2015; Published online: 18 May 2015
DOI: 10.14416/j.kmutnb.2015.01.002 © 2015 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

**Abstract**

The retrieval of information to meet user's requirement is crucial in this Big Data era. Due to the increasing volume of data, retrieval suffers from the processing delays and the inaccuracy of the retrieved data. The objective of this research is to combine a non-relational database, which features the management of large volumes of data, with a big data retrieval system based on NoSQL. Ontology structure, a semantic web technology, is employed to enhance the retrieval process by storing in MongoDB, which is a non-relational database. The result shows that the response time of the query operation using 100 threads with 100,000 rows is 36.42 seconds. For the updating and deleting operations, the difference between the thread number and the data size has a minimal effect on the processing time, which is less than 1 second.

**Keywords**: Information Retrieval, NoSQL, Ontology, Semantic Web, Non-relational Database

Please cite this article as: P. Boonrawd and P. Srisuktaksin, "Retrieving Big Data Using NoSQL and Semantic Web Technology," *J. KMUTNB*, Vol. 25, No. 2, pp. 255 - 264, May. - Aug. 2015 (in Thai). http://dx.doi.org/10.14416/j.kmutnb.2015.01.002

**1. บทนำ**

ในปัจจุบันข้อมูลมีเป็นจำนวนมาก และมีแนวโน้มเพิ่มขึ้นอย่างรวดเร็ว จากผลการสำรวจผู้ใช้งานอินเทอร์เน็ตทั่วโลกในปี ค.ศ. 2013 จากรายงาน Global Internet Report 2014 คิดเป็น 37.9% [1] และผลการสำรวจพฤติกรรมการใช้งานเครือข่ายสังคมออนไลน์ของชาวอเมริกันในงานวิจัย The Social Habit 2012 ตั้งแต่ปี ค.ศ. 2008 - 2012 พบว่ามีจำนวนผู้ใช้งานเพิ่มขึ้นจาก 24% เป็น 56% [2] ซึ่งหากการค้นคืนข้อมูลได้ผลลัพธ์ไม่ตรงตามความต้องการของผู้ใช้งาน จะทำให้ได้ข้อมูลขยะมากยิ่งขึ้น ดังนั้น การบริหารจัดการข้อมูลให้มีประสิทธิภาพจึงเป็นเรื่องสำคัญ โดยฐานข้อมูลไม่สัมพันธ์ (Non-relational Database) เริ่มมีบทบาทกับระบบสารสนเทศขนาดใหญ่เพิ่มมากขึ้น เนื่องจากรองรับข้อมูลปริมาณมาก และมีคุณสมบัติรองรับการทำงานหนักได้ดี [3]

การค้นคืนสารสนเทศ (Information Retrieval) เป็นกระบวนการคัดเลือกข้อมูลหรือสารสนเทศที่ผู้ใช้งานต้องการ และทำการคัดกรองข้อมูลที่ไม่ต้องการออกไป ซึ่งมีหลายวิธีการถูกนำมาใช้ที่แตกต่างกันตามประเภทของข้อมูล เช่น ข้อความ รูปภาพ วีดิทัศน์ หรือเสียง เป็นต้น โดยแนวคิดหลักที่ถูกนำมาประยุกต์ใช้เพื่อสามารถค้นคืนข้อมูลให้ตรงตามความต้องการของผู้ใช้งาน และมีการประมวลผลที่รวดเร็ว ตัวอย่างเช่น แบบจำลองบูลีน (Boolean Model) แบบจำลองเวกเตอร์สเปซ (Vector Space Model) หรือแบบจำลองความน่าจะเป็น (Probabilistic Model) เป็นต้น [4]

คณะผู้วิจัยจึงตระหนักถึงการบริหารจัดการข้อมูลขนาดใหญ่อย่างมีประสิทธิภาพ โดยเน้นการค้นคืนข้อมูลให้ตรงตามความต้องการของผู้ใช้งาน และประหยัดเวลาในการประมวลผล ดังนั้นวัตถุประสงค์ของการวิจัยนี้จึงได้ทำการออกแบบฐานข้อมูลไม่สัมพันธ์โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้าง และโครงสร้างออนโทโลยีที่เป็นเทคโนโลยีเว็บเชิงความหมาย (Semantic Web) เพื่อนำมาช่วยในการค้นคืนข้อมูลให้มีประสิทธิภาพยิ่งขึ้น จากนั้นทำการทดสอบความเร็วในการประมวลผลข้อมูล ซึ่งในบทความวิจัยนี้ประกอบด้วย 5 หัวข้อหลัก ได้แก่ หัวข้อที่ 1 บทนำ หัวข้อที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง หัวข้อที่ 3 วิธีการดำเนินการวิจัย หัวข้อที่ 4 ผลการดำเนินการวิจัย และหัวข้อที่ 5 สรุป

**2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง**

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องเพื่อนำมาประยุกต์ใช้กับงานวิจัย สามารถแสดงรายละเอียดต่าง ๆ ได้ดังต่อไปนี้

**2.1 ฐานข้อมูลไม่สัมพันธ์**

ฐานข้อมูลไม่สัมพันธ์มีลักษณะข้อมูลที่ไม่มีความสัมพันธ์กัน ซึ่งใช้ภาษาสอบถามแบบไม่มีโครงสร้าง (Not Only SQL: NoSQL) ในการจัดการข้อมูล โดยเน้นการเข้าถึงที่รวดเร็ว และฐานข้อมูลถูกออกแบบมาเพื่อรองรับข้อมูลปริมาณมาก แต่อาจไม่รับรองคุณสมบัติ ACID ได้แก่ ความเป็นหนึ่งเดียว (Atomicity) ความสอดคล้อง (Consistency) ความเป็นอิสระ (Isolation) และความคงทน (Durability) ของฐานข้อมูลเชิงสัมพันธ์ (Relational Database)

2.1.1 ประเภทของฐานข้อมูลไม่สัมพันธ์ สามารถแบ่งออกได้เป็น 4 ประเภท ตามลักษณะการจัดเก็บข้อมูลที่แตกต่างกัน ได้แก่

ก) ฐานข้อมูลแบบคีย์อ้างอิง (Key Value Store Databases) เป็นการจัดเก็บข้อมูลอย่างง่ายที่ใช้คีย์ (Keys) และค่าผลลัพธ์ (Values) ซึ่งมีลักษณะเฉพาะไม่ซ้ำกับข้อมูลตัวอื่น ๆ เพื่อใช้อ้างอิงและค้นหาข้อมูล

ข) ฐานข้อมูลแบบคอลัมน์ (Column Oriented Databases) เป็นการจัดเก็บข้อมูลในรูปแบบคอลัมน์ เพื่อรองรับข้อมูลขนาดใหญ่ และการเข้าถึงข้อมูลที่รวดเร็ว

ค) ฐานข้อมูลแบบเอกสาร (Document Oriented Databases) เป็นการจัดเก็บเอกสารในรูปแบบไม่มีโครงสร้างหรือกึ่งโครงสร้าง เช่น JSON หรือ XML เป็นต้น [5] โดยฐานข้อมูลแบบเอกสารมีรูปแบบการจัดเก็บคล้ายกับอาร์เรย์แบบหลายมิติ

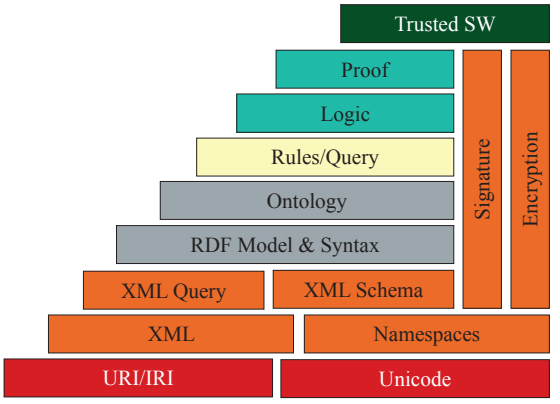
ง) ฐานข้อมูลแบบกราฟ (Graph Databases) เป็นแนวคิดของทฤษฎีกราฟที่จัดเก็บข้อมูลในลักษณะของความสัมพันธ์โดยมีโหนดเป็นตัวแทนของข้อมูล เส้น และคุณสมบัติเป็นองค์ประกอบ [6]

2.1.2 ตัวอย่างของฐานข้อมูลไม่สัมพันธ์ ซึ่งในปัจจุบันมีเครื่องมือที่สามารถนำมาประยุกต์ใช้ในการพัฒนาฐานข้อมูลไม่สัมพันธ์เพิ่มมากขึ้น โดยตัวอย่างสามารถแสดงได้ดังตารางที่ 1

**ตารางที่ 1** ตัวอย่างฐานข้อมูลไม่สัมพันธ์แต่ละประเภท

| ฐานข้อมูล | ประเภท | คุณสมบัติ |
|---|---|---|
| Berkeley DB | คีย์อ้างอิง | เป็นฐานข้อมูลแบบฝังตัว (Embedded Database) ขนาดเล็ก เหมาะสำหรับการใช้งานอย่างง่าย และทรัพยากรจำกัด รองรับการทำงานพร้อมกัน และสามารถสำรองข้อมูลได้ [3] |
| BigTable | คอลัมน์ | ฐานข้อมูลขนาดใหญ่ที่มีหลายมิติ ทั้งแถว คอลัมน์ และมิติของเวลา มีประสิทธิภาพในการประมวลผลได้เป็นอย่างดีและรองรับการขยายฐานข้อมูลในอนาคต [3], [7] |
| MongoDB | เอกสาร | โครงสร้างฐานข้อมูลคล้ายกับฐานข้อมูลเชิงสัมพันธ์ สามารถค้นหาข้อมูลจำนวนมากได้อย่างรวดเร็ว รองรับการขยายฐานข้อมูล และการทำงานหนักได้ดี [3], [7] |
| Neo4j | กราฟ | โครงสร้างฐานข้อมูลอยู่ในรูปแบบกราฟ ซึ่งช่วยให้มีความคล่องตัวและมีความรวดเร็วในการพัฒนา รองรับข้อมูลขนาดใหญ่ และการขยายเครื่องแม่ข่าย |

ภาษาสอบถามแบบไม่มีโครงสร้างเป็นองค์ประกอบหนึ่งที่สำคัญของฐานข้อมูลไม่สัมพันธ์ ซึ่งมีรูปแบบคำสั่งในการจัดการข้อมูลที่ไม่แน่นอนขึ้นอยู่กับประเภทและ




**รูปที่ 1** สถาปัตยกรรมของเว็บเชิงความหมาย

ผู้ผลิตฐานข้อมูล เน้นการ Denormalize เพื่อการทำงานที่รวดเร็วโดยไม่มีการ Join ทั้งนี้ในส่วนการเชื่อมต่อจะง่ายกว่าฐานข้อมูลเชิงสัมพันธ์ที่ไม่ต้องมีค่าใช้จ่ายสำหรับการจัดซื้ออุปกรณ์ฮาร์ดแวร์เพิ่มเติม แต่จำเป็นต้องมี IP และ Port สำหรับการใช้งาน [8]

## 2.2 เว็บเชิงความหมาย

เว็บเชิงความหมายเป็นเทคโนโลยีเว็บ 3.0 (Web 3.0) ที่เน้นการเชื่อมโยงข้อมูลกันมากขึ้น โดยทำให้คอมพิวเตอร์สามารถเข้าใจข้อมูลที่สอดคล้องกับความเข้าใจของมนุษย์ ซึ่งสามารถแลกเปลี่ยนข้อมูล และนำไปประมวลผลต่อได้อย่างอัตโนมัติ โดยสถาปัตยกรรมของเว็บเชิงความหมายจะเกี่ยวข้องกับการระบุและรูปแบบของทรัพยากรที่ต้องการใช้งาน (URI/IRI and Unicode) ภาษาเอ็กซ์เอ็มแอล (XML Language) เนมสเปซ (Namespaces) โครงร่างภาษาเอ็กซ์เอ็มแอล (XML Schema) ภาษาสอบถามสำหรับเอ็กซ์เอ็มแอล (XML Query) แนวคิดของอาร์ดีเอฟและไวยากรณ์ (RDF Model and Syntax) ออนโทโลยี (Ontology) กฎ (Rules) ภาษาสอบถาม (Query Language) ตรรกะ (Logics) การตรวจสอบข้อมูลตามกฎ (Proof) และความปลอดภัยของซอฟต์แวร์ (Trusted SW) เช่น ลายเซ็นอิเล็กทรอนิกส์ (Digital Signature) และการเข้ารหัส (Encryption) ซึ่งสามารถแสดงได้ดังรูปที่ 1 [9]

ออนโทโลยี (Ontology) เป็นแนวคิดของเทคโนโลยีเว็บเชิงความหมายที่อธิบายรูปแบบโครงสร้างความสัมพันธ์ระหว่างข้อมูลในขอบเขต (Domain) ที่สนใจ [10] โดยประกอบด้วยแนวคิด (Concepts) คุณสมบัติ (Properties) ความสัมพันธ์ (Relationships) ข้อกำหนดการสร้างความสัมพันธ์ (Axioms) และตัวอย่างข้อมูล (Instances) โดยในงานวิจัยนำโครงสร้างออนโทโลยีมาช่วยในการค้นคืนข้อมูลให้ตรงตามความต้องการของผู้ใช้งานมากยิ่งขึ้น

### 2.3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยต่างๆ เกี่ยวกับฐานข้อมูลไม่สัมพันธ์ ตัวอย่างเช่น งานวิจัยที่นำฐานข้อมูล NoSQL มาประยุกต์ใช้ในการวิเคราะห์และจัดหมวดหมู่ข้อมูลที่มีขนาดใหญ่ [11] การเปรียบเทียบระหว่างฐานข้อมูลเชิงสัมพันธ์และไม่สัมพันธ์ [12] - [14] ซึ่งจากงานวิจัยแสดงให้เห็นว่าฐานข้อมูลไม่สัมพันธ์เหมาะสมในการประยุกต์ใช้กับข้อมูลขนาดใหญ่ แต่ยังขาดมาตรฐานที่นำมาบังคับใช้กับภาษาสอบถาม [13] หรือบางงานวิจัยที่นำเสนอการนำ HPQS มาช่วยในการประมวลผลฐานข้อมูลที่ต่างแพลตฟอร์มระหว่างฐานข้อมูลเชิงสัมพันธ์และฐานข้อมูลไม่สัมพันธ์ [15] หรือการทดสอบความเร็วการประมวลผลระหว่าง MySQL และ MongoDB [16] ซึ่งพบว่า MongoDB เร็วกว่า MySQL หลายเท่าตัว เป็นต้น

งานวิจัยที่เกี่ยวข้องกับออนโทโลยีและเว็บเซอร์วิส ตัวอย่าง เช่น การออกแบบสถาปัตยกรรมออนโทโลยีเพื่อบูรณาการข้อมูลสำหรับการบริหารงานที่ใช้กรณีศึกษาข้อมูลกิจการนักศึกษา มหาวิทยาลัยมหาสารคาม โดยแบ่งสถาปัตยกรรมออกเป็น 3 ระดับ ได้แก่ ระดับข้อมูล ระดับสื่อกลาง และระดับนำเสนอ จากผลการวิจัยพบว่าออนโทโลยีทำให้การค้นคืนมีประสิทธิภาพสูงขึ้น [17] งานวิจัยที่เกี่ยวข้องกับการจัดการองค์ความรู้ตามแนวคิด SECI ซึ่งนำออนโทโลยีมาช่วยจัดการองค์ความรู้และค้นคืนเชิงความหมาย [18] หรือการบูรณาการสารสนเทศขององค์กรชุมชนด้วยออนโทโลยี ซึ่งในงานวิจัยประกอบด้วย 2 ส่วนหลัก ได้แก่ ส่วนแรกเกี่ยวข้องกับการจัดการข้อมูลในระดับ Local View และ Global View และส่วนที่สองเกี่ยวข้องกับการให้บริการค้นคืนข้อมูล [19] หรืองานวิจัยที่ทำการสร้างระบบบัญชีจำแนกโรคระหว่างประเทศ ฉบับภาษาไทย (ICD-10-TM) โดยประยุกต์ใช้ออนโทโลยีสำหรับจัดเก็บและการค้นคืนข้อมูล [20] และตัวอย่างงานวิจัยการพัฒนาระบบผู้เชี่ยวชาญเพื่อวินิจฉัยและให้คำแนะนำ ซึ่งประยุกต์ใช้กฎและฐานความรู้ออนโทโลยีเข้ามาช่วยในการวินิจฉัย และให้คำแนะนำแก่ผู้ป่วยได้เป็นอย่างดี ซึ่งมีค่าความแม่นยำ (Precision) เท่ากับ 96.89% และค่าความระลึก (Recall) เท่ากับ 95.39% [21]

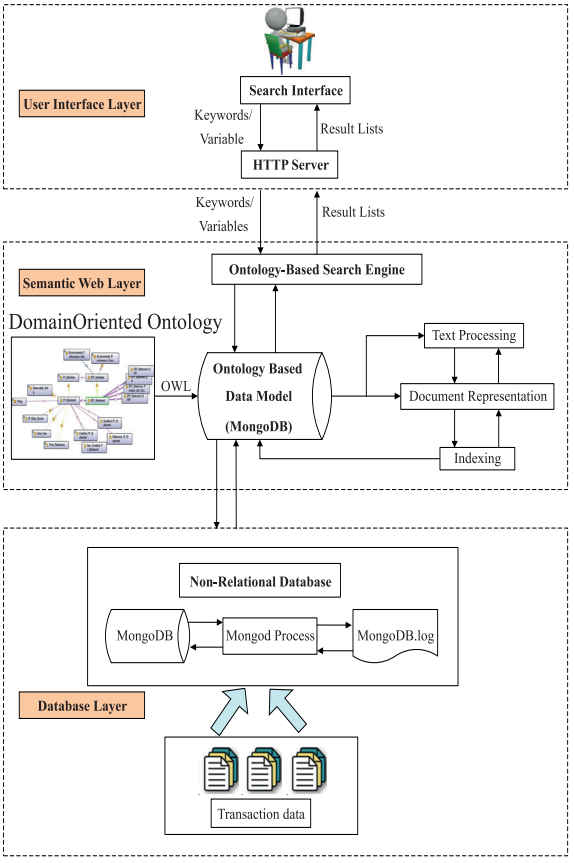
## 3. วิธีการดำเนินการวิจัย

### 3.1 การศึกษาและรวบรวมข้อมูล

คณะผู้วิจัยได้ทำการศึกษา รวบรวมข้อมูล และวิเคราะห์ปัญหาต่างๆ เกี่ยวกับการจัดการข้อมูลขนาดใหญ่ ภาษาสอบถามแบบไม่มีโครงสร้าง ซอฟต์แวร์ที่ใช้ในการพัฒนาฐานข้อมูลไม่สัมพันธ์ แนวคิด และวิธีการต่างๆ ที่นำมาประยุกต์ใช้ในการค้นคืนข้อมูล ซึ่งพบว่าในปัจจุบันข้อมูลมีเป็นจำนวนมาก และมีแนวโน้มเพิ่มขึ้นอย่างรวดเร็ว อีกทั้งการค้นคืนข้อมูลอาจได้ผลลัพธ์ไม่ตรงตามความต้องการของผู้ใช้งาน และหากข้อมูลมีจำนวนมากอาจทำให้การประมวลผลล่าช้า

### 3.2 การวิเคราะห์และออกแบบกรอบแนวคิดของการวิจัย

หลังจากทำการศึกษา รวบรวมข้อมูล และวิเคราะห์ปัญหา วัตถุประสงค์ของงานวิจัยนี้จึงได้ทำการออกแบบฐานข้อมูลไม่สัมพันธ์โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างและโครงสร้างออนโทโลยี เพื่อนำมาช่วยในการจัดเก็บและค้นคืนข้อมูล ซึ่งใช้กรณีศึกษาการออกใบแจ้งค่าใช้บริการ ได้แก่ ข้อมูล ลูกค้า ประเภทลูกค้า ผลิตภัณฑ์ ประเภทบริการ และหน่วยงานรับผิดชอบ




**รูปที่ 2** กรอบแนวคิดของการวิจัย

กรอบแนวคิดของการวิจัยสามารถแสดงได้ดังรูปที่ 2 ซึ่งสามารถแบ่งการทำงานออกเป็น 3 ส่วนหลัก คือ ชั้นการติดต่อกับผู้ใช้งาน (User Interface Layer) เป็นส่วนสำหรับผู้ใช้งานกรอกข้อมูล เพื่อใช้ในการค้นคืนข้อมูลในส่วนชั้นเว็บเชิงความหมาย (Semantic Web Layer) เป็นส่วนการประมวลผลการค้นคืนข้อมูลเว็บเชิงความหมาย โดยใช้โครงสร้างออนโทโลยีกำหนดความสัมพันธ์ระหว่างข้อมูลตามขอบเขตที่สนใจ เพื่อช่วยให้การค้นคืนข้อมูลมีความถูกต้องมากขึ้น โดยจัดเก็บข้อมูลลงใน MongoDB ที่เป็นฐานข้อมูลไม่สัมพันธ์แบบเอกสาร และชั้นฐานข้อมูล (Database Layer) ใช้สำหรับจัดเก็บข้อมูลการประมวลผลรายการที่ใช้ฐานข้อมูลไม่สัมพันธ์แบบเอกสารเช่นเดียวกับในส่วนของ Semantic Web Layer

Json Text of Resulting Documents:

```
{
    “subject” : “http://www.w3.org/subject”,
    “predicate” : “http://www.w3.org/predicate”,
    “object” : “http://www.w3.org/ object”
}
```

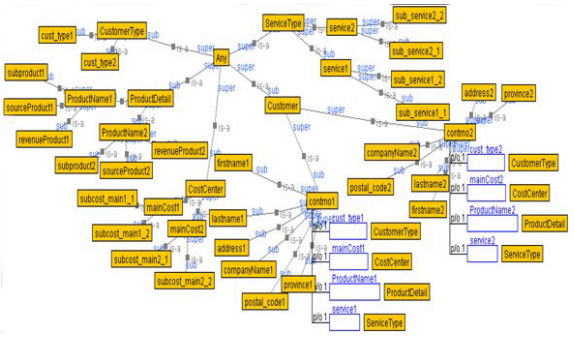
**รูปที่ 3** ตัวอย่างโครงสร้างเอกสาร Json เพื่อจัดเก็บข้อมูลลงใน MongoDB

### 3.3 การพัฒนาฐานข้อมูลไม่สัมพันธ์ร่วมกับเทคโนโลยีเว็บเชิงความหมาย

ขั้นตอนนี้ทำการพัฒนาฐานข้อมูลไม่สัมพันธ์เพื่อจัดเก็บและรองรับการค้นคืนข้อมูลขนาดใหญ่ โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับออนโทโลยีที่เป็นเทคโนโลยีเว็บเชิงความหมาย ซึ่งมีรายละเอียดต่าง ๆ ดังนี้

งานวิจัยทำการพัฒนาออนโทโลยีตามวงจรชีวิตออนโทโลยี (Ontology Life Cycle) [22] ซึ่งใช้ซอฟต์แวร์ Hozo: Ontology Editor เวอร์ชัน 5.2.36 โดยโครงสร้างออนโทโลยีถูกจัดเก็บลงใน MongoDB เวอร์ชัน 2.4.5 ซึ่งเป็นซอฟต์แวร์โอเพนซอร์ส (Open Source) และเป็นฐานข้อมูลไม่สัมพันธ์แบบเอกสารที่มีรูปแบบใกล้เคียงกับฐานข้อมูลเชิงสัมพันธ์ และทำการออกแบบโครงสร้างเอกสารในรูปแบบ Json โดยแบ่งข้อมูลออกเป็น 3 ส่วน ตามมาตรฐานของ World Wide Web Consortium (W3C) ที่ประกอบด้วยชื่อ (Subject) คุณสมบัติ (Predicate) และวัตถุ (Object) ดังรูปที่ 3 [23]

โครงสร้างออนโทโลยีที่ได้จากการพัฒนาสามารถแสดงได้ดังรูปที่ 4 ซึ่งประกอบด้วยข้อมูลประเภทลูกค้า 99 แถว ลูกค้า 3,169,164 แถว ประเภทบริการ 198 แถว ผลิตภัณฑ์ 2,856 แถว หน่วยงานที่รับผิดชอบการบริการ 656 แถว และข้อมูลการออกใบแจ้งค่าใช้บริการ 14,055,993 แถว ได้แก่ ข้อมูลเลขที่ลูกค้า (Contrno) เลขที่ใบแจ้งค่าใช้บริการ (BillReference) บริการ (BillGroup) รหัสผลิตภัณฑ์ (ProductCode) วันออกใบแจ้งค่าใช้บริการ (InvDate) วันครบกำหนดชำระ (DueDate) งวดการใช้บริการ



**รูปที่ 4** โครงสร้างออนโทโลยีการออกใบแจ้งค่าใช้บริการ

(Period) จำนวนเงินไม่รวมภาษี (Amount) ภาษี (Vat) จำนวนเงินรวมภาษี(Total) ประเภทลูกค้า(CustomerGroup) สกุลเงิน (Curr) และรหัสหน่วยงานที่รับผิดชอบลูกค้า (CostCenter)

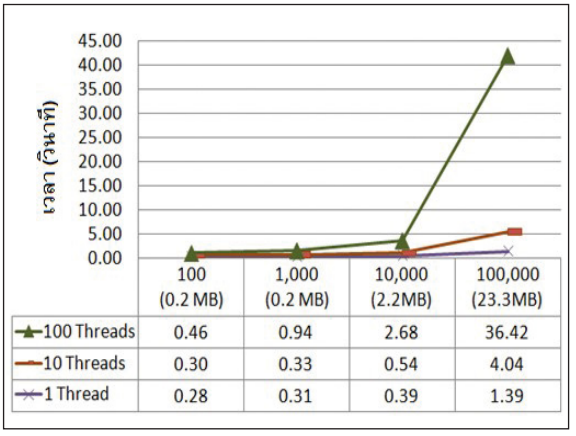
### 3.4 การทดสอบความเร็วในการประมวลผล

การทดสอบความเร็วในการประมวลผลแบ่งการทดสอบโดยใช้ 3 ตัวดำเนินการ (Operators) ได้แก่ การอ่าน (Select) การแก้ไข (Update) และการลบ (Delete) ซึ่งใช้จำนวนเธรดและจำนวนข้อมูลที่แตกต่างกัน โดยเธรดคือการทำงานได้มากกว่าหนึ่งโปรแกรม ซึ่งรายละเอียดการทดสอบความเร็วการประมวลผลสามารถแสดงได้ดังตารางที่ 2

**ตารางที่ 2** จำนวนเธรดและข้อมูลที่ใช้ในการทดสอบ

| Operators | No of Threads | | | No of Selected Rows | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 100 | 100 | 1,000 | 10,000 | 100,000 |
| Select | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Update | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Delete | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

การทดสอบความเร็วในการประมวลผล จะคำนวณจากเวลาทั้งหมดที่ใช้ในการประมวลผล (Total Time) ซึ่งสามารถแสดงได้ดังสมการที่ 1 โดยการทดสอบแต่ละกรณีจะทำการทดลอง 3 ครั้งและหาค่าเฉลี่ยเลขคณิต (Arithmetic Mean)




| | 100 (0.2 MB) | 1,000 (0.2 MB) | 10,000 (2.2MB) | 100,000 (23.3MB) |
|---|---|---|---|---|
| 100 Threads | 0.46 | 0.94 | 2.68 | 36.42 |
| 10 Threads | 0.30 | 0.33 | 0.54 | 4.04 |
| 1 Thread | 0.28 | 0.31 | 0.39 | 1.39 |

**รูปที่ 5** ความเร็วในการอ่านข้อมูล

$$Total_Time = End_Time - Begin_Time \quad (1)$$

Total_Time คือเวลาทั้งหมดที่ใช้ในการประมวลผล
End_Time คือเวลาที่ใช้ในการ Select และ Fetch ข้อมูล
Begin_Time คือเวลาเริ่มต้นในการประมวลผลข้อมูล

## 4. ผลการวิจัย

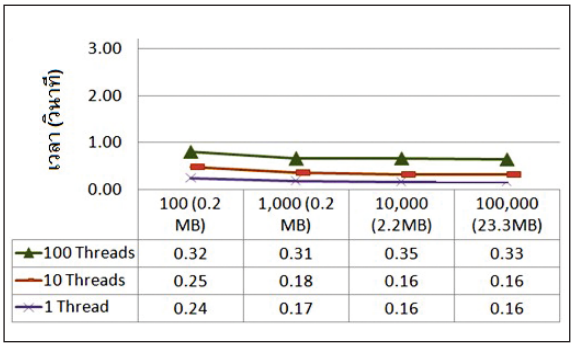
ผลการทดสอบความเร็วในการประมวลผลข้อมูลของแต่ละตัวดำเนินการ โดยพิจารณาจากจำนวนเธรดหรือจำนวนการทำงานได้มากกกว่าหนึ่งโปรแกรมในเวลาเดียวกัน จำนวนข้อมูล และเวลาทั้งหมดที่ใช้ในการประมวลผล ซึ่งผลการวิจัยมีรายละเอียดดังต่อไปนี้

### 4.1 ผลการทดสอบความเร็วในการอ่านข้อมูล

จากผลการทดสอบความเร็วในการอ่านข้อมูล (Select) ของฐานข้อมูลไม่สัมพันธ์ พบว่าจำนวนเธรดเท่ากับ 1 และ 10 ที่จำนวนข้อมูล 100, 1,000 และ 10,000 แถวฐานข้อมูลใช้เวลาในการอ่านข้อมูลใกล้เคียงกัน แต่ที่จำนวนเธรดเท่ากับ 100 ที่จำนวนข้อมูล 100,000 แถวใช้เวลาในการประมวลผลมากที่สุด คือ 36.42 วินาที ซึ่งสามารถแสดงได้ดังรูปที่ 5

| | 100 (0.2 MB) | 1,000 (0.2 MB) | 10,000 (2.2MB) | 100,000 (23.3MB) |
|---|---|---|---|---|
| 100 Threads | 0.31 | 0.32 | 0.32 | 0.32 |
| 10 Threads | 0.18 | 0.16 | 0.17 | 0.16 |
| 1 Thread | 0.15 | 0.16 | 0.16 | 0.15 |

**รูปที่ 6** ความเร็วในการแก้ไขข้อมูล




| | 100 (0.2 MB) | 1,000 (0.2 MB) | 10,000 (2.2MB) | 100,000 (23.3MB) |
|---|---|---|---|---|
| 100 Threads | 0.32 | 0.31 | 0.35 | 0.33 |
| 10 Threads | 0.25 | 0.18 | 0.16 | 0.16 |
| 1 Thread | 0.24 | 0.17 | 0.16 | 0.16 |

**รูปที่ 7** ความเร็วในการลบข้อมูล

### 4.2 ผลการทดสอบความเร็วในการแก้ไขข้อมูล

จากผลการทดสอบความเร็วในการแก้ไขข้อมูล (Update) ของฐานข้อมูลไม่สัมพันธ์ พบว่าจำนวนเธรดและจำนวนข้อมูลที่แตกต่างกัน ฐานข้อมูลใช้เวลาในการประมวลผลใกล้เคียงกัน โดยแต่ละการทดสอบใช้เวลาในการประมวลผลไม่ถึง 1 วินาที ซึ่งสามารถแสดงได้ดังรูปที่ 6

### 4.3 ผลการทดสอบความเร็วในการลบข้อมูล

จากผลการทดสอบความเร็วในการลบข้อมูล (Delete) ของฐานข้อมูลไม่สัมพันธ์ พบว่าจำนวนเธรดและจำนวนข้อมูลที่แตกต่างกัน ฐานข้อมูลใช้เวลาในการประมวลผลใกล้เคียงกัน โดยแต่ละการทดสอบใช้เวลาประมวลผลไม่ถึง 1 วินาที ซึ่งสามารถแสดงได้ดังรูปที่ 7

## 5. สรุป

งานวิจัยนี้ได้ทำการออกแบบและพัฒนาฐานข้อมูลไม่สัมพันธ์ เพื่อรองรับการจัดเก็บและการค้นคืนข้อมูลขนาดใหญ่ โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับโครงสร้างออนโทโลยีที่เป็นเทคโนโลยีเว็บเชิงความหมาย ซึ่งใช้กรณีศึกษาการออกใบแจ้งค่าใช้บริการทำการจัดเก็บข้อมูลลงใน MongoDB ที่เป็นฐานข้อมูลไม่สัมพันธ์แบบเอกสาร โดยจัดเก็บในรูปแบบเอกสาร Json ในการวิจัยได้ทำการทดสอบความเร็วในการประมวลผลข้อมูลจาก 3 ตัวดำเนินการ ได้แก่ การอ่าน (Select) การแก้ไข (Update) และการลบ (Delete)

จากการศึกษาพบว่าผลการทดสอบความเร็วในการอ่านข้อมูลที่จำนวนเธรดเท่ากับ 1 และ 10 ที่จำนวนข้อมูล 100, 1,000 และ 10,000 แถว ใช้เวลาในการอ่านข้อมูลใกล้เคียงกัน แต่ที่จำนวนเธรดเท่ากับ 100 ที่จำนวนข้อมูล 100,000 แถว ใช้เวลาในการประมวลผลมากที่สุด คือ 36.42 วินาที ส่วนการแก้ไขและการลบข้อมูล พบว่าจำนวนเธรดและจำนวนข้อมูลที่แตกต่างกันฐานข้อมูลไม่สัมพันธ์ใช้เวลาในการประมวลผลใกล้เคียงกัน โดยแต่ละการทดสอบใช้เวลาประมวลผลไม่ถึง 1 วินาที ดังนั้นจะเห็นได้ว่าฐานข้อมูลไม่สัมพันธ์ MongoDB มีความเหมาะสมกับการนำมาใช้จัดเก็บและค้นคืนข้อมูลขนาดใหญ่ ซึ่งสามารถประมวลผลได้อย่างรวดเร็ว มีรูปแบบโครงสร้างคล้ายคลึงกับฐานข้อมูลเชิงสัมพันธ์ ทำให้ง่ายต่อการเรียนรู้และการใช้งาน แต่ฐานข้อมูลดังกล่าวยังไม่รองรับข้อมูลเชิงความหมายเท่าที่ควร โดยทางคณะผู้วิจัยจึงต้องทำการออกแบบเพื่อให้สอดคล้องกับแนวคิดเว็บเชิงความหมาย

งานวิจัยในอนาคตจะทำการทดสอบความถูกต้องและความแม่นยำในการค้นคืนข้อมูลนอกเหนือจากการทดสอบความเร็วในการประมวลผลเพียงอย่างเดียว และนำฐานข้อมูลไม่สัมพันธ์ประเภทอื่นๆ มาประยุกต์ใช้ตัวอย่างเช่น ฐานข้อมูลไม่สัมพันธ์แบบกราฟที่มีคุณสมบัติรองรับการทำงานแบบโหนดมาช่วยในการค้นคืนข้อมูลเชิงความหมาย เพื่อให้เกิดประสิทธิภาพการค้นคืนข้อมูลมากยิ่งขึ้น

**เอกสารอ้างอิง**

[1] K. Brown, "Global Internet Report 2014," Internet Society, 2014.

[2] T. Webster, "The Social Habit 2012," Edison Research, 2012.

[3] C. Tauro, et al., "Comparative Study of the New Generation, Agile, Scalable, High Performance NOSQL Databases," *International Journal of Computer Applications*, vol. 48, no. 20, pp. 20-24, 2012.

[4] M. Sharma and R. Patel, "A Survey on Information Retrieval Models, Techniques and Applications, *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 11, pp. 542-545, 2013.

[5] A. K. Zaki, "NoSQL Databases: New Millennium Database for Big Data, Big Users, Cloud Computing and Its Security Challenges," *International Journal of Research in Engineering and Technology (IJRET)*, vol. 3, no. 15, pp. 403-409, 2014.

[6] V. Sharma and M. Dave, "SQL and NoSQL Databases," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 8, pp. 20-27, 2012.

[7] N. Grover and R. Wason, "Comparative Analysis of Pagerank and HITS Algorithms," *International Journal of Engineering Research & Technology (IJERT)*, vol. 1, no. 8, pp. 1-15, 2012.

[8] K. Orend, "Analysis and Classification of NoSQL Databases and Evaluation of their Ability to Replace an Object-Relational Persistence Layer," Master's Thesis, Software Engineering for Business Information Systems, Technische Universitat Muchen, Germany, 2010.

[9] M. Kifer, et al., "A Realistic Architecture for the Semantic Web," *Rules and Rule Markup Languages for the Semantic Web Lecture Notes in Computer Science*, vol. 3791, pp. 17-29, 2005.

[10] W. Chotirat, P. Boonrawd, and S. Na Wichian, "Developing an Ontology Knowledge Based for Automatic Online News Analysis," *Information Technology Journal*, vol. 7, no. 14, pp. 13-18, 2011.

[11] A. B. M. Moniruzzaman and S. A. Hossain, "NoSQL Database: New Era of Databases for Big Data Analytics-Classification, Characteristics and Comparison," *International Journal of Database Theory and Application*, vol. 6, no. 4, pp. 1-13, 2013.

[12] M. Mohamed, O. Altrafi, and M. Ismail, "Relational vs. NoSQL Database: A Survey," *International Journal of Computer and Information Technology*, vol. 3, no. 3, pp. 598-601, 2014.

[13] A. Nayak, A. Poriya, and D. Poojary, "Type of NOSQL Databases and Its Comparison with Relational Databases," *International Journal of Applied Information Systems (IJAIS)*, vol. 5, no. 4, pp. 16-19, 2013.

[14] P. Srisuktaksin and P. Boonrawd, "Analysis of Data Processing Performance between Relational and Non-Relational Databases of Documents," *Proceeding of the 10th National Conference on Computing and Information Technology (NCCIT 2014)*, 2014.

[15] P. T. Thant and T. T. Naing, "Hybrid Query Processing System (HQPS) for Heterogeneous Database (Relational and NoSQL)," in *Proceeding of the International Conference on Computer Networks and Information Technology*, 2014, pp. 53-58.

[16] A. Joshi, S. Haradhvala, and C. Lamb, "Oracle NoSQL Database-Scalable, Transactional Key-Value Store," in *Proceeding of the 2nd International Conference on Advances in Information Mining and Management (IMMM 2012)*, 2013, pp. 75-78.

[17] J. Phuboon-ob and W. Sangkatip, "Ontology-Based Architecture Design to Integrate Information for Student Affair Administration of Mahasarakham University," *Journal of Science and Technology Mahasarakham University*, pp. 271-280, 2013.

[18] A. Keawlaiod and P. Boonrawd, "Semantic Knowledge Management Using SECI Model for Computerized Auditing," in *Proceeding of the 9th National Conference on Computing and Information Technology (NCCIT 2013)*, 2013.

[19] Y. Jiydumrong and G. Sriharee, "An Information Integration of Community Organization Using Ontology," in *Proceeding of the 2013 International Computer Science and Engineering Conference (ICSEC 2013)*, 2013, pp. 460-465.

[20] W. Paoin and S. Nitsuwat. "Development of Experience Base Ontology to Increase Competency of Semi-automated ICD-10-TM Coding System," *Information Technology Journal*, vol. 8, no. 2, pp. 95-99, 2012.

[21] J. Tiansaard and O. Imsombut. "An Ontology Based Expert System for Diagnosis and Recommendation in Hemodialysis Patient," *Journal of Information Science and Technology (JIST)*, vol. 3, no. 2, pp. 22-30, 2012.

[22] B. Kapoor and S. Sharma, "A Comparative Study Ontology Building Tools for Semantic Web Applications," *International Journal of Web & Semantic Technology (IJWesT)*, vol.1, no.3, pp. 1-13, 2010.

[23] J. Li, "Building Distributed Index for Semantic Web Data," in *Proceeding of the 23rd IEEE International Conference on Advanced Information Networking and Applications (AINA-09)*, 2009, pp.660-667.