

การจัดกลุ่มเว็บเซอร์วิสเพื่อการสืบค้นเชิงความหมาย Web Service Clustering for Semantic Web Service Discovery

จตุพล โพธิ์คำพอก (Jatupon Pocompok)^{1*} ดร.สมจิตร อาจอินทร์ (Dr.Somjit Arch-int)**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบสถาปัตยกรรมการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย โดยใช้หลักวิธีการประมวลผลภาษาธรรมชาติ (natural language processing) มาวิเคราะห์การสอดคล้องกันเชิงความหมายของเว็บเซอร์วิสโดยใช้ขั้นตอนวิธี ONION ร่วมกับพจนานุกรมเวิร์ดเน็ตในการเชื่อมโยงเว็บเซอร์วิสกับโดเมนออนโทโลยี แล้วทำการวัดระดับความสัมพันธ์เชิงแนวคิดของเว็บเซอร์วิส (Degree of Match) เพื่อหาค่าคะแนนความเหมือนระหว่างเว็บเซอร์วิส ดังนั้นเว็บเซอร์วิสที่มีค่าคะแนนความเหมือนมากที่สุดจะถูกจัดให้อยู่กลุ่มเดียวกัน ผลการวิจัยพบว่ากระบวนการเชื่อมโยงโดเมนออนโทโลยีทำให้โดเมนออนโทโลยีเกิดการเรียนรู้และทำให้การจัดกลุ่มมีประสิทธิภาพมากยิ่งขึ้น

ABSTRACT

The purpose of this research was to design architecture of semantic web service grouping by using method of natural language processing to analyze the harmony semantically of web service. The process was using ONION with the words net dictionary to connect web service and domain ontology. After that it was measured the conceptual relationship of web service to find the degree of matching, therefore web service, which had the most degree of matching, was classified in the same group. The research's result found that connection process of domain ontology was caused learning process of domain ontology and caused better capacity of grouping.

คำสำคัญ : เว็บเซอร์วิสเชิงความหมาย การจัดกลุ่ม โดเมนออนโทโลยี

Key Words : Semantic web service, Clustering, Domain ontology

¹ Correspondent author : jatuponp@gmail.com

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

** รองศาสตราจารย์ ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

บทนำ

การพัฒนาด้านเทคโนโลยีสารสนเทศปัจจุบันเจริญรุดหน้าไปอย่างรวดเร็ว แต่ละวันข้อมูลข่าวสารถูกผลิตและเผยแพร่ผ่านเครือข่ายอินเทอร์เน็ตเป็นจำนวนมาก โดยเฉพาะเว็บเซอร์วิส (Web Service) ที่ผลิตเพื่อสนับสนุนการทำงานระหว่างคอมพิวเตอร์กับคอมพิวเตอร์ที่มีระบบสารสนเทศแตกต่างกันให้สามารถทำงานร่วมกันอย่างอัตโนมัติ นักพัฒนาระบบต้องใช้เวลามากขึ้นในการคัดเลือกเว็บเซอร์วิสที่การทำงานสอดคล้องหรือตรงตามความต้องการมากที่สุด ทำให้เกิดปัญหาด้านการสืบค้นที่นักพัฒนาระบบได้เว็บเซอร์วิสที่ไม่ตรงตามความต้องการ ประกอบกับที่ผ่านมามีการสืบค้นและการจัดหมวดหมู่ใช้หลักการพื้นฐานของการเปรียบเทียบคำหลัก (Keyword based) เป็นส่วนใหญ่ ซึ่งไม่ได้ให้ความสำคัญกับการพิจารณาความหมายของเว็บเซอร์วิสเท่าที่ควร ปัจจุบันมีการค้นคว้าวิจัยอย่างกว้างขวางเพื่อพัฒนาระบบการสืบค้นให้มีประสิทธิภาพทั้งในด้านการสืบค้นและด้านการจัดกลุ่มเว็บเซอร์วิส

งานวิจัยนี้ได้นำเสนอแนวทางหนึ่งสำหรับการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย (Semantic web service) [8] โดยประยุกต์ใช้ความรู้ด้านการประมวลผลภาษาธรรมชาติ (natural language processing) เพื่อวิเคราะห์การสอดคล้องกันเชิงความหมายของเว็บเซอร์วิส และในงานวิจัยนี้ยังใช้ความรู้ด้านการวัดระดับความสัมพันธ์เชิงแนวคิด (Degree of Match) เพื่อวัดความเหมือนระหว่างเว็บเซอร์วิสสำหรับจัดกลุ่มเว็บเซอร์วิสตามความหมาย

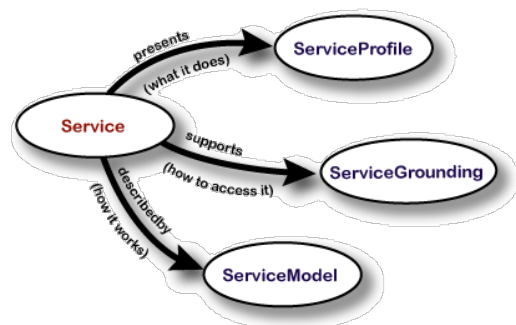
วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์ เพื่อออกแบบสถาปัตยกรรมการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย (Semantic Web Service Clustering) ด้วยวิธีเว็บเซอร์วิสแบบสื่อความหมาย เพื่อสร้างฐานความรู้ที่เหมาะสมสำหรับการสืบค้น และช่วยให้เพิ่มประสิทธิภาพการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. เว็บเซอร์วิสเชิงความหมาย (Semantic Web Service) คือเว็บเซอร์วิสที่มีลักษณะการอธิบายความหมายหรือคำจำกัดความของข้อมูลไว้ รวมถึงการกำหนดโครงสร้างของข้อมูล ทำให้คอมพิวเตอร์สามารถเข้าใจข้อมูลที่สอดคล้องกับความเข้าใจของมนุษย์ ทำให้สามารถรองรับการทำงานแบบอัตโนมัติ เช่น ภาษา OWL-S (Ontology Web Language for Web Service)

ภาษา OWL-S เป็นภาษาที่พัฒนาต่อมาจากภาษา OWL พัฒนาขึ้นมาเพื่อสนับสนุนการอธิบายเว็บเซอร์วิสแบบสื่อความหมายมีองค์ประกอบ 3 ส่วนคือ ServiceProfile, ServiceModel และ ServiceGrounding ดังภาพที่ 1



ภาพที่ 1 องค์ประกอบของภาษา OWL-S [8]

2. ออนโทโลยี (Ontology) เป็นแนวคิดที่ปรากฏอยู่ในหลายสาขาวิชา ในทางคอมพิวเตอร์แล้วได้ให้คำจำกัดความของออนโทโลยีไว้หลากหลาย แต่คำจำกัดความที่ได้รับการยอมรับกันอย่างแพร่หลายคือ “ข้อกำหนดเกี่ยวกับแนวคิด (The Specification of a conceptualization)” [1] แนวความคิดของออนโทโลยีคือการอธิบายองค์ความรู้อย่างมีขอบเขตและความสัมพันธ์ของข้อมูลตามสิ่งที่สนใจ ซึ่งสามารถแสดงออกมาในรูปของระบบสัญลักษณ์ (Notation) เช่น คลาส (Class) อินสแตนซ์ (Instance) ความสัมพันธ์ (Relation) คุณสมบัติ

(Property) และกฎ (Rule) โดยใช้ภาษาสำหรับแสดงความรู้ (Knowledge Representation Language) ซึ่งมีความชัดเจนและเที่ยงตรงมากกว่าการอธิบายถึงสิ่งของในแง่มุมต่างๆ ทั้งนี้การใช้ระบบสัญลักษณ์จะช่วยสื่อความหมาย (Semantics) ให้ซอฟต์แวร์หรือคอมพิวเตอร์เข้าใจและนำไปประมวลผลต่อไป

3. พจนานุกรมเวิร์ดเน็ต (WordNet Thesaurus) [5] คือฐานข้อมูลคำศัพท์ภาษาอังกฤษที่มีความสัมพันธ์ระหว่างคำศัพท์ซึ่งเป็นคำศัพท์ทั่วไปไม่เฉพาะเจาะจง ฐานความรู้เวิร์ดเน็ตมีการจัดรวบรวมข้อมูลคำศัพท์ตามเกณฑ์ทางความหมายมากกว่าจะดูจากรูปคำโดยอาศัยความสัมพันธ์ทางความหมายของคำ ฐานความรู้เวิร์ดเน็ตมีการจัดแบ่งชนิดของคำออกเป็น 4 ประเภท ได้แก่ คำนาม (Nouns) คำกริยา (Verbs) คำวิเศษณ์ (Adverbs) และคำคุณศัพท์ (Adjectives)

4. ขั้นตอนวิธีของ ONION [6] เป็นขั้นตอนวิธีในการหาค่าคะแนนความคล้ายคลึงระหว่างคำโดยอาศัยพจนานุกรมเวิร์ดเน็ตช่วยหาความสัมพันธ์ทางความหมายของคำ โดยมีการพิจารณา 3 เงื่อนไข เงื่อนไขแรกพิจารณาว่าคำทั้งสองคำนั้น เป็นคำเดียวกันหรือไม่ ซึ่งหากเป็นคำเดียวกันจะมีค่าคะแนนความคล้ายคลึงเท่ากับ 1 เงื่อนไขที่สอง พิจารณาคำพ้องความหมาย synonyms จะให้ค่าคะแนนความคล้ายคลึงเป็น 1 เช่นกัน เงื่อนไขที่ 3 กรณีที่ไม่เข้าสองเงื่อนไขแรก จะพิจารณาที่นิยามของคำศัพท์หรือเวิร์ดเซนซ์ โดยพิจารณาคำที่อยู่ในเวิร์ดเซนซ์ของสองคำว่ามีคำใดบ้างที่เป็นคำเดียวกัน หรือพ้องความหมายกัน ซึ่งหากเหมือนกันจะนับค่าคะแนนเป็น 1 และจะตรวจสอบทุกคู่ของคำในเวิร์ดเซนซ์จนครบแล้วหาผลรวมของคะแนน หาด้วยจำนวนคำในเวิร์ดเซนซ์ที่มีจำนวนน้อยที่สุด

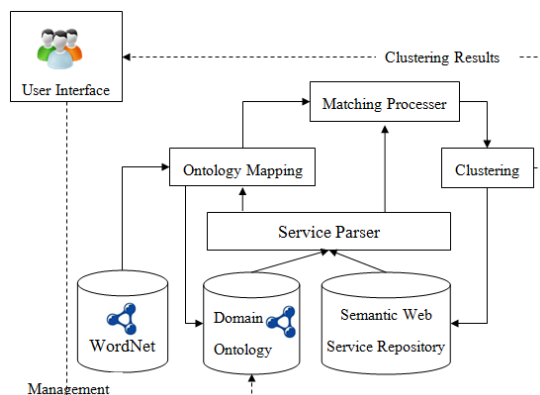
งานวิจัยที่เกี่ยวข้อง

งานวิจัย [4] นำเสนอแนวคิดในการค้นหาบริการเว็บเซอร์วิสเชิงความหมาย โดยอาศัยการจับคู่เชิงความหมาย (Semantic Match) ระหว่างเว็บเซอร์วิสที่ให้บริการกับคำร้องจากผู้ใช้งาน โดยความสามารถของเว็บเซอร์วิส (Web Services Capabilities) ซึ่งมีการพิจารณาเฉพาะข้อมูลเข้าและข้อมูลออกของเว็บเซอร์วิส ดังนั้นวิธีการนี้ผู้ใช้สามารถค้นหาได้ตามข้อมูลเข้าและข้อมูลออกเท่านั้น ในบางครั้งฟังก์ชันบนเว็บเซอร์วิสที่มีการประมวลผลต่างกันอาจมีข้อมูลเข้า (Input) หรือข้อมูลออก (Output) เหมือนกันได้ เช่นเว็บเซอร์วิส HotelBooking และเว็บเซอร์วิส BookPrice ต่างมีข้อมูลออกชนิดเดียวกัน คือ Price ทั้งที่ทั้งสองเว็บเซอร์วิสมีการทำงานที่แตกต่างกัน ซึ่งผลลัพธ์ที่ได้จากการค้นหาจะไม่ถูกต้อง เนื่องจากพิจารณาเพียงข้อมูลเข้าและข้อมูลออกเพียงอย่างเดียว ดังนั้นงานวิจัยนี้ได้เพิ่มการพิจารณาชื่อบริการ (Service Name) ทั้งนี้เพราะชื่อบริการจะบ่งบอกถึงการให้บริการของเว็บเซอร์วิส

งานวิจัย [2] ได้นำเสนอวิธีการประยุกต์ใช้ระดับการจับคู่ของเว็บเซอร์วิส (Degree of Match) เพื่อวัดค่าคะแนนความคล้ายคลึงของเว็บเซอร์วิส โดยเพิ่มการพิจารณาถึงระยะห่างระหว่างแนวคิดเชิงความหมาย (Semantic distance) มีขั้นตอนการทำงานดังนี้ ขั้นตอนแรก วัดระดับการจับคู่ของเว็บเซอร์วิส (Degree of Match) จะได้ระดับความสัมพันธ์ต่าง ๆ คือ Exact, Plugin, Subsumes และ Fail ขั้นตอนที่สอง นำผลลัพธ์ที่ได้จากขั้นตอนแรก มาวัดค่าคะแนนความคล้ายคลึงเชิงความหมายระหว่างเว็บเซอร์วิสที่นำมาพิจารณา ซึ่งงานวิจัยนี้ได้พัฒนาจากแนวคิดของงานวิจัย [2] ซึ่งมีการเพิ่มกระบวนการเชื่อมโยงบริการเว็บเซอร์วิสกับโดเมนออนโทโลยีเพื่อให้โดเมนออนโทโลยีหลักนั้นสามารถเรียนรู้ได้เองจากเว็บเซอร์วิสที่นำมาเปรียบเทียบ

วิธีการดำเนินการวิจัย

1. การออกแบบสถาปัตยกรรมการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย สามารถแสดงการทำงานของระบบ ได้ดังภาพที่ 2



ภาพที่ 2 สถาปัตยกรรมการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย

Service Parser ทำหน้าที่ในการวิเคราะห์ข้อมูลบริการของเว็บเซอร์วิสที่อยู่ในฐานข้อมูล (Semantic Web Service Repository) โดยใช้ภาษาที่สามารถค้นหบริการจากข้อกำหนดเชิงโครงสร้าง โดยการระบุค่าของอินสแตนซ์เป็นเงื่อนไขในการค้นหา และระบุคุณสมบัติของอินสแตนซ์ด้วยภาษารีดคิวเอล (RDQL)

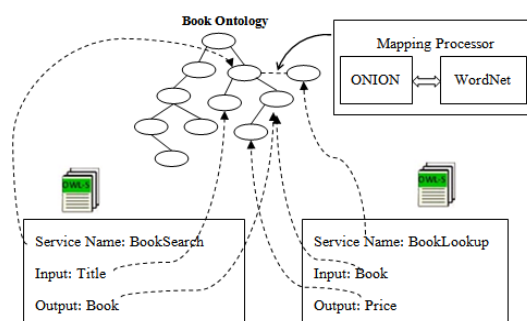
Ontology Mapping ทำหน้าที่เชื่อมโยงบริการเว็บเซอร์วิสกับโดเมนออนโทโลยีหลัก (Domain Ontology) การเชื่อมโยงออนโทโลยีนั้นใช้ขั้นตอนวิธีของ ONION เพื่อวัดค่าความคล้ายคลึงระหว่างคำเชิงความหมาย ร่วมกับพจนานุกรมเวิร์ดเน็ต (WordNet)

Matching Processor ทำหน้าที่ในการวัดระดับความสัมพันธ์ของเว็บเซอร์วิส (Degree of Match) ร่วมกับโดเมนออนโทโลยี (Domain Ontology) เมื่อได้ระดับความสัมพันธ์ระหว่างเว็บเซอร์วิสที่นำมาเปรียบเทียบแล้ว ก็ทำการวัดค่าคะแนนความคล้ายคลึงระหว่างเว็บเซอร์วิส

Clustering ทำหน้าที่ในการจัดกลุ่มเว็บเซอร์วิส โดยการนำค่าคะแนนความคล้ายคลึงระหว่างเว็บเซอร์วิสว่าเหมือนกันมากน้อยเพียงใด ถ้าเหมือนกันจะถูกจัดเป็นกลุ่มเดียวกัน

2. การเตรียมความรู้ เป็นขั้นตอนการประมวลผลเพื่อเตรียมข้อมูลบริการเว็บเซอร์วิสและโดเมนออนโทโลยีที่เหมาะสมสำหรับการนำไปเชื่อมโยงออนโทโลยีและการจับคู่เว็บเซอร์วิส เป้าหมายของขั้นตอนการเตรียมความรู้คือ การได้ข้อมูลบริการเว็บเซอร์วิสได้แก่ ชื่อเว็บเซอร์วิส (Service Name) ข้อมูลเข้า (Input) ข้อมูลออก (Output) และแนวคิดต่าง ๆ บนโดเมนออนโทโลยีทั้งหมด ขั้นตอนนี้ประกอบด้วย การคิวรี่ข้อมูล การตัดคำและการสกัดคำหลัก (Keyword Extraction)

3. กระบวนการเชื่อมโยงออนโทโลยี (Ontology Mapping Processor) ทำหน้าที่ในการเชื่อมโยงเว็บเซอร์วิสกับออนโทโลยีหลัก ด้วยการวัดค่าความคล้ายคลึงระหว่างคำเชิงความหมายโดยใช้ขั้นตอนวิธีของ ONION ร่วมกับพจนานุกรม WordNet เพื่อทำการพิจารณาความเหมือนระหว่างคำศัพท์ที่นำมาเปรียบเทียบ สามารถอธิบายได้ดังภาพที่ 3



ภาพที่ 3 แสดงแนวคิดการเชื่อมโยงออนโทโลยี

จากภาพที่ 3 แสดงแนวคิดในการเชื่อมโยงเว็บเซอร์วิสกับโดเมนออนโทโลยีโดยอาศัยขั้นตอนวิธี ONION ร่วมกับพจนานุกรมเวิร์ดเน็ต ทำการพิจารณาความคล้ายคลึง พิจารณา 3 เงื่อนไข ได้แก่

เงื่อนไขแรกพิจารณาคำทั้งสองคำนั้น เป็นคำเดียวกันหรือไม่ เช่น “Computer” และ “Computing” ถือว่าเป็นคำเดียวกัน ให้ค่าคะแนนความคล้ายคลึงเป็น 1 เงื่อนไขที่สอง พิจารณาคำพ้องความหมาย เช่น “Compute” และ “Calculate” เป็นคำพ้องความหมาย จะให้ค่าคะแนนความคล้ายคลึงเป็น 1 เช่นกัน ส่วนเงื่อนไขที่สามนั้น จะเป็นเงื่อนไขที่ไม่เข้าทั้งสองเงื่อนไขแรก โดยจะเพิ่มการพิจารณาที่นิยามของคำศัพท์หรือเวิร์ดเซนซ์ โดยพิจารณาคำที่อยู่ในเวิร์ดเซนซ์ของทั้งสองคำว่ามีคำใดบ้างที่เป็นคำเดียวกัน หรือพ้องความหมายกัน ซึ่งถ้าหากเหมือนกันจะให้ค่าคะแนนเป็น 1 และจะตรวจสอบทุกคู่ของคำในเวิร์ดเซนซ์จนครบ แล้วหาผลรวมคะแนน ทหารด้วยจำนวนคำในเวิร์ดเซนซ์ที่มีจำนวนน้อยที่สุด (ไม่นับคำที่ไม่มีความหมาย เช่น the, an, of เป็นต้น)

ตารางที่ 1 การหาค่าคะแนนความคล้ายคลึงระหว่างคำ

BookSearch	BookLookup	SimilarityScore
Book	Book	1
Book	Lookup	0
Search	Book	0
Search	Lookup	1

จากตารางที่ 1 แสดงการหาค่าความคล้ายคลึงระหว่างคำตามขั้นตอนวิธี ONION ซึ่งเป็นตัวอย่างของการเปรียบเทียบระหว่างคำว่า “BookSearch” กับคำว่า “BookLookup” โดยทำการเปรียบเทียบหาค่าคะแนนความคล้ายคลึงของแต่ละคู่คำแล้ว นำคะแนนที่ได้มาหาผลรวมและหารด้วยจำนวนคำที่น้อยที่สุด จะได้ค่าคะแนนความคล้ายคลึงเท่ากับ $(1.0+0.0+0.0+1.0)/2 = 1.0$ ดังนั้นค่าคะแนนความคล้ายคลึงระหว่าง “BookSearch” และ “BookLookup” มีค่าเป็น 1.0

ดังนั้นเมื่อทำการวัดค่าคะแนนความคล้ายคลึงระหว่าง “BookSearch” และ “BookLookup” มีค่าเท่ากับ 1.0 ซึ่งมีค่ามากกว่า 0.5

(ค่าคะแนนความคล้ายคลึงจะมีค่าระหว่าง 0.0–1.0 ดังนั้นค่าคะแนนที่เหมาะสมจึงต้องมีค่าเกินครึ่งหนึ่ง นั่นคือ 0.5) จากนั้นทำการเชื่อมโยง “BookLookup” เข้าไปในโดเมนออนโทโลยี โดยการสร้างคลาส “BookLookup” และกำหนดให้คลาส “BookLookup” มีความสัมพันธ์แบบ EquivalentClass กับคลาส “BookSearch” สามารถอธิบายได้ดังภาพที่ 4

```
<owl:Class rdf:about="#BookSearch">
  <owl:equivalentClass>
    <rdfs:Class rdf:resource="#BookLookup"/>
  <owl:equivalentClass>
</owl:Class>
```

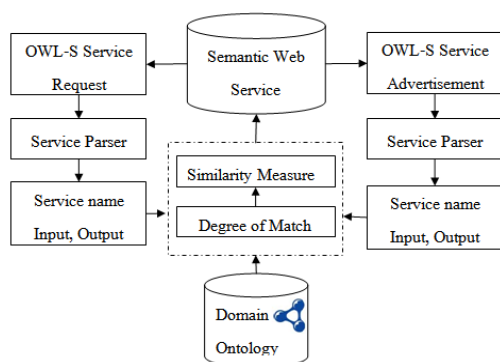
ภาพที่ 4 แสดงตัวอย่างการเพิ่มคลาส BookSearch

เมื่อได้ทำการเพิ่มคลาสเข้าไปในโดเมนออนโทโลยีแล้วจะทำให้ได้โดเมนออนโทโลยีใหม่ที่มีคลาส “BookLookup” เป็นคลาสย่อย ๆ ในโดเมนออนโทโลยี ซึ่งขั้นตอนวิธีดังกล่าวมานี้จะทำให้โดเมนออนโทโลยีเกิดการเรียนรู้ (Ontology Learning) จากข้อมูลของเว็บเซอร์วิสที่นำมาเชื่อมโยง หลังจากนั้นทำการกำหนดโดเมนออนโทโลยีให้กับเว็บเซอร์วิสนั้น เพื่อนำไปใช้ในกระบวนการหาระดับความสัมพันธ์ของเว็บเซอร์วิสต่อไปได้ สามารถอธิบายได้ดังภาพที่ 5

```
<owl:Ontology rdf:about="">
  <owl:imports rdf:resource="http://www.daml.org/services/owl-s/1.1/Service.owl" />
  <owl:imports rdf:resource="http://www.daml.org/services/owl-s/1.1/Process.owl" />
  <owl:imports rdf:resource="http://www.daml.org/services/owl-s/1.1/Profile.owl" />
  <owl:imports rdf:resource="http://www.daml.org/services/owl-s/1.1/Grounding.owl" />
  <owl:imports rdf:resource="http://127.0.0.1/ontology/books.owl" />
</owl:Ontology>
```

ภาพที่ 5 การเพิ่มโดเมนออนโทโลยีให้กับเว็บเซอร์วิส

4. ขั้นตอนวิธีการจับคู่เว็บเซอร์วิส (Semantic web service matching) เป็นกระบวนการหา ระดับความสัมพันธ์ของการจับคู่เว็บเซอร์วิสที่ เปรียบเทียบ โดยงานวิจัยนี้ได้พัฒนามาบนพื้นฐาน สถาปัตยกรรมการจับคู่ของงานวิจัย Improving Grade Match to Value Match for Semantic Web Service Discovery กระบวนการจับคู่เว็บเซอร์วิส พิจารณาจาก ชื่อเว็บเซอร์วิส ข้อมูลเข้า (input) และข้อมูลออก (output) ตามลำดับ โดยมีระดับของ การจับคู่เว็บเซอร์วิส 4 ระดับ ได้แก่ Exact, Plugin, Subsume และ fail สามารถอธิบายขั้นตอนวิธีการ จับคู่เว็บเซอร์วิสได้ดังต่อไปนี้



ภาพที่ 6 กระบวนการจับคู่เว็บเซอร์วิสเชิงความหมาย

จากภาพที่ 6 เป็นกระบวนการในการจับ คู่เว็บเซอร์วิสเริ่มจากการนำเว็บเซอร์วิสที่อยู่ใน ฐานข้อมูลเว็บเซอร์วิส (Semantic Web Service Repository) มาพิจารณาเป็นคู่ ๆ หลังจากนั้น Service Parser จะทำการอนุมานข้อมูลเว็บเซอร์วิส ออกมาเป็นข้อมูลพารามิเตอร์ที่จะนำไปใช้ใน กระบวนการจับคู่ (Matching Engine) เพื่อหา ระดับความสัมพันธ์ของคู่เว็บเซอร์วิส (Degree of Match) เมื่อได้ระดับความสัมพันธ์แล้วก็นำระดับ ความสัมพันธ์ที่ได้ไปผ่านขั้นตอนการวัดค่าคะแนน ความคล้ายคลึง (Similarity Measure) จะได้ค่า คะแนนความคล้ายคลึงของเว็บเซอร์วิสที่นำมาเปรียบ

เทียบ ซึ่งสามารถอธิบายระดับความสัมพันธ์ของ เว็บเซอร์วิสได้ดังนี้

```

degreeOfMatch(outR, outA)
    if outA = outR then return exact
    if outR is a direct subclassof outA then return exact
    if outA subsumes outR and outR is not a direct subclassof outA
        then return plugin
    if outR subsumes outA then return subsumes
    else return fail
  
```

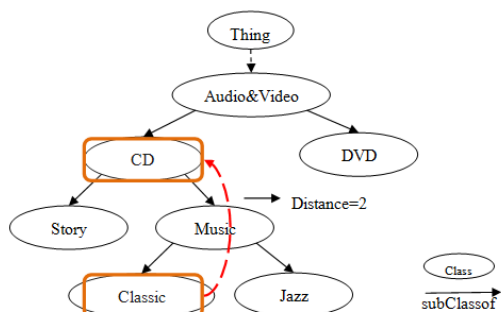
ภาพที่ 7 ระดับความสัมพันธ์ของบริการเว็บเซอร์วิส

จากภาพที่ 7 เป็นระดับความสัมพันธ์ของ เว็บเซอร์วิสสำหรับข้อมูลออก (output) ซึ่งในส่วน ของข้อมูลอื่น ๆ ก็มีกระบวนการการจับคู่เช่นเดียวกัน เพื่อหา ระดับของการจับคู่ให้สามารถนำไปวัดค่าความ คล้ายคลึงระหว่างเว็บเซอร์วิสโดยอาศัยระยะห่าง ระหว่างแนวคิดในออนโทโลยี โดยสามารถอธิบาย ระดับของการจับคู่แต่ละระดับได้ดังนี้

Exact ($outA=outR$) หรือ ($outR$ is direct SubClassof $outA$) เป็นระดับการจับคู่ เว็บเซอร์วิสที่มีความแม่นยำที่สุด ซึ่งมีระดับความ สัมพันธ์แบบสมมูลกัน หรือมีความสัมพันธ์ในแบบ SubClassof

Plug-in ($outR$ subclassof $outA$)
เมื่อ $outR$ มีความสัมพันธ์กับ $outA$ แบบ subclassof
Subsume ($outA$ subclassof $outR$)
เมื่อ $outA$ มีความสัมพันธ์กับ $outR$ แบบ subclassof
Fail คือไม่มีความสัมพันธ์กันเลย

การหา ระดับความสัมพันธ์เชิงแนวคิด ระหว่างเว็บเซอร์วิสนั้นจะอาศัยโดเมนออน โทโลยี ตัวอย่างบางส่วนของออนโทโลยีเกี่ยวกับ Audio&Video ดังภาพที่ 8



ภาพที่ 8 ตัวอย่างบางส่วนของออนโทโลยี Audio&Video

ดังนั้นเมื่อต้องการระดับความสัมพันธ์ของบริการเว็บเซอร์วิสใด ๆ ที่นำมาเปรียบเทียบ (Service request) ที่มีข้อมูลเข้า คือ year และมีข้อมูลออกคือ computer book, classic music และ dvd ผลลัพธ์ที่ได้จากการหาระดับความสัมพันธ์ของเว็บเซอร์วิสแสดงดังตารางที่ 2 และตารางที่ 3 ตามลำดับ

ตารางที่ 2 ผลการหาระดับความสัมพันธ์ของเว็บเซอร์วิส

Advertisement	Request	Degree of Match
Book&CDSelling Service	RequestBook&CD&DVDsellingService	Exact

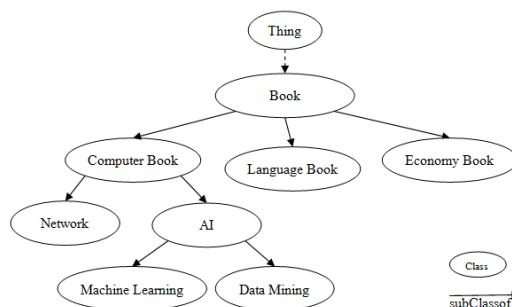
ตารางที่ 3 ผลการหาระดับความสัมพันธ์ของข้อมูลเข้า

Advertisement	Request	Degree of Match
Year	Year	Exact

ตารางที่ 4 ผลการหาระดับความสัมพันธ์ของข้อมูลออก

Advertisement	Request	Degree of Match
Computer Book	Book	Exact
Computer Book	CD	Fail
Classic Music	Book	Fail
Classic Music	CD	Subsume
DVD	Book	Fail
DVD	CD	Fail

จากตารางที่ 4 เป็นผลลัพธ์ที่ได้จากการหาระดับความสัมพันธ์ระหว่างเว็บเซอร์วิสโดยอาศัยออนโทโลยีหนังสือช่วยในการหาระดับความสัมพันธ์เพิ่มเติม ดังแสดงในภาพที่ 9



ภาพที่ 9 ตัวอย่างบางส่วนของออนโทโลยีหนังสือ

5. กระบวนการวัดค่าคะแนนความคล้ายคลึงระหว่างเว็บเซอร์วิสเป็นกระบวนการการแทนค่าคะแนนในแต่ละคู่ที่นำมาเปรียบเทียบ โดยที่คะแนนที่แทนค่าลงไปนั้นได้มาจากการวัดระยะห่างระหว่างเว็บเซอร์วิสที่นำมาเปรียบเทียบ ว่ามีระยะห่างเท่าไรก็แทนค่าลงไปในสมการ ดังต่อไปนี้

นิยามที่ 1 คือการวัดค่าความคล้ายคลึงเชิงความหมาย (Semantic Similarity) ระหว่างออนโทโลยี สามารถนิยามได้ดังนี้

$$sim < outR, outA > = \begin{cases} 1, & \text{if } outR = outA \\ \frac{1}{(distance(outR, outA) + 1)}, & \text{if } outA \subset outR \\ \frac{1}{2} + \frac{1}{(distance(outR, outA) + 1)}, & \text{if } outR \subset outA \\ 0, & \text{if } outR \not\subset outA \text{ and } outR \not\supset outA \end{cases}$$

จากนิยามที่ 1 สามารถหาค่าคะแนนความคล้ายคลึง $sim<outR,outA>$ ได้ดังนี้

- ถ้า $outR=outA$ หรือมีระดับความสัมพันธ์เป็น Exact กำหนดค่าคะแนนเป็น 1
- ถ้า $outA \subset outR$ หรือมีระดับความสัมพันธ์เป็น Subsumes กำหนดค่าคะแนนเป็น $1/((distance(outR,outA)+1))$
- ถ้า $outR \subset outA$ หรือมีระดับความสัมพันธ์เป็น Plugin กำหนดค่าคะแนนเป็น $1/2 + 1/((distance(outR,outA)+1))$
- ถ้า $outR \not\subset outA$ and $outR \not\supset outA$ หรือมีระดับความสัมพันธ์เป็น Fail กำหนดค่าคะแนนเป็น 0

โดยที่

- $outR$ คือข้อมูลออกของเว็บเซอร์วิสที่ร้องขอ (Request)
- $outA$ คือข้อมูลออกของเว็บเซอร์วิสที่ประกาศเอาไว้ (Advertisement)

- $distance(outR,outA)$ คือระยะห่างเชิงแนวคิดระหว่าง $outR$ กับ $outA$ โดยที่ระยะห่างระหว่างแนวคิดนี้มีเงื่อนไขต่างๆ ดังต่อไปนี้

1. ถ้า $outR$ กับ $outA$ มีระดับความสัมพันธ์เป็น Exact แล้ว $distance(outR,outA)=0$
2. ถ้า $outR$ กับ $outA$ มีระดับความสัมพันธ์เป็น Fail แล้ว $distance(outR,outA)=\infty$
3. ถ้า $outR$ กับ $outA$ มีระดับความสัมพันธ์เป็น Subsume หรือ Plugin แล้ว $distance(outR,outA)$ คือจำนวนระยะห่างระหว่างเส้นทางจากโหนด $outR$ ไปถึง $outA$ ถ้าหากมีหลายเส้นทางให้นับระยะทางที่สั้นที่สุด (Shortest path)

หลังจากผ่านกระบวนการจับคู่แล้วก็นำมาพิจารณาแทนค่าตามนิยามที่ 1 จะได้ดังตารางที่ 5 ตารางที่ 6 และตารางที่ 7

ตารางที่ 5 แสดงการวัดค่าคะแนนความคล้ายคลึงของซื้อบริการเว็บเซอร์วิส

Advertisement	Request	Degree of Match	Sim<Req,Adv>	max(Req, Adv)
Book&CDSelling Service	RequestBook&CD&DVDsSellingService	Exact	1	1

ตารางที่ 6 แสดงการวัดค่าคะแนนความคล้ายคลึงของข้อมูลเข้า

Advertisement	Request	Degree of Match	Sim<inR,inA>	max(inR,inA)
year	year	Exact	1	1

ตารางที่ 7 แสดงการวัดค่าคะแนนความคล้ายคลึงของข้อมูลออก

Advertisement	Request	Degree of Match	Sim<outR,outA>	max(outR,outA)
Computer Book	Book	Exact	1	1
Computer Book	CD	Fail	0	
Classic Music	Book	Fail	0	5/6
Classic Music	CD	Subsume	5/6	
DVD	Book	Fail	0	0
DVD	CD	Fail	0	

การคำนวณหาค่าคะแนนความคล้ายคลึงนั้น สามารถคำนวณได้ด้วยการนำค่าที่มากที่สุดของแต่ละคู่ที่วัดระดับความสัมพันธ์และระดับความเหมือนกัน มาคูณด้วยค่าน้ำหนักของแต่ละพารามิเตอร์

$$\text{degreeOfServiceMatch} = w_0 * \text{sim}(\text{book}) + w_2 * \text{sim}(\text{year}) + w_1 * \text{sim}(\text{computer book}) + w_1 * \text{sim}(\text{class music}) + w_1 * \text{sim}(\text{DVD})$$

โดยที่

w_0 = ค่าน้ำหนักของ Service Name

w_1 = ค่าน้ำหนักของข้อมูลออก output

w_2 = ค่าน้ำหนักของข้อมูลเข้า Input

ดังนั้น

$$\text{degreeOfServiceMatch} = ((1*1) + (0.7*1) + (0.9*1) + (0.9*5/6) + (0.9*0)) / 5 = 0.67$$

เพราะฉะนั้นค่าที่ได้จากการวัดค่าความคล้ายคลึงระหว่างเว็บเซอร์วิสที่นำมาเปรียบเทียบจะมีค่าประมาณ 0.67

ผลการศึกษาวิจัยและการอภิปรายผล

จากการออกแบบและทำการทดลองการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย ในงานวิจัยนี้ได้ทำการจำลองข้อมูลเว็บเซอร์วิสที่ได้จากงานวิจัย [7] จำนวน 17 ชุด และใช้ข้อมูลโดเมนออนโทโลยีจาก OWL-S service retrieval test collection [3] จำนวน 10 โดเมนออนโทโลยี

ผลการทดลองการจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย พบว่าการนำพารามิเตอร์ของเว็บเซอร์วิสมาทำการเปรียบเทียบหาระดับความสัมพันธ์เชิงแนวคิดแล้วนำระดับความสัมพันธ์ที่ได้ไปทำการวัดค่าคะแนนความคล้ายคลึงเชิงความหมายระหว่างคู่ของเว็บเซอร์วิสที่เปรียบเทียบในแต่ละคู่จนครบทุกคู่ พบว่าสามารถจัดกลุ่มของเว็บเซอร์วิสได้ทั้งหมด 3 กลุ่ม

ตารางที่ 8 แสดงการวัดประสิทธิภาพการจัดกลุ่มเว็บเซอร์วิส

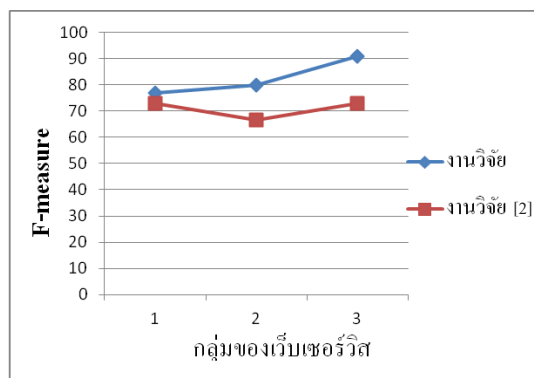
Cluster	Recall	Precision	F-measure
1	83.33%	71.43%	76.92%
2	80.00%	80.00%	80.00%
3	100%	83.33%	90.91%

การวัดประสิทธิภาพการจัดกลุ่มในแต่ละกลุ่มด้วยการวัดค่าความถูกต้อง (Precision) ค่าความครบถ้วน (Recall) และ ค่าเอฟเมเชอร์ (F-measure) ของการจัดกลุ่ม

ตารางที่ 9 แสดงการเปรียบเทียบการทำงานระหว่าง
วิทยานิพนธ์และงานวิจัย [2]

คำอธิบาย	งานวิจัย [2]	วิทยานิพนธ์
- การเชื่อมโยง ออนโทโลยี	ไม่มี	ทำการเชื่อมโยง ออนโทโลยี
- พารามิเตอร์ที่ นำมาพิจารณา	ข้อมูลเข้า, ข้อมูลออก	ชื่อเว็บเซอร์วิส, ข้อมูลเข้าและ ข้อมูลออก
- วิธีการถ่วง น้ำหนัก	หาค่าเฉลี่ย	กำหนดน้ำหนัก ความสำคัญของ พารามิเตอร์
- จำนวนตัวอย่าง เว็บเซอร์วิส	17 ชุด	17 ชุด
- จำนวนโดเมน ออนโทโลยี	10 โดเมน	10 โดเมน

ผลการทดสอบประสิทธิภาพระหว่าง
วิทยานิพนธ์และงานวิจัย [2] โดยใช้ค่าเอฟเมเชอร์
แสดงค่าดังตารางที่ 1 ซึ่งพบว่าค่าเอฟเมเชอร์ของ
วิทยานิพนธ์สูงกว่าค่าเอฟเมเชอร์ของงานวิจัย [2]
จากภาพที่ 10 แสดงกราฟเปรียบเทียบประสิทธิภาพ



ภาพที่ 10 กราฟแสดงการเปรียบเทียบประสิทธิภาพ
ด้วยค่าเอฟเมเชอร์

สรุปผลการศึกษวิจัย

งานวิจัยนี้ได้นำเสนอแนวทางหนึ่งสำหรับ
การจัดกลุ่มเว็บเซอร์วิสเชิงความหมาย (Semantic
web service) [8] โดยประยุกต์ใช้ความรู้ด้านการ
ประมวลผลภาษาธรรมชาติ (natural language
processing) เพื่อวิเคราะห์การสอดคล้องกันเชิง
ความหมายของเว็บเซอร์วิสโดยใช้ฐานความรู้เวิร์ด
เน็ตมาช่วยในการวิเคราะห์คำศัพท์และหาความหมาย
ระหว่างคำศัพท์ที่นำมาวิเคราะห์ งานวิจัยนี้อาศัย
กระบวนการการจับคู่เว็บเซอร์วิส (Web Service
Matching) โดยเปรียบเทียบระดับความสัมพันธ์
ระหว่างเว็บเซอร์วิส (Degree of Match) หลังจากนั้น
ทำการหาค่าคะแนนความคล้ายคลึงกันของเว็บ
เซอร์วิส

กระบวนการจัดกลุ่มเว็บเซอร์วิสเชิงความ
หมายทำให้ได้ฐานความรู้เว็บเซอร์วิสที่มีการจัดแบ่งหมวด
หมู่ที่ถูกต้องตรงตามความหมายมากขึ้น ซึ่งพิจารณา
ได้จากการวัดประสิทธิภาพด้วยค่าเอฟเมเชอร์ โดยค่า
เอฟเมเชอร์มีค่าสูงขึ้นเมื่อเทียบกับวิธีการวัดค่าคะแนน
ความคล้ายคลึงในแบบเดิม นอกจากนี้ผู้วิจัยได้ทำการ
วัดประสิทธิภาพของการจัดกลุ่มเว็บเซอร์วิสเชิงความ
หมายเปรียบเทียบกับงานวิจัย [2] จากการทดลองพบ
ว่ามีความแตกต่างกันโดยสรุป ดังนี้

1. การเพิ่มความสามารถในการเชื่อมโยง
กับออนโทโลยีหลัก ทำให้โดเมนออนโทโลยีนั้นเกิด
การเรียนรู้และเพิ่มประสิทธิภาพในการวัดค่าคะแนน
ความคล้ายคลึงเชิงความหมายมากขึ้น

2. การพิจารณาชื่อของเว็บเซอร์วิส (Service
Name) และการถ่วงน้ำหนักเพื่อพิจารณาค่าคะแนน
ความคล้ายคลึงนั้น ช่วยให้การหาค่าคะแนนความ
คล้ายคลึงของเว็บเซอร์วิสมีประสิทธิภาพมากขึ้น

อย่างไรก็ตามงานวิจัยนี้มีรายละเอียดที่ควร
ได้รับการพัฒนาต่อ ซึ่งหากเพิ่มขอบเขตของข้อมูล
ที่นำมาพิจารณาก็จะทำให้ได้ระดับความสัมพันธ์
เชิงความหมายมากขึ้น เช่น คำอธิบายเว็บเซอร์วิส
(Service Description) เงื่อนไขของเว็บเซอร์วิส

(precondition, effect) เป็นต้น เพื่อให้สามารถนำมาจัดกลุ่มได้อย่างถูกต้องเหมาะสมมากขึ้น และนอกจากนี้ การพัฒนาโดเมนออนโทโลยี (Domain Ontology) ให้ครอบคลุมในหลายสาขาวิชา เพื่อให้สามารถบรรยายแนวคิดของโดเมน (Domain) หรือขอบเขตความสนใจใดๆ ให้ถูกต้องครบถ้วน เนื่องจากโดเมนออนโทโลยี เป็นแนวคิดกลางที่ใช้ในงานวิจัยนี้ในการวัดระดับความสัมพันธ์ของแนวคิดที่พิจารณา ช่วยให้กำหนดขอบเขต และช่วยเพิ่มความหมายในกระบวนการวัดค่าคะแนนความคล้ายคลึงเชิงความหมายมากขึ้น

เอกสารอ้างอิง

1. Gruber, T. 1993. A Translational Approach to Prototypical Ontologies. Knowledge Acquisition, Vol. 5 No. 2, 1993:199-220.
2. Hui peng et al. Improving Grade Match to Value Match for Semantic Web Service Discovery. In Natural computation, 2008. ICNC'08 Fourth International Conference. Vol.5, 2008: 232-236
3. Matthias Klusch, OWL-S Service Retrieval Test Collection [online] 2005. Available from: <http://www.semwebcentral.org/projects/owls-tc/>
4. Massimo Paolucci et al. Semantic Matching of Web Service Capabilities. In Proceedings of the 1st International Semantic Web Conference, 2002.
5. Prince-ton University Cognitive Science Laboratory. Wordnet-a lexical database for English [online] 2008. Available from: <http://wordnet.princeton.edu>
6. Prasenjit Mitra, Gio Wiederhold. Resolving Terminological Heterogeneity in Ontologies, In: Proceedings of Workshop on Ontologies and Semantic Interoperability at the 15th European Conference on Artificial Intelligence (ECAI); July 2002;2002.
7. Viriya K. Semantic Web Service Discovery and Integration using Service Search Crawler. NCSEC, 2008; 12: 626-631 Thai.
8. W3C, World Wide Web Consortium. OWL-S: Semantic Markup for Web Services [Online] 2004. Available from: <http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/>.