การออกแบบสถาปัตยกรรมออนโทโลยีเพื่อบูรณาการข้อมูลสำหรับการบริหารงานกิจการนิสิต มหาวิทยาลัยมหาสารคาม

Ontology-based Architecture Design to Integrate Information for Student Affair Administration of Mahasarakham University

จิรัฏฐา ภูบุญอบ,[1*] วรวิทย์ สังฆทิพย์[2]
Jiratta Phuboon-ob,[1] Worawith Sangkatip[2]

บทคัดย่อ

ข้อมูลการบริหารงานกิจการนิสิตประกอบด้วยข้อมูลด้านต่างๆ เช่น ข้อมูลนักศึกษา ข้อมูลด้านทุนการศึกษา ข้อมูลโครงการ ข้อมูลด้านวินัยนิสิต ข้อมูลบริการนิสิต เป็นต้น ซึ่งงานด้านกิจการนิสิตถูกกระจายให้ฝ่ายพัฒนานิสิตแต่ละคณะเป็นผู้รับผิดชอบ จึงทำให้แต่ละคณะต่างพัฒนาระบบสารสนเทศขึ้นมาใช้งาน ส่งผลให้ข้อมูลสารสนเทศมีความหลากหลายแตกต่างกัน เมื่อต้องการนำระบบสารสนเทศมาใช้งานร่วมกันจึงไม่สามารถทำได้ เนื่องจากข้อมูลสารสนเทศมีความหลากหลายของแหล่งข้อมูล ก่อให้เกิดความขัดแย้งเชิงความหมายและเชิงโครงสร้างของข้อมูล จากปัญหาดังกล่าวจึงได้วิจัยและพัฒนาการบูรณาการข้อมูลเข้าด้วยกัน โดยทำการออกแบบสถาปัตยกรรมการบูรณาการออนโทโลยีแบบผสม ซึ่งแบ่งเป็นสามระดับคือ ระดับของแหล่งข้อมูล ระดับสื่อกลาง และระดับนำเสนอ เพื่อรองรับการบูรณาการข้อมูลจากแหล่งข้อมูลที่มีหลากหลาย พร้อมนำเสนอการเชื่อมโยงออนโทโลยีที่มีความหลากหลายของแหล่งข้อมูลแบบกึ่งอัตโนมัติ ด้วยการหาความคล้ายคลึงเชิงความหมายระหว่างออนโทโลยี และใช้ฐานข้อมูลเวิร์ดเน็ตเพื่อหาความสัมพันธ์ระหว่างคำศัพท์ของข้อมูล ซึ่งการเชื่อมโยงได้พิจารณาถึงระดับคลาสและคุณสมบัติของคลาสเพื่อให้การบูรณาการและการสืบค้นข้อมูลมีประสิทธิภาพมากยิ่งขึ้น ภายหลังจากทำการทดสอบด้วยค่าความระลึก ค่าความแม่นยำ และค่าเอฟ พบว่าได้ค่าอยู่ที่ร้อยละ 94 ร้อยละ 83 และร้อยละ 88 ตามลำดับ จากผลลัพธ์ดังกล่าวแสดงให้เห็นว่าการสืบค้นมีประสิทธิภาพสูง

คำสำคัญ: สถาปัตยกรรมออนโทโลยี การเชื่อมโยงออนโทโลยี ความคล้ายคลึงเชิงความหมาย ฐานข้อมูลเวิร์ดเน็ต

Abstract

Student Affairs Administrative database consist of various types of information such as student profiles, scholarships, projects, disciplinary and information services etc. The student affairs are delegated to

[1] อาจารย์ สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150

[2] อาจารย์ สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม อำเภอเมือง จังหวัดมหาสารคาม 44000

[1] Lecturer, Department of Information Technology, Faculty of Informatics, Mahasarakham University, Kantarawichai District, Maha Sarakham 44150, Thailand.

[2] Lecturer, Department of Information Technology, Faculty of Information Technology, Rajabhat Mahasarakham University, Mueang District, Maha Sarakham 44000, Thailand.

* Corresponding author: Jiratta Phuboon-ob, Department of Information Technology, Faculty of Informatics, Mahasarakham University, Kantarawichai District, Maha Sarakham 44150, Thailand.

eachfaculty. This leads to various information systems in these faculties. As a result, it turns to be more difficult than attempting to gather the data together because it brings out the conflict in terms of semantic structures. Therefore, the research on data integration has been conducted for serving the data from different sources. Architectural integration was designed by using ontology combination form which can be divided into three levels; including sources, channels and presentations of data. This system can serve the data from different sources by presenting the ontology connection. It can look for the similarity of semantic ontology, and the WordNet can be used to look for the correlation between vocabularies. The connection has been reviewed by considering the class levels and qualifications of class properties. The results show that the accuracies are at 94%, 83%, and 88%, respectively using recall, precision and F-measure. This may demonstrate that the data searching can be more effective.

Keyword: Ontology Architecture, Ontology Mapping, Semantic Similarity, WordNet

บทนำ

ในปัจจุบันข้อมูลด้านการบริหารงานกิจการนิสิตมีปริมาณมาก ประกอบด้วยข้อมูลด้านต่างๆ เช่น ข้อมูลนักศึกษา ข้อมูลด้านทุนการศึกษา ข้อมูลโครงการ ข้อมูลด้านวินัยนิสิต ข้อมูลบริการนิสิตด้านอื่นๆ เป็นต้น ข้อมูลถูกจัดเก็บในรูปแบบที่หลากหลาย ทั้งในรูปแบบของเอกสาร อีเมล์ เว็บไซต์ หรือฐานข้อมูล ข้อมูลเหล่านี้มีความจำเป็นในการบริหารงานเป็นอย่างมาก แต่เนื่องจากงานด้านกิจการนิสิตถูกกระจายให้ฝ่ายพัฒนานิสิตแต่ละคณะเป็นผู้รับผิดชอบ จึงทำให้แต่ละคณะต่างพัฒนาระบบสารสนเทศขึ้นมาใช้งาน จึงส่งผลให้ข้อมูลสารสนเทศมีความหลากหลายแตกต่างกัน (Heterogeneous Information System) เมื่อผู้บริหารต้องการทราบข้อมูลในภาพรวมระหว่างแต่ละคณะจึงไม่สามารถกระทำได้ เนื่องจากไม่มีการวางแผนในการจัดการและใช้ข้อมูลร่วมกัน ทำให้ข้อมูลเรื่องเดียวกันของคณะเดียวกันหรือต่างคณะมีความซ้ำซ้อนและขัดแย้งไม่ตรงกัน ก่อให้เกิดความสับสนในการนำข้อมูลไปใช้งานและการตีความข้อมูลส่งผลให้เกิดปัญหากับการสืบค้นข้อมูลที่ต้องการจากระบบ ผลลัพธ์ที่ได้ออกมาไม่ตรงกับความต้องการของผู้ใช้งาน ทำให้เกิดความล่าช้าในการบริหารงาน ส่วนหนึ่งเป็นผลมาจากระบบสืบค้นข้อมูลยังไม่สามารถตอบสนองความต้องการของผู้ใช้ได้อย่างเพียงพอ โดยระบบสืบค้นข้อมูลส่วนใหญ่เป็นการค้นหาโดยใช้คำสำคัญ (Keyword Search) กล่าวคือ เมื่อผู้ใช้ป้อนคำค้นให้ระบบ ระบบจะทำการค้นคืนเอกสาร โดยค้นหาจากฐานข้อมูลที่มีคำดังกล่าวปรากฏอยู่ ข้อมูลที่ได้จากการสืบค้นจะออกมาจำนวนมาก ผู้ค้นหาต้องใช้เวลาในการคัดเลือกข้อมูลเพื่อที่จะได้ข้อมูลที่สอดคล้องตรงกับความต้องการ

ปัญหาดังกล่าวสามารถแก้ได้โดยการบูรณาการข้อมูล (Data Integration) เพื่อให้เกิดการใช้งานร่วมกันระหว่างแต่ละคณะด้วยการออกแบบฐานออนโทโลยีโดยกำหนดนิยามความหมายที่เป็นเรื่องเดียวกันของแต่ละแหล่งข้อมูล (Data Source) ให้มีความหมายความสัมพันธ์ที่สอดคล้องตรงกัน เพื่อเป็นฐานในการบูรณาการข้อมูลและการเชื่อมโยงแหล่งข้อมูลที่หลากหลายเข้าด้วยกัน เพื่อแก้ปัญหาความหลากหลายของข้อมูล พร้อมทั้งยังช่วยให้การสืบค้นข้อมูลมีประสิทธิภาพมากยิ่งขึ้นเนื่องจากเป็นการสืบค้นเชิงความหมาย (Semantic)

ดังนั้นการวิจัยครั้งนี้จึงเสนอวิธีการออกแบบสถาปัตยกรรมออนโทโลยีเพื่อบูรณาการข้อมูลจากแหล่งข้อมูลที่มีความหลากหลาย พร้อมเสนอวิธีการเชื่อมโยงออนโทโลยีเพื่อให้สามารถเรียกใช้ข้อมูลกิจการนิสิตร่วมกันได้ระหว่างแต่ละคณะภายในมหาวิทยาลัย ซึ่งจะส่งผลให้การสืบค้นข้อมูลมีประสิทธิภาพมากยิ่งขึ้น

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ออนโทโลยี (Ontology)

ออนโทโลยี คือแนวคิดในการบรรยายองค์ความรู้และแสดงสิ่งที่เราสนใจอย่างมีขอบเขตตามโครงสร้างและความสัมพันธ์หรือการนิยามคำศัพท์และความหมายของคำศัพท์สำหรับบรรยายความรู้ที่สนใจ ซึ่งโครงสร้างความสัมพันธ์ดังกล่าว เครื่องคอมพิวเตอร์สามารถเข้าใจและแปลความได้ โดยหาความสัมพันธ์ระหว่างคลาส หมายรวมถึงลำดับชั้นของคลาส (Class) และคุณสมบัติของคลาส (Properties) ซึ่งการแลกเปลี่ยนข้อมูลระหว่างกันโดยใช้ออนโทโลยีทำให้ไม่เกิดความแตกต่างทางด้านคำศัพท์ [1]

ภาษา OWL (Web Ontology Language)

ภาษา OWL เป็นภาษาที่ใช้สำหรับการอธิบายออนโทโลยีและกำหนดความสัมพันธ์ระหว่างข้อมูลตามขอบเขตที่สนใจ ซึ่งพัฒนาต่อมาจากภาษา RDF และสืบทอดมาจากภาษา DAML+OIL โดยภาษา OWL ได้นำเอาคลาสและคุณสมบัติของคลาสจากภาษา RDF มาใช้ รวมทั้งเพิ่มส่วนของการกำหนดชนิดข้อมูล การบรรยายข้อมูลเชิงตรรกะและการกำหนดขนาดข้อมูลเข้าไป ทำให้ข้อมูลที่ถูกแทนที่มีความหมายมากยิ่งขึ้น ลักษณะการบรรยายจะอยู่ในรูปคลาส คุณสมบัติของคลาสและความสัมพันธ์ระหว่างคลาส เพื่ออธิบายเอ็นทิตี้ (Entity) และความสัมพันธ์ต่างๆ ที่เกิดขึ้น ภาษา OWL มีประสิทธิภาพอย่างมากในการอธิบายเนื้อหาต่างๆ ตามขอบเขตซึ่งคอมพิวเตอร์สามารถอ่านค่าและเข้าใจถึงความหมายของข้อมูล ภาษา OWL แบ่งออกเป็น 3 ประเภท คือ OWL LITE, OWL DL, OWL FULL โดยแต่ละประเภทถูกออกแบบให้เหมาะสำหรับการใช้งานตามกลุ่มการใช้งาน [2]

การเชื่อมโยงออนโทโลยี (Ontology Mapping)

การเชื่อมโยงออนโทโลยี (Ontology Mapping) ในงานด้านการสืบค้นเชิงความหมายจะมีรูปแบบหรือโครงสร้างของออนโทโลยีที่เหมาะสมกับงานนั้นๆ เมื่อต้องการใช้งานออนโทโลยีร่วมกันซึ่งจะเกิดปัญหาในเรื่องของการแทนข้อมูลหรือแนวความคิดที่แตกต่างกัน เช่น ออนโทโลยี A จะแทนผู้ชายให้เป็น Subclass ของคน ส่วนออนโทโลยี B แทนผู้ชายเป็นเพียงแค่ attribute หนึ่งเท่านั้นซึ่งทั้งสองออนโทโลยีนี้จะ equivalence กัน ดังนั้นวิธีการในการจับคู่ระหว่างออนโทโลยีจึงเป็นสิ่งหนึ่งที่สำคัญในการใช้งานออนโทโลยีร่วมกันเพื่อเชื่อมโยงความสัมพันธ์ระหว่างออนโทโลยี การเชื่อมโยงออนโทโลยีสามารถแบ่งออกเป็น 4 ประเภท [3] ดังนี้

1) Ontology Mapping คือ การเชื่อมโยงความสัมพันธ์ระหว่างออนโทโลยีที่มีแนวความคิด (Concept) เหมือนกัน

2) Ontology Alignment คือ กระบวนการวิเคราะห์แนวความคิดที่เหมือนกันและแตกต่างกันระหว่างหลายออนโทโลยี

3) Ontology Merging คือ การผสานออนโทโลยีที่มีความสัมพันธ์กันระหว่างสองออนโทโลยีแล้วสร้างออนโทโลยีขึ้นมาใหม่

4) Ontology Integration คือ การนำออนโทโลยีมารวมกันให้เป็น 1 ออนโทโลยี โดยที่ออนโทโลยีที่นำมารวมกันอาจจะมี Subject ที่แตกต่างกันก็ได้

สถาปัตยกรรมออนโทโลยี (Ontology Architectures)

สถาปัตยกรรมออนโทโลยี คือการออกแบบออนโทโลยีในการใช้งานร่วมกัน ทั้งขอบเขตความรู้มีทั้งกว้างและเฉพาะเจาะจง การจัดการที่ดีที่สุดโดยการแบ่งเป็นขอบเขตความรู้ย่อย (Sub-Domain) แต่ละขอบเขตความรู้ย่อยมีออนโทโลยีย่อยร่วมกัน (Sub-Ontology) เพราะออนโทโลยีย่อยสามารถรวบรวมจัดเก็บองค์ความรู้ได้สมบูรณ์ ทั้งนี้สถาปัตยกรรมออนโทโลยี [4] ได้นำเสนอเป็นสามกลุ่มแนวคิด ดังนี้

1) แนวคิดออนโทโลยีเดี่ยว (Single Ontology Approach) แนวคิดนี้เป็นการจัดเก็บรวบรวมและใช้คำศัพท์เฉพาะที่มีความหมายร่วมกันของแหล่งข้อมูลที่หลากหลายในขอบเขตความรู้เรื่องใดเรื่องหนึ่ง เชื่อมโยงกันในลักษณะออนโทโลยีหลัก (Global Ontology) แนวคิดออนโทโลยีหลักสามารถนำไปประยุกต์ใช้ในการแก้ปัญหาและบูรณาการออนโทโลยีของแหล่งข้อมูลที่มีขอบเขตความรู้เหมือนกันหรือใกล้เคียงกัน

2) แนวคิดออนโทโลยีแบบหลากหลาย (Multiple Ontologies Approach) แนวคิดนี้แต่ละแหล่งข้อมูลถูกอธิบายความหมายด้วยออนโทโลยี

พื้นฐาน (Local Ontology) เฉพาะของแต่ละแหล่งข้อมูลและออนโทโลยีพื้นฐานสามารถเชื่อมโยงกับออนโทโลยีพื้นฐานเฉพาะของแหล่งข้อมูลอื่นๆ ได้ด้วยคำศัพท์ที่สัมพันธ์กัน (Related Terms) ซึ่งแนวคิดนี้มีการบูรณาการงานได้ชัดเจนมากและสนับสนุนการเปลี่ยนแปลง

3) แนวคิดออนโทโลยีแบบผสม (Hybrid Approach) แนวคิดนี้พยายามครอบคลุมข้อเสียของแนวคิดออนโทโลยีเดี่ยวและแนวคิดออนโทโลยีแบบหลากหลาย โดยแต่ละแหล่งข้อมูลสามารถให้รายละเอียดออนโทโลยีพื้นฐานของตนเอง และเชื่อมโยงคำศัพท์ที่ใช้ร่วมกันเป็นออนโทโลยีหลัก

ฐานความรู้เวิร์ดเน็ต (WordNet)

ฐานข้อมูลเวิร์ดเน็ต คือ ฐานข้อมูลคำศัพท์ภาษาอังกฤษจัดทำโดยมหาวิทยาลัยพรินซ์ตัน (Princeton) โดยมีการจัดแบ่งตามชนิดของคำศัพท์ คือ คำนาม (Nouns) คำกริยา (Verb) คำวิเศษ (Adverb) และคำคุณศัพท์ (Adjective) ภายในฐานข้อมูลเวิร์ดเน็ตจะมีความสัมพันธ์ที่สำคัญ เช่น ความพ้องกันเชิงความหมาย (Synonymy) โดยที่เวิร์ดเน็ตจะนิยามคำพ้องในลักษณะต่างๆ ซึ่งชุดคำพ้องแต่ละชุดนั้นจะมีหมายเลขเฉพาะกำกับ เรียกว่า synset number แต่ละ synset จะแทนลำดับชั้นของแนวคิดเพียงคำศัพท์เดียว และจะเชื่อมโยงไปยัง synset อื่นในsemantic network ด้วยหมายเลขของความสัมพันธ์ความสัมพันธ์ที่แตกต่างกันจะถูกนิยามระหว่าง synset ที่อาศัยระดับชั้นความเหมือนที่มีอยู่ ตัวอย่างเช่น คำกริยาส่วนใหญ่ถูกจัดอยู่ในส่วนคำพ้องที่เรียกว่า Troponymy คำคุณศัพท์และคำกริยาวิเศษจัดอยู่ในส่วนคำตรงข้าม (Antonymy) และคำพ้องเชิงความหมาย (Synonymy) ส่วนคำนามมีจุดเด่นของความสัมพันธ์ที่คำพ้อง (Synonymy) และคำที่เป็นลักษณะ Hyponymy หรือ Hypernymy [5]

งานวิจัยที่เกี่ยวข้อง

1) งานวิจัยของ Changqing และ Tok Wang Ling [6] ได้ทำการค้นหาความขัดแย้งเชิงความหมายของข้อมูลที่จะนำมารวมกันหรือข้อมูลที่แตกต่างกันในโครงสร้างภาษา OWL โดยเสนอวิธีการตรวจสอบความขัดแย้งในเชิงความหมายที่ทำให้เกิดความขัดแย้งทั้งหมด 7 กรณี พร้อมทั้งบอกถึงขั้นตอนวิธีการแก้ปัญหา และวิธีการบูรณาการข้อมูลร่วมกันระหว่างออนโทโลยีแบบกึ่งอัตโนมัติ งานวิจัยนี้เน้นวิเคราะห์หาความขัดแย้งของข้อมูลจากโครงสร้างและคุณสมบัติของภาษา OWL แต่ยังขาดการวิเคราะห์ข้อมูลเชิงความหมายที่แท้จริงระหว่างคอนเซ็บหรือคลาส

2) งานวิจัยของ Suwanmanee และคณะ [7] มีกระบวนการในการเชื่อมโยงออนโทโลยีที่แตกต่างกันระหว่างสองออนโทโลยี โดยใช้คุณสมบัติของภาษา OWL ในการนิยามออนโทโลยีเพื่อใช้ในการอธิบายความสัมพันธ์ระหว่างคลาสและคุณสมบัติของคลาส โดยแบ่งการเชื่อมโยงออนโทโลยีเป็น 2 รูปแบบ คือ เชื่อมโยงระหว่างคลาสและเชื่อมโยงระหว่างคุณสมบัติของคลาส งานวิจัยนี้ทำการพิจารณาการเชื่อมระหว่างคลาสและคุณสมบัติของคลาส โดยอาศัยโครงสร้างภาษา OWL เข้ามาพิจารณา แต่ยังขาดการพิจารณาถึงความหมายของคลาสและคุณสมบัติของคลาส

3) งานวิจัยของ Corley และ Mihalcea [8] เป็นการวัดความคล้ายคลึงกันระหว่างข้อความโดยใช้ฐานข้อมูลเวิร์ดเน็ต โดยผู้วิจัยกล่าวถึงสมการที่มีรูปแบบการหาค่าความคล้ายคลึงโดยใช้เวิร์ดเน็ตหลายวิธี และมีสมการของ Wu and Palmer เป็นวิธีการหนึ่งในการวัดความคล้ายคลึงแบบเมตริก

4) งานวิจัยของ วรวิทย์ สังฆทิพย์และคณะ [9] ได้เสนอวิธีการการบูรณาการออนโทโลยีด้วยเทคนิคเชื่อมโยงออนโทโลยี (Mapping Ontology) โดยใช้สมการของ Wu and Palmer ในการหาค่าความคล้ายคลึงเชิงความหมายระหว่างคลาสและคุณสมบัติของคลาส และทำการวัดประสิทธิภาพความถูกต้องของการจับคู่ความสัมพันธ์ของคลาสและคุณสมบัติของคลาส แต่ยังไม่ได้ทำการทดลองในส่วนของการสืบค้นข้อมูลที่ได้จากการจับคู่ความสัมพันธ์ของออนโทโลยี

การดำเนินการวิจัย

1) วิเคราะห์ข้อมูล

ได้ทำการศึกษาข้อมูลการบริหารงานกิจการนิสิตของคณะวิทยาการสารสนเทศ และคณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม เป็นกรณีศึกษา โดยทำการวิเคราะห์ข้อมูลทุนการศึกษาและโครงการกิจกรรมนิสิตทั้งสองคณะ พบว่ามีการเก็บข้อมูลที่ที่แตกต่างกัน เช่น การนิยามศัพท์ที่แตกต่างกันแต่มีความหมายเหมือนกัน การนิยามศัพท์เหมือนกันแต่มีความหมายที่แตกต่างกัน ทำให้เกิดความขัดแย้งของข้อมูลเกิดขึ้นเป็นต้น

2) การออกแบบสถาปัตยกรรมการบูรณาการข้อมูลออนโทโลยี

การออกแบบสถาปัตยกรรมเพื่อบูรณาการข้อมูลกิจการนิสิตได้ยึดหลักการออกแบบสถาปัตยกรรมแบบผสม [4] เพื่อให้รองรับแหล่งข้อมูลที่มีความหลากหลายของงานกิจการนิสิตแต่ละคณะที่เพิ่มขึ้นภายในมหาวิทยาลัยมหาสารคามจึงได้ออกแบบกระบวนการเป็นระดับชั้น ดัง Figure 1

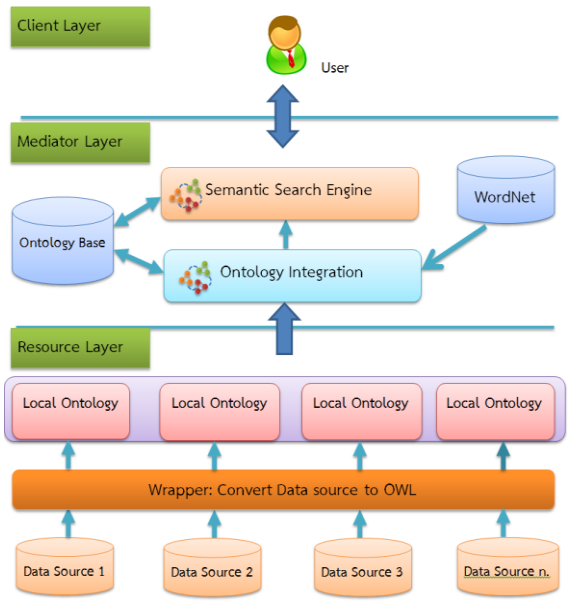

Figure 1 Ontology-based Architecture to Integrate Information for Student Affair

จาก Figure1 การออกแบบสถาปัตยกรรมการบูรณาการออนโทโลยีการบริหารงานกิจการนิสิตโดยมุ่งเน้นการบูรณาการข้อมูลกิจการนิสิตเข้าด้วยกันของแต่ละคณะ ซึ่งแบ่งได้เป็นสามระดับ คือ ระดับที่หนึ่งระดับแหล่งข้อมูล ระดับที่สองระดับสื่อกลาง และระดับที่สามระดับการนำเสนอ โดยสามารถอธิบายรายละเอียดของแต่ละระดับได้ ดังนี้

2.1) ระดับแหล่งข้อมูล (Resource Layer) เป็นระดับชั้นที่ทำหน้าที่ในการรับเข้าข้อมูลกิจการนิสิตจากทั้งสองคณะ โดยแต่ละคณะสามารถสร้างระบบสารสนเทศขึ้นเองได้โดยไม่ได้จำกัดรูปแบบแหล่งข้อมูล ซึ่งแหล่งข้อมูลอาจจะอยู่ในรูปแบบของฐานข้อมูลเชิงสัมพันธ์ หรืออยู่ในรูปแบบของไฟล์ Extensible Markup Language (xml) โดยในชั้นนี้จะมีขั้นตอนการแปลงข้อมูลรูปแบบต่างๆ ให้อยู่ในรูปแบบของออนโทโลยีที่นิยามด้วยภาษา OWLเท่านั้น ภายในระดับนี้มีส่วนของการทำงานอยู่สามส่วน ดังนี้

2.1.1) Data Source คือ ส่วนที่เป็นแหล่งข้อมูลที่มีโครงสร้างข้อมูลที่แตกต่างกัน รูปแบบการจัดเก็บข้อมูลแตกต่างกัน โดยแต่ละแหล่งข้อมูลถูกออกแบบโครงสร้างข้อมูลมาให้เหมาะสมกับการทำงานของหน่วยงานนั้นๆ

2.1.2) Wrapper คือ ส่วนที่ทำหน้าที่ในการแปลงข้อมูลแต่ละแหล่งข้อมูลที่มีความหลากหลาย ให้อยู่ในรูปแบบของออนโทโลยีที่นิยามด้วยภาษา OWL เพื่อที่จะนำไปบูรณาการข้อมูลออนโทโลยีในระดับชั้นต่อไปได้

2.1.3) Local Ontology คือ เป็นออนโทโลยีพื้นฐานที่ผ่านขั้นตอนการแปลงรูปแบบข้อมูลจากแหล่งข้อมูลที่มีความหลากหลายจากชั้น Wrapper ให้เป็นออนโทโลยีพื้นฐานที่นิยามด้วยภาษา OWL แต่ยังคงเป็นออนโทโลยีที่มีโครงสร้างพื้นฐานเฉพาะตามแหล่งข้อมูลนั้นๆ

2.2) ระดับสื่อกลาง (Mediator Layer) เป็นระดับที่ทำหน้าที่ในการบูรณาการออนโทโลยีที่มาจากระดับแหล่งข้อมูล และเป็นส่วนการรับข้อมูลจากระดับการใช้งาน (Client Layer) ภายในระดับนี้มีส่วนของการทำงานอยู่สามส่วน ดังนี้

2.2.1) Ontology Integration คือ ส่วนที่ทำหน้าที่ในการบูรณาการออนโทโลยีเข้าด้วยกันโดยใช้วิธีการเชื่อมโยงความสัมพันธ์เชิงความหมาย ซึ่งข้อมูลที่นำมาใช้ในการเชื่อมโยงเป็นข้อมูลที่ได้มาจากระดับแหล่งข้อมูลและออนโทโลยีจะต้องอยู่ในรูปแบบภาษา OWL ในส่วนของการเชื่อมโยงนั้นจะต้องหาความคล้ายคลึงเชิงความหมายของศัพท์ที่นิยามภายในออนโทโลยี โดยอาศัยฐานข้อมูลคำศัพท์เข้าช่วยในการหาความสัมพันธ์เชิงความหมายของคลาสต่างๆ ในออนโทโลยี หลังจากทำการเชื่อมโยงข้อมูลออนโทโลยีเข้าด้วยกันแล้ว ข้อมูลจะถูกนำไปเก็บไว้ที่ฐานออนโทโลยีโดยการเก็บข้อมูลจะมีสองกรณี กรณีแรกคือ ถ้าไม่มีข้อมูลออนโทโลยีใดๆ อยู่ภายในฐานออนโทโลยีจะทำการเก็บข้อมูลใส่ในฐานออนโทโลยีแต่ถ้ามีข้อมูลออนโทโลยีอยู่ภายในฐานออนโทโลยีอยู่แล้วจะทำการเชื่อมโยงข้อมูลออนโทโลยีภายในฐานออน

โทโลยีกับข้อมูลออนโทโลยีเข้ามาใหม่เข้าด้วยกัน โดยจะใช้ข้อมูลออนโทโลยีที่อยู่ภายในฐานออนโทโลยีเป็นตัวเก็บข้อมูลเชื่อมโยง

2.2.2) Ontology Base คือ ส่วนของฐานเก็บออนโทโลยีหลังจากที่บูรณาการออนโทโลยีเข้าด้วยกันแล้ว และยังมีหน้าที่ในการส่งข้อมูลระหว่าง semantic search engine อีกด้วย

2.2.3) Semantic Search Engine คือ ส่วนที่รับข้อมูลมาจากระดับการใช้งาน (Client Layer) แล้วนำข้อมูลไปประมวลผลเพื่อสืบค้นภายในฐานออนโทโลยี เมื่อได้ข้อมูลที่ต้องการแล้วจะส่งข้อมูลกลับไปเพื่อทำการแสดงผลในส่วนของการสืบค้นจะสืบค้นด้วยภาษา SPARQL[10] โดยจะใช้ Pellet[11] ซึ่งเป็น Reasoned Engine ที่ช่วยในการวิเคราะห์หาความสัมพันธ์ภายในโครงสร้างของออนโทโลยี

2.3) ระดับการใช้งาน (Client Layer) ทำหน้าที่แสดงผลโต้ตอบกับผู้ใช้ในการสืบค้นข้อมูลในระบบ โดยผู้ใช้สามารถสืบค้นข้อมูลและถูกส่งไปที่ระดับสื่อกลางของส่วน Semantic Search Engine เพื่อทำการประมวลผลคำสั่งตามความต้องการของผู้ใช้ จากนั้นก็จะส่งขึ้นมาแสดงข้อมูลให้ผู้ใช้เห็นในระดับนี้

3) พัฒนาฐานความรู้ออนโทโลยี

จากผลการวิเคราะห์ข้อมูลกิจการนิสิตของทั้งสองคณะ พบว่าทั้งสองแหล่งข้อมูลมีการนิยามความหมายให้กับข้อมูลที่แตกต่างกัน ดังนั้นจึงใช้วิธีการพัฒนาออนโทโลยีโดยกำหนดนิยามความหมายของคำศัพท์ เพื่อการแปลความหมายระหว่างคำศัพท์ที่แตกต่างกันของแหล่งข้อมูลที่หลากหลายให้สอดคล้องตรงกัน

โดยได้ออกแบบออนโทโลยีตามหลักวงจรการออกแบบออนโทโลยี[12]ให้อยู่ในรูปแบบของกราฟผ่านโปรแกรม Hozo-Ontology Editor [13] และสามารถส่งออกผลลัพธ์จากโปรแกรมเป็นภาษา OWL ได้ดังแสดงใน Figure 2-5 ตามลำดับ

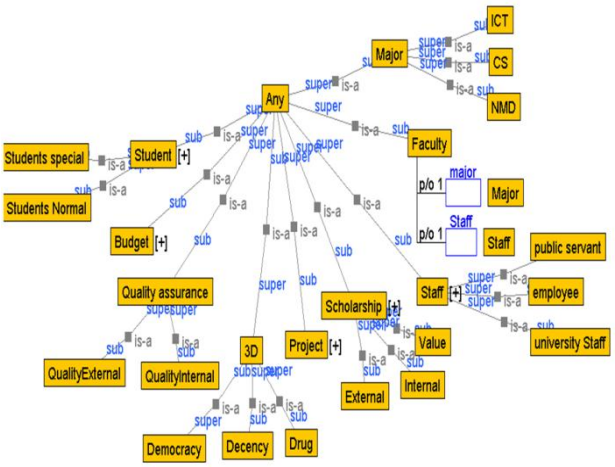

Figure 2 Ontology Student Affairs Administration for Faculty of Informatics

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDFxmlns:It="http://www.it.msu.ac.th#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#">
<rdf:Descriptionrdf:about="http://www.it.msu.ac.th#Stud
ent">
        <rdf:type>
        <rdf:Descriptionrdf:about="http://www.w3.org
/2002/07/owl#Class"/>
        </rdf:type>
        <rdfs:subClassOf>

        <rdf:Descriptionrdf:about="http://www.it.msu.
ac.th#StudentType"/>
                </rdfs:subClassOf>
```

Figure 3 Language of OWL Ontologies Student Administration for Faculty of Informatics

จาก Figure 2,3 เป็นโครงสร้างออนโทโลยีงานกิจการนิสิตของคณะวิทยาการสารสนเทศ โดยทำการจับกลุ่มข้อมูลคำศัพท์หลักๆ ที่สามารถสร้างเป็นคลาสหรือคอนเซพท์ได้ เช่น Student เป็นคลาสที่แสดงถึงข้อมูลหรือแนวคิดของนิสิต Faculty เป็นคลาสที่แสดงถึงข้อมูลหรือแนวคิดของคณะ เป็นต้น และออนโทโลยีที่อยู่ในรูปแบบของภาษา OWL ซึ่งตั้ง Prefix ของออนโทโลยีคณะวิทยาการสารสนเทศเป็น rdf:RDFxmlns:It="http://www.it.msu.ac.th#"

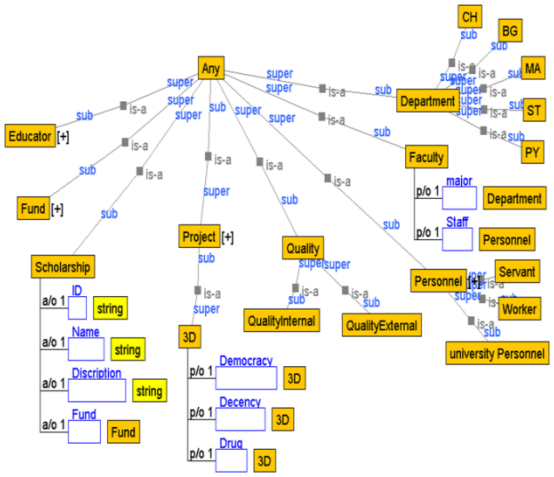

Figure 4 Ontology Student Affairs Administration for Faculty of Science

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDFxmlns:lt="http://www. science.msu.ac.th#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#">
<rdf:Descriptionrdf:about="http://www.science.msu.ac
.th#Educator">
<rdf:type>
<rdf:Descriptionrdf:about="http://www.w3.org/2002/07
/owl#Class"/>
</rdf:type>
<rdfs:subClassOf>
        <rdf:Descriptionrdf:about="http://www.
        science.msu.ac.th#EducatorType"/>
</rdfs:subClassOf>
```

Figure 5 Language of OWL Ontology Student Administration for Faculty of Science

จาก Figure 4,5 เป็นโครงสร้างออนโทโลยีงานกิจการนิสิตของคณะวิทยาศาสตร์ โดยทำการจับกลุ่มข้อมูลคำศัพท์หลักๆ ที่สามารถสร้างเป็นคลาสหรือคอนเซพท์ได้ เช่น Educator เป็นคลาสที่แสดงถึงข้อมูลหรือแนวคิดของนิสิต Project เป็นคลาสที่แสดงถึงข้อมูลหรือแนวคิดของโครงการเป็นต้นและออนโทโลยีที่อยู่ในรูปแบบของภาษา OWL ซึ่งตั้ง Prefix ของออนโทโลยีคณะวิทยาศาสตร์ เป็น rdf:RDFxmlns:lt="http://www.science.msu.ac.th#"

4) การบูรณาการออนโทโลยี

เป็นขั้นตอนที่อยู่ในส่วนของ Ontology Integration Engine ในชั้นของ Mediator Layer ทำหน้าที่ในการบูรณาการออนโทโลยี ดังนั้นจึงได้ออกแบบขั้นตอนการบูรณาการออกเป็นห้าขั้นตอน ดังแสดงใน Figure 6 สามารถอธิบายรายละเอียดของแต่ละขั้นตอนได้ ดังนี้

4.1) Input Ontology เป็นการนำเข้าออนโทโลยีที่อยู่ในรูปแบบของภาษาOWL เพื่อเข้าสู่ขั้นตอนถัดไป

4.2) Ontology Classification เป็นการจำแนกโครงสร้างของออนโทโลยีเพื่อให้เห็นคลาสและพร็อพเพอร์ตี้

4.3) Ontology Analysisเป็นขั้นตอนของการวิเคราะห์ออนโทโลยีซึ่งจะได้ค่าความลึก (depth) ค่าความสัมพันธ์ของคลาสแม่ (SuperClass) ค่าความสัมพันธ์ของคลาสลูก (Sub Class) ค่า Synset ที่เก็บกลุ่มคำเหมือนเชิงความหมายที่ได้จากฐานข้อมูลเวิร์ดเน็ต

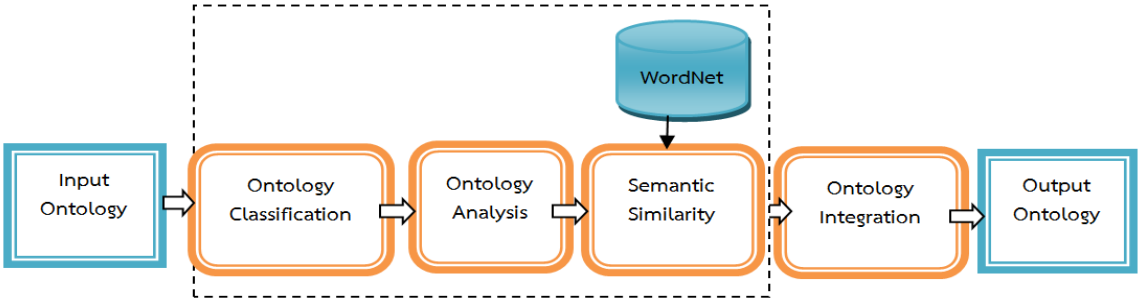

Figure 6 Ontology-Based Data Integration Method

4.4) Semantic Similarity เป็นขั้นตอนการหาความคล้ายคลึงเชิงความหมายระหว่างสองออนโทโลยี ทำได้โดยนำคอนเซ็ปหรือคลาสจากออนโทโลยีทั้งสองแหล่งข้อมูลมาเปรียบเทียบหาค่าความคล้ายคลึงเชิงความหมาย รวมไปถึงเปรียบเทียบคู่พร็อพเพอร์ตี้จากทั้งสองออนโทโลยีด้วย

การวิจัยนี้ได้ใช้สมการของ Wu and Palmer (wup) ซึ่งเป็นสมการที่หาค่าความคล้ายคลึงแบบเมตริก โดยสมการจะหาค่าความคล้ายคลึงระหว่างแต่ละคลาสของออนโทโลยีที่หนึ่งกับแต่ละคลาสของออนโทโลยีที่สองโดยใช้เวิร์ดเน็ตเพื่อหาความสัมพันธ์ ซึ่งจะพิจารณาความลึกของสอง synset ภายในโครงสร้าง Taxonomy ของเวิร์ดเน็ต

$$Sim(c1,c2) = \frac{2*depth(LCS)}{depth(concept1)+depth(concept2)} \quad (1)$$

จากสมการที่ (1) ค่า depth คือระยะทางจากคอนเซ็ปไปยัง root ของโครงสร้างลำดับชั้นของเวิร์ดเน็ต

lcs (Least Common Subsumer) โดย lcs(c1,c2) หมายถึง คอนเซ็ปร่วมที่เป็น subsumer ของคอนเซ็ป c1 และ c2 ที่อยู่ใกล้กับ c1 และ c2 มากที่สุด

ค่าความคล้ายคลึงจะมีค่าเท่ากับ 0<Simwup<=1 เนื่องจากค่าความลึกที่น้อยที่สุด (LCS) ไม่เท่ากับศูนย์ และถ้าค่า Sim(c1,c2) เท่ากับหนึ่ง แสดงว่าคอนเซ็ป 1 และ c2 นั้นจะอยู่ใน synset เดียวกัน ซึ่งมีความหมายเหมือนกันแม้จะใช้คำที่ต่างกันก็ตาม

เมื่อคำนวณความคล้ายคลึงเชิงความหมายระหว่างคลาสของทั้งสองออนโทโลยี แสดงให้เห็นว่ามีคลาสที่แตกต่างกันแต่มีความคล้ายคลึงกันเชิงความหมาย เมื่อต้องการนำมาเชื่อมโยงกันระหว่างคลาสก็จะอาศัยคุณสมบัติของภาษา OWL เข้ามาช่วยแก้ปัญหาความขัดแย้งเชิงความหมาย ได้แก่ owl:equivalentClassและ owl:equivalentProperty โดยแสดงใน Table 1,2

Table 1 the similarity between the two classes of ontology

C1	C2	Sim(c1,c2)
Project	Project	1
Student	Educator	0.66
Culture	Culture	1
Staff	Personnel	0.93
Employee	Worker	0.92
Scholarship	Scholarship	1
Major	Department	0.55
Staff	Lecturer	0.50

Table 2 the similarity between the two Properties of ontology

C1	C2	Sim(c1,c2)
name	title	0.93
address	address	1
procedures	description	0.75
schedule	fund	0.3
student_id	student_id	1
Amount	Capital	0.85
Event	Place	0.8

4.5) Ontology Integrationเป็นการเชื่อมโยงระหว่างคลาสจะพิจารณาจากค่าคำนวณความคล้ายคลึงของคลาสจากทั้งสองออนโทโลยีว่ามีความคล้ายคลึงกันหรือไม่ ถ้ามีความคล้ายคลึงจะทำการเชื่อมโยงโดยใช้คุณสมบัติการเท่ากันของภาษา OWL(owl:equivalentClass) เช่น คลาส Staff ในออนโทโลยีที่หนึ่งมีความคล้ายคลึงกับคลาส Personnel ในออนโทโลยีที่สองดังแสดงใน Figure 7

```
<owl:Class rdf:ID="Staff">
<rdfs:subClassOfrdf:resource="#Faculty"/>
<owl:equivalentClassrdf:resource="#Personnel
"/>
</owl:Class>
```

Figure 7 Ontology Mapping based on Concept

การเชื่อมโยงระหว่างคลาสจะพิจารณาจากค่าคำนวณความคล้ายคลึงของคลาสจากทั้งสองออนโทโลยีว่ามีความคล้ายคลึงกันหรือไม่ ถ้ามีความคล้ายคลึงจะทำการเชื่อมโยงโดยใช้คุณสมบัติการเท่ากันของภาษา OWL(owl:equivalentProperty) หมายถึง ความคล้ายคลึงระหว่างพร็อพเพอร์ตี้ เช่น คุณสมบัติ procedures อยู่ในออนโทโลยีที่หนึ่งมีความคล้ายคลึงกับ คุณสมบัติ description อยู่ในออนโทโลยีที่สอง ดังแสดงใน Figure 8

```
<owl:DatatypePropertyrdf:ID="procedures">
<owl:equivalentPropertyrdf:resource="#schol
arship"/>
<rdfs:domainrdf:resource="#Subject"/>
<rdfs:rangerdf:resource=" #string"/>
</owl:DatatypeProperty>
```

Figure 8 Ontology Mapping based on Property

5) วัดประสิทธิภาพการสืบค้น

เมื่อออนโทโลยีจากแหล่งข้อมูลที่หลากหลายผ่านขั้นตอนการเชื่อมโยงออนโทโลยีเข้าด้วยกันแล้ว หลังจากนั้นจะถูกเก็บไว้ในฐานออนโทโลยี ซึ่งจะสามารถวัดประสิทธิภาพด้วยวิธีการสืบค้นข้อมูลเชิงความหมายด้วยภาษา SPARQL ในการสืบค้นจะใช้คำค้นที่มีข้อมูลอยู่ในออนโทโลยีที่หนึ่งแต่ระบบจะสามารถสืบค้นข้อมูลจากออนโทโลยีที่สองขึ้นมาแสดงได้ด้วย เนื่องจากออนโทโลยีถูกเชื่อมโยงข้อมูลเชิงความหมายและเชิงโครงสร้างเข้าด้วยกันไว้แล้ว ซึ่งการวิจัยนี้ใช้คำสำคัญที่แตกต่างกันในการสืบค้น และใช้ค่า Recall ดังสมการที่ (2) ค่า Precision ดังสมการที่ (3) และ F-Measure ดังสมการที่ (4) สามารถแสดงสมการได้ ดังนี้

A คือ จำนวนข้อมูลที่สืบค้นได้และถูกต้องตามความสนใจ

B คือ จำนวนข้อมูลที่ถูกต้องตามความสนใจแต่ไม่ถูกสืบค้น

C คือ จำนวนข้อมูลที่สืบค้นได้แต่ไม่ถูกต้องตามความสนใจ

$$Recall = \frac{A}{A+B} \; x \; 100\% \qquad (2)$$

$$Precisions = \frac{A}{A+C} \; x \; 100\% \qquad (3)$$

$$F\text{-}measure = 2\left(\frac{Precision*Recall}{Precision+Recall}\right) \qquad (4)$$

จากการทดลองสืบค้นข้อมูล ได้ผลการวัดประสิทธิภาพได้ค่า Recall เท่ากับ 94% ค่า Precision 83% ส่งผลให้ได้ F-Measure เท่ากับ 88% ซึ่งการทดลองในครั้งนี้ทำให้ทราบถึงประสิทธิภาพในการสืบค้นอยู่ในระดับที่ดี แสดงให้เห็นถึงความถูกต้องของการบูรณาการข้อมูลออนโทโลยีที่มีความหลากหลายทั้งเชิงความหมายและเชิงโครงสร้าง

ผลการวิจัยและสรุปผล

งานวิจัยนี้ได้ออกแบบสถาปัตยกรรมออนโทโลยีเพื่อบูรณาการงานกิจการนิสิต โดยออกแบบเป็นแบบผสมและทำการแบ่งระดับชั้นภายในสถาปัตยกรรมออกเป็นสามระดับเพื่อให้รองรับแหล่งข้อมูลที่มีความหลากหลายโดยนิยามความรู้ผ่านการสร้างออนโทโลยีด้วยภาษา OWL จากนั้นนำออนโทโลยีที่มีความแตกต่างกันเชิงความหมายและเชิงโครงสร้างมาบูรณาการเข้าด้วยกันตามสถาปัตยกรรมของระบบที่ออกแบบไว้ ซึ่งการบูรณาการใช้วิธีการวัดความคล้ายคลึงเชิงความหมายระหว่างออนโทโลยีด้วยสมการของ Wu and Palmer เพื่อแก้ปัญหาความขัดแย้งในเชิงความหมาย และนำฐานข้อมูลเวิร์ดเน็ตเข้ามาหาความสัมพันธ์เชิงความหมาย ซึ่งการบูรณาการออนโทโลยีได้พิจารณาถึงระดับคลาสและคุณสมบัติของคลาส เพื่อให้การสืบค้นมีประสิทธิภาพมากยิ่งขึ้น และในส่วนของการสืบค้นข้อมูลเลือกใช้ภาษา SPARQL เป็นภาษาสำหรับสืบค้นข้อมูลในออนโทโลยี

จากการทดลองบูรณาการออนโทโลยีการบริหารงานกิจการนิสิตจากสองแหล่งข้อมูล จากนั้นทำการวัดประสิทธิภาพการบูรณาการข้อมูลจากการสืบค้นข้อมูลในออนโทโลยี โดยได้สืบค้นข้อมูลจากคำสำคัญที่แตกต่างกัน ได้ค่า Recall ค่า Precision และค่า F-measure อยู่ที่ร้อยละ 94 ร้อยละ 83 และร้อยละ 88 ตามลำดับ จากผลลัพธ์ดังกล่าวแสดงให้เห็นว่าการค้นหาข้อมูลมีประสิทธิภาพดีขึ้น

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยหลักสูตรเทคโนโลยีสารสนเทศคณะวิทยาการสารสนเทศประจำปีงบประมาณ พ.ศ. 2555 และขอขอบพระคุณฝ่ายกิจการนิสิตคณะวิทยาการสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม ที่อนุเคราะห์ข้อมูลเพื่อศึกษาวิจัยในครั้งนี้

เอกสารอ้างอิง

[1] Uschold M, Grüninger M. Ontologies: Principles, methods and applications. Knowledge Engineering Review 1996 June;11:93-136.

[2] W3C. Web Ontology Language (OWL). [serial online]. 2007 September; Available from : http://www.w3.org/2004/OWL/. Accessed August 13, 2013

[3] Li L, Huang L, Guan Q, Xu D. Research on Tree Segmentation-based Ontology Mapping. IEEE Computer Society; 2009. p. 89-92.

[4] WacheH, Vögele T, Visser U, Stuckenschmidt H, Schuster G, Neumann H, et al. Ontology-Based Integration of Information - A Survey of Existing Approaches.Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing; 2001. p. 108-117.

[5] WordNet A lexical database for English . What is WordNet [serial online]. 2012 December; Available from : http://wordnet.princeton.edu/wordnet/. Accessed August 13, 2013

[6] Changqing Li, Tok Wang Ling. OWL-Based Semantic Conflicts Detection and Resolution for Data Interoperability. Proceeding of International Conference on Entity Relational Approach Workshops; 2004 p. 266-277

[7] Suwanmanee S, Benslimane D, Thiran P. OWL-Based approach for Semantic Interoperability. International Conference on Advanced Information Networking and Applications. IEEE Computer Society: 2005. p. 145-50.

[8] Corley C, Mihalcea R. Measuring the semantic similarity of texts. Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment; 2005. p. 13-8.

[9] วรวิทย์ สังฆทิพย์, จิรัฏฐา ภูบุญอบ, ฉัตรตระกูล สมบัติธีระ. การเชื่อมโยงออนโทโลยีการบริหารงานกิจการนิสิตด้วยวิธีการวัดความคล้ายคลึงเชิงความหมาย. ใน: เอกสารการประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ ครั้งที่ 4. มหาวิทยาลัยเทคโนโลยีมหานคร; 2555. หน้า 443-7.

[10] W3C. SPARQL Query Language for RDF. [serial online]. 2008January; Available from :http://www.w3.org/TR/rdf-sparql-query/.Accessed August 13, 2013

[11] Parsia C. Pellet: OWL 2 Reasoner for Java. [serial online]. Available from :http://clarkparsia.com/pellet/.Accessed August 13, 2013

[12] Noy NF, mcguinness DL. Ontology Development 101: A Guide to Creating Your First Ontology. [serial online]. Available from :http://protege. stanford.edu/publications/ontology_development/ontology101.pdf.Accessed August 13, 2013

[13] Kozaki K, Kitamura Y, Ikeda M, Mizoguchi R. Hozo: An Environment for Building/Using Ontologies Based on aFundamental Consideration of “Role” and “Relationship”.Springer Berlin Heidelberg; 2002. p. 155-163.