

## ระบบการสืบค้นเว็บเซอร์วิสโดยใช้เวกเตอร์สเปซโมเดล\*

### Web Service Retrieval by Using Vector Space Model

วีณาดี ม่วงอัน (Weenawadee Muangon)<sup>\*\*</sup>

คทาวุธ แก้วบรรจง (Katawut Kaewbanjong)<sup>\*\*\*</sup>

#### บทคัดย่อ

ปัจจุบันมีการนำเทคโนโลยีเว็บเชิงความหมายมาทำการประยุกต์ใช้ในการค้นหาเว็บเซอร์วิสเพื่อให้ได้รับบริการได้ตรงตามความต้องการของผู้ใช้งาน แต่เนื่องจากปริมาณเว็บเซอร์วิสที่มีมากมายในปัจจุบันจึงเป็นไปได้ยากที่จะสามารถค้นหาเว็บเซอร์วิสได้ตรงตามความต้องการของผู้ใช้ งานวิจัยนี้ได้นำเสนอระบบสืบค้นเว็บเซอร์วิสด้วยการพิจารณารายละเอียดเชิงความหมายของเว็บเซอร์วิส โดยทำการสร้างดัชนีจากส่วนประวัติการบริการ, ส่วนแบบจำลองการบริการ และส่วนพื้นฐานการบริการ ของเว็บเซอร์วิส โดยแบบจำลองการสืบค้นนี้จะทำการแทนค่าแต่ละเว็บเซอร์วิสบนพื้นฐานของเวกเตอร์สเปซโมเดลเพื่อให้ได้ผลลัพธ์ที่มีความใกล้เคียงกับคำสอบถามมากที่สุด ผลการทดลองเบื้องต้นสำหรับการใช้แบบจำลองการสืบค้นนี้ชี้ให้เห็นว่ามีค่าเฉลี่ยความแม่นยำที่ร้อยละ 63.60 %

**คำสำคัญ:** ระบบการสืบค้นสารสนเทศ การสืบค้นเว็บเซอร์วิส คำอธิบายเว็บเซอร์วิสเชิงความหมายเวกเตอร์สเปซโมเดล

---

\* บทความวิจัยนี้แนะนำระบบการสืบค้นเว็บเซอร์วิส ด้วยการสร้างดัชนีในการสืบค้นจากการคำอธิบายรายละเอียดเชิงความหมายของเว็บเซอร์วิสเพื่อรองรับการสืบค้นจากผู้ใช้งานด้วยคำสำคัญในรูปแบบของภาษาธรรมชาติ

This research presented web service retrieval model that consider the details of web service through the profile of semantic web services to support the search for a keyword in form of natural language.

\*\* อาจารย์ประจำภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร 6 ถนนราชมรรคาใน อ.เมือง นครปฐม 73000 Email: weenawadee@hotmail.com

Department of Computing. Faculty of Science. Silpakorn University. 6 Rajamankha Nai Road, Inbound District, Nakhon Pathom, 73000, Thailand.

\*\*\* อาจารย์ประจำหลักสูตรระบบสารสนเทศทางคอมพิวเตอร์ คณะบริหารธุรกิจและศิลปศาสตร์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน 59 หมู่ 13 ต.ฝายแก้ว อ.ภูเพียง น่าน 55000 Email: katawutk@gmail.com

Business Information System Department. Faculty of Business Administration and Liberal Art. Rajamangala University of Technology Lanna Nan. 59 Moo 13, Fai kaeo, Phu Phiang District, Nan, 55000, Thailand.

## Abstract

Currently, semantic web technology provided to support web services searching system. However, web services are created rapidly, therefore, it is difficult to retrieve appropriately web services to users. This research presented web service retrieval model that consider the details of web service through the profile of semantic web services. The web service retrieval model starts with indexing construction. It uses semantic web service description as index resource such as the history of the service, the service model and service fundamentals of web service. Further, each of the web services is represented on the basis of the vector space model to achieve web service results that are the most similar with user's query. The preliminary results of this retrieving model shows that the average accuracy rate of 63.60%.

**Keyword:** Information retrieval system, Web service retrieval, Semantic web service description, Vector space model

## บทนำ

ในยุคแห่งระบบสารสนเทศ ข้อมูล และบริการต่างๆ ได้กระจายอยู่ทั่วไปในระบบอินเทอร์เน็ต และมีแนวโน้มที่จะเพิ่มมากขึ้นอย่างรวดเร็วต่อไปในอนาคต เว็บเซอร์วิสเป็นเทคโนโลยีหนึ่งที่ถูกนำมาใช้ในการนำเสนอ บริการขององค์กร หน่วยงาน ทั้งภาครัฐและภาคเอกชน ผู้ใช้บริการผ่านระบบเครือข่ายอินเทอร์เน็ต เช่น การให้บริการรายงานสภาพข้อมูลอากาศ การให้บริการรายงานความเคลื่อนไหวของราคาหลักทรัพย์ของตลาดหลักทรัพย์แห่งประเทศไทย เป็นต้น

อย่างไรก็ตามการมีเว็บเซอร์วิสจำนวนมาก ทำให้เกิดปัญหาในการสืบค้น นั่นคือ ผู้ใช้ไม่สามารถที่จะ ค้นหาข้อมูลได้อย่างถูกต้องแม่นยำ ในปัจจุบันมีการนำเสนอระบบการสืบค้นเว็บเซอร์วิสหลายงานวิจัย โดย โครงสร้างการเก็บข้อมูลของงานวิจัยส่วนใหญ่จะเก็บประเภทข้อมูลในรูปแบบต่างๆ เช่น ประเภทข้อมูลนำเข้า (Input Data Type), ประเภทข้อมูลนำออก (Output Data Type), จำนวนตัวแปรของข้อมูลนำเข้า (Input Parameters), จำนวนตัวแปรของข้อมูลนำออก (Output Parameters) และชื่อของเว็บเซอร์วิส เป็นต้น โดย ระบบสืบค้นจะทำการรับคำสอบถามตามโครงสร้างที่จัดทำขึ้นซึ่งอาจเป็นการยากต่อผู้ใช้งานทั่วไป หรือแม้กระทั่ง นักพัฒนาเว็บเซอร์วิส (Web Service Developer) เนื่องจากต้องรู้จักเว็บเซอร์วิสทั้งหมด รวมทั้งต้องจดจำ พารามิเตอร์เพื่อมาทำการป้อนข้อมูลให้สอดคล้องกับโครงสร้างการจัดเก็บของเว็บเซอร์วิส

แนวทางหนึ่งในการแก้ปัญหา คือ การให้ผู้ใช้งานได้สืบค้นเว็บเซอร์วิสตามลักษณะความต้องการโดยใช้คำสำคัญ (Keyword) โดยกระบวนการสืบค้นด้วยคำสำคัญนี้ ระบบจะเริ่มต้นทำการสร้างดัชนีจากการวิเคราะห์คำอธิบายรายละเอียดเชิงความหมายของเว็บเซอร์วิส (Semantic web service description) ซึ่งถูกนำเสนอด้วยภาษาโอดับเบิลวอลเอส (OWL-S) ซึ่งดัชนีเหล่านี้จะถูกแทนที่ด้วยเวกเตอร์เพื่อนำไปเปรียบเทียบกับ

เวกเตอร์คำสอบถาม โดยทำการวัดค่าความคล้ายคลึงตามวิธีของเวกเตอร์สเปซโมเดล โดยผลลัพธ์ที่ได้ คือ เว็บเซอร์วิสที่ถูกเรียงลำดับตามค่าความคล้ายคลึง โดยเว็บเซอร์วิสที่มีความคล้ายมากที่สุดจะถูกจัดอยู่ในลำดับต้นๆ

### งานวิจัยที่เกี่ยวข้อง

Aielo et al (2006) และ Adam and Aleksander (2014) นำเสนอระบบการสืบค้นเว็บเซอร์วิสด้วยการวิเคราะห์คำอธิบายรายละเอียดที่ถูกนำเสนอด้วยภาษาดัชนีเอสดีแอล ในการสร้างดัชนีการสืบค้น โดย Aielo et al (2006) จัดเก็บดัชนีในระบบแฮชเทเบิล (Hash Table) เพื่อรองรับการสืบค้นเว็บเซอร์วิสด้วยคำสอบถามที่เกี่ยวข้องกับฟังก์ชันในการทำงานเป็นหลัก และ Adam and Aleksander (2014) ได้เสนอวิธีการจัดกลุ่มดัชนีก่อนทำการสืบค้น อย่างไรก็ตามงานวิจัยนี้มีข้อจำกัดในด้าน ประสิทธิภาพของการสืบค้นซึ่งขึ้นอยู่กับจำนวนดัชนีที่ได้จากการจัดกลุ่มเว็บเซอร์วิส โดยอาจมีจำนวนดัชนีมากขึ้นอยู่กับจำนวนเว็บเซอร์วิสในแต่ละกลุ่ม

Tamilarasi and Ramakrishnan (2015) นำเสนอการพัฒนาระบบจัดเก็บเว็บเซอร์วิส (Web Service Registry) ด้วยการจัดเก็บคำอธิบายรายละเอียดข้อมูลเว็บเซอร์วิสที่ทำการลงทะเบียนไว้กับระบบในลักษณะของดัชนีแทนคำสำคัญ ร่วมกับข้อมูลคุณภาพในการให้บริการ (Quality of Service: QoS) โดยผลการทดลองจากงานวิจัยนี้ชี้ให้เห็นว่าระบบสามารถกำจัดเว็บเซอร์วิสที่ไม่เกี่ยวข้องออกได้อย่างมีประสิทธิภาพและใช้เวลาในการสืบค้นอย่างรวดเร็ว

Zhou et al (2009) นำเสนอระบบการสืบค้นเว็บเซอร์วิสด้วยการสร้างดัชนีผกผัน (Inverted index) จากข้อมูลประวัติการบริการของคำอธิบายรายละเอียดเชิงความหมาย โดยผลการทดลองจากงานวิจัยนี้ชี้ให้เห็นประสิทธิภาพในการสืบค้นเร็วกว่าการจับคู่ตามลำดับ (Sequential match making) และให้ค่าเรียกกลับ (Recall) ในการสืบค้นที่ดีกว่าระบบการสืบค้นที่ไม่ได้นำเอาดัชนีผกผันมาใช้

### แนวคิดและทฤษฎีที่เกี่ยวข้อง

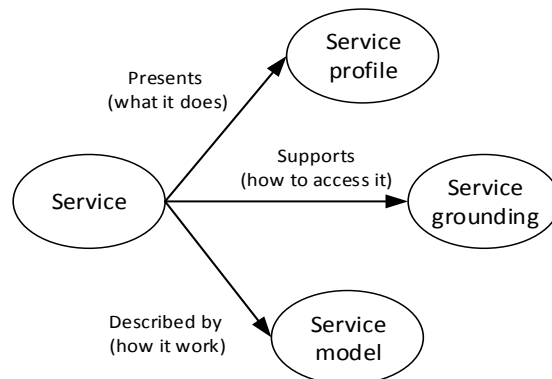
#### เว็บเซอร์วิส (Web service)

คือ โปรแกรมที่ทำหน้าที่อย่างใดอย่างหนึ่งในลักษณะการให้บริการผ่านระบบเครือข่ายอินเทอร์เน็ต โดยเป็นตัวกลางในการสื่อสารระหว่างผู้ให้บริการ (Service provider) และผู้ร้องขอบริการ (Service Requestor) โดยผู้ให้บริการจะทำการลงทะเบียนข้อมูลเว็บเซอร์วิสกับผู้รับลงทะเบียนเว็บเซอร์วิส (Service registry) โดยมีคำอธิบายรายละเอียดและรูปแบบการเรียกใช้งานสำหรับกระบวนการเรียกใช้งาน และเพื่อสนับสนุนการสืบค้น แต่อย่างไรก็ตามด้วยข้อจำกัดของคำอธิบายซึ่งไม่สามารถแยกแยะความแตกต่างระหว่างเว็บเซอร์วิสที่มีความคล้ายคลึงกันได้เท่าที่ควร ดังนั้นการเพิ่มคำอธิบายเชิงความหมายจึงเป็นถูกพัฒนาขึ้นมา (สมเกียรติ สอนवल และคณะ, 2556)

#### เว็บเซอร์วิสเชิงความหมาย (Semantic web service)

เว็บเซอร์วิสเชิงความหมาย หมายถึง เว็บเซอร์วิสที่มีคำอธิบายรายละเอียดข้อมูลในลักษณะคำจำกัดความ โดยมีโครงสร้างข้อมูลที่ทำให้คอมพิวเตอร์สามารถทำความเข้าใจข้อมูลให้สอดคล้องกับการตีความของมนุษย์ ซึ่งประกอบด้วย ข้อมูลประวัติการให้บริการ (Service profile) แบบจำลองการบริการ (Service model)

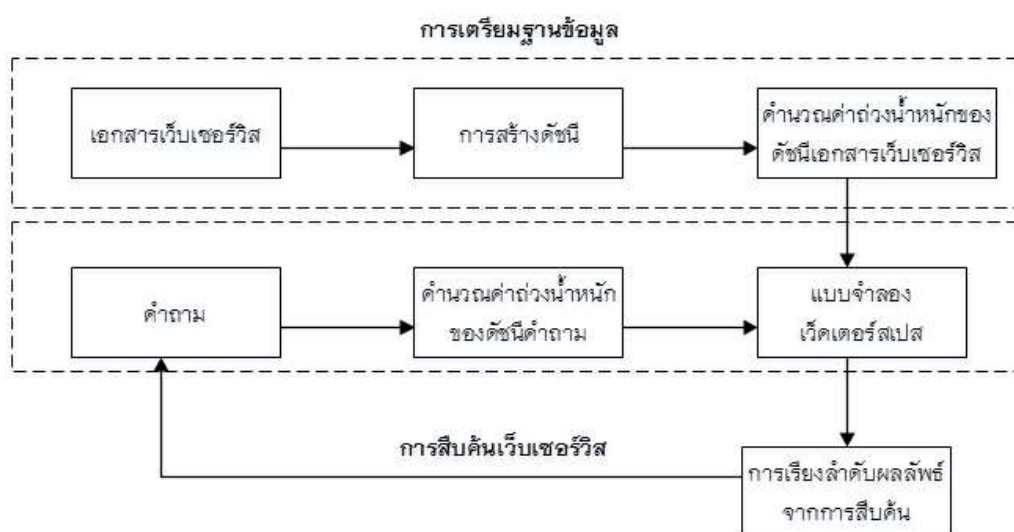
และพื้นฐานการบริการ (Service grounding) (จตุพล โพธิ์คำพก และดร.สมจิตร อาจอินทร์, 2555) ซึ่งทั้ง 3 ส่วนนั้นอยู่ภายใต้ คลาส (Class) ที่เรียกว่าเซอร์วิส (Service) (David, 2008) ดังภาพที่ 1



ภาพที่ 1. องค์ประกอบของคำอธิบายเว็บเซอร์วิสเชิงความหมาย

### การสืบค้นสารสนเทศ (Information Retrieval)

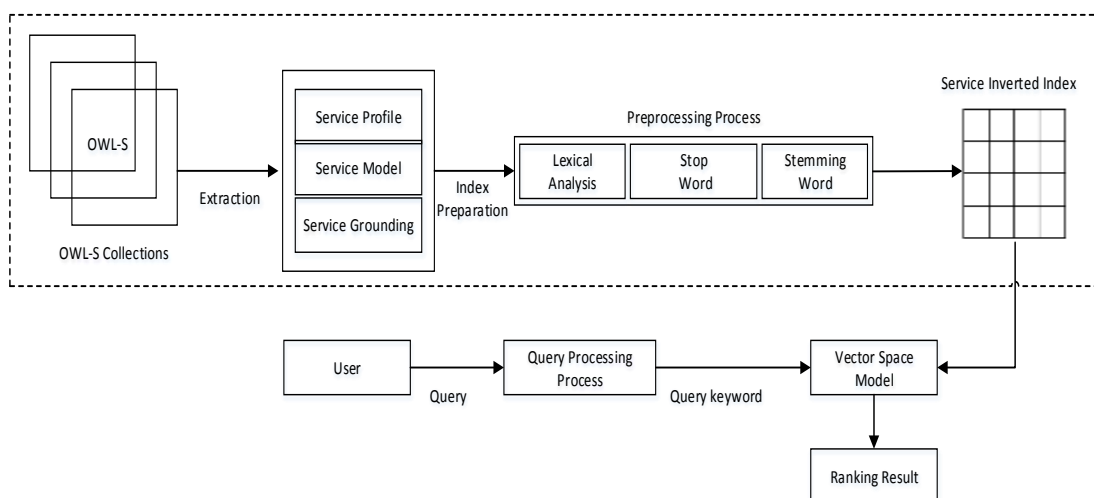
การสืบค้นสารสนเทศ คือ ศาสตร์ทางด้านวิทยาการคอมพิวเตอร์ ซึ่งมีจุดประสงค์เพื่อการสืบค้นเอกสารให้ตรงกับความต้องการของผู้ใช้งานมากที่สุด ซึ่งประกอบด้วยขั้นตอนหลักๆ ได้แก่ การเตรียมเอกสารเพื่อให้เหมาะสมในการเป็นดัชนี ซึ่งดัชนีเหล่านี้จะเป็นตัวแทนของเอกสารจะถูกแทนที่ด้วยเวกเตอร์ โดยเวกเตอร์เหล่านี้จะถูกนำไปทำการวัดความคล้ายคลึงระหว่างเอกสารและคำสอบถาม เพื่อทำการจัดลำดับผลลัพธ์ ด้วยหลักการของเวกเตอร์แทนเอกสารและคำสอบถาม โดยที่แกนแต่ละแกนของคือ ดัชนีที่ปรากฏอยู่ในเอกสาร และใช้หลักการค่าถ่วงน้ำหนักของดัชนี ในการกำหนดขนาดของเวกเตอร์ ดังนั้นในแบบจำลองดังกล่าวนี้จะประกอบด้วยเวกเตอร์เอกสาร และเวกเตอร์คำสอบถาม ซึ่งแต่ละเวกเตอร์จะประกอบด้วยขนาดและทิศทาง ดังนั้นหากเวกเตอร์เอกสารใดที่ชี้ไปในทิศทางเดียวกันกับเวกเตอร์คำสอบถาม จะหมายถึงเอกสารนั้นมีเนื้อหาที่ใกล้เคียงกับคำสอบถามและควรเป็นคำตอบสำหรับคำสอบถามนั้น กระบวนการแสดงดังภาพที่ 2



ภาพที่ 2. กระบวนการสืบค้นสารสนเทศ

## วิธีการวิจัย

แบบจำลองระบบการสืบค้นเว็บเซอร์วิสในงานวิจัยนี้ประกอบไปด้วย 4 ขั้นตอน ดังต่อไปนี้ ขั้นตอนการสกัดข้อมูลจากคำอธิบายเอกสารเว็บเซอร์วิส, การสร้างดัชนีเว็บเซอร์วิส, การหาค่าถ่วงน้ำหนักของดัชนี, และการหาค่าความคล้ายคลึงระหว่างเอกสารที่เป็นตัวแทนเว็บเซอร์วิสและคำถามของผู้ใช้ แนวคิดดังกล่าวแสดงดังภาพที่ 3.



ภาพที่ 3. แสดงแบบจำลองของการสืบค้นเว็บเซอร์วิสโดยใช้เวกเตอร์สเปซโมเดล

## ขั้นตอนการสกัดข้อมูลจากคำอธิบายเอกสารเว็บเซอร์วิส

ในงานวิจัยนี้ จะทำการวิเคราะห์ข้อมูลจากเว็บเซอร์วิสเชิงความหมายที่ถูกรวบรวมด้วยโครงสร้างของภาษาโอดับบลิวแอลเอสซึ่งประกอบด้วยคลาสประวัติการบริการ, คลาสรูปแบบการบริการ, และคลาสพื้นฐานการบริการ ภาพที่ 4 แสดงตัวอย่างโครงสร้างการอธิบายเว็บเซอร์วิส Book\_OracleService.owl (Matthias, 2010) แสดงดังภาพที่ 4 งานวิจัยนี้สกัดส่วนรายละเอียดข้อมูลจากคลาสประวัติการบริการที่เหมาะสมที่จะนำไปเป็นดัชนีเพื่อการสืบค้นเว็บเซอร์วิสได้ 2 คุณสมบัติ (Property) และ 4 ฟังก์ชัน (Function) ดังนี้

1. serviceName คือ ชื่อของเว็บเซอร์วิสซึ่งเป็นค่าบ่งชี้ที่เป็นลักษณะเฉพาะที่ทำให้ผู้ใช้งานสามารถเข้าใจได้ถึงความสามารถในการให้บริการในลักษณะของภาพรวม
2. textDescription คือ กลุ่มของคำอธิบายที่ทำหน้าที่สื่อถึงรายละเอียดเชิงลึกที่เกี่ยวข้องจุดประสงค์ในการจัดสร้าง หน้าที่ ข้อมูลสำหรับติดต่อผู้พัฒนาเว็บเซอร์วิส
3. hasOutput คือ คำอธิบายของการนำข้อมูลออกของเว็บเซอร์วิส ซึ่งเป็นผลลัพธ์ที่ได้จากการประมวลผลเมื่อผู้ใช้งานทำการกำหนดข้อมูลนำเข้า
4. hasInput คือ คำอธิบายของการรับข้อมูลนำเข้าของเว็บเซอร์วิสเพื่อทำการประมวลผลตามลักษณะหรือจำนวนของข้อมูลนำเข้า

5. hasPrecondition คือ คำอธิบายที่เกี่ยวกับเงื่อนไขที่จำเป็นต้องได้รับการตรวจสอบหรือการประมวลผลก่อนการประมวลผลข้อมูลนำเข้า

6. hasEffect คือ ผลลัพธ์จากการประมวลผลที่ขึ้นอยู่กับข้อมูลที่ปรากฏใน hasPrecondition

```
<profile:Profile rdf:ID="_BOOK_PROFILE">
<service:isPresentedBy rdf:resource="#_BOOK_SERVICE"/>
<profile:serviceName xml:lang="en">
Computer Networking Training Service
</profile:serviceName>
<profile:textDescription xml:lang="en">
This service returns books information to teach and help exams of Oracle networking course
</profile:textDescription>
<profile:hasOutput rdf:resource="#_BOOK"/>
<profile:has_process rdf:resource="_BOOK_PROCESS" /></profile:Profile>
```

ภาพที่ 4. ส่วนประกอบของข้อมูลจากคลาสประวัติการบริการ

```
<!--<process:ProcessModel rdf:ID="_BOOK_PROCESS_MODEL">
<service:describes rdf:resource="#_BOOK_SERVICE"/>
<process:hasProcess rdf:resource="#_BOOK_PROCESS"/>
</process:ProcessModel-->
<process:AtomicProcess rdf:ID="_BOOK_PROCESS">
<service:describes rdf:resource="#_BOOK_SERVICE"/>
<process:hasOutput rdf:resource="#_BOOK"/>
</process:AtomicProcess>
<process:Output rdf:ID="_BOOK">
<process:parameterType rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
<rdfs:label></rdfs:label>
</process:Output>
```

ภาพที่ 5. ส่วนประกอบของข้อมูลจากคลาสแบบจำลองการบริการ

จากภาพที่ 5 งานวิจัยนี้สกัดส่วนรายละเอียดข้อมูลจากคลาสแบบจำลองการบริการที่เหมาะสมที่จะนำใช้เป็นดัชนีเพื่อการสืบค้นเว็บเซอร์วิสได้ 4 ฟังก์ชัน ดังนี้

1. hasPrecondition คือ คำอธิบายที่เกี่ยวกับกระบวนการทำงานของเว็บเซอร์วิสที่จำเป็นต้องทำการตรวจสอบเงื่อนไข

2. hasInput คือ คำอธิบายที่เกี่ยวกับกระบวนการทำงานของเว็บเซอร์วิสที่เกี่ยวข้องกับการรับข้อมูลนำเข้าของเว็บเซอร์วิสเพื่อทำการประมวลผลตามลักษณะหรือจำนวนของข้อมูลนำเข้า

3. hasOutput คือ คำอธิบายที่เกี่ยวกับกระบวนการทำงานของเว็บเซอร์วิสที่เกี่ยวข้องกับการนำข้อมูลออกของเว็บเซอร์วิสเพื่อทำการประมวลผลตามลักษณะหรือจำนวนของข้อมูลนำออก

4. hasEffect คือ ผลลัพธ์จากการประมวลผลของกระบวนการทำงานที่ขึ้นอยู่กับข้อมูลที่ปรากฏใน hasPrecondition

```
<grounding:wSDLOutput>
  <grounding:WSDLOutputMessageMap>
    <grounding:owlsParameter rdf:resource="#_BOOK"/>
    <grounding:wSDLMessagePart rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"
      >http://127.0.0.1/wSDL/BookComputer#_BOOK</grounding:wSDLMessagePart>
    <grounding:xsltTransformationString>None (XSL)</grounding:xsltTransformationString>
  </grounding:WSDLOutputMessageMap>
</grounding:wSDLOutput>
```

ภาพที่ 6. ส่วนประกอบของข้อมูลจากคลาสพื้นฐานการบริการ

จากภาพที่ 6 งานวิจัยนี้สกัดส่วนรายละเอียดข้อมูลจากคลาสแบบพื้นฐานการบริการที่เหมาะสมที่จะนำใช้เป็นดัชนีเพื่อการสืบค้นเว็บเซอร์วิสได้ 2 ฟังก์ชัน ดังนี้

1. wsdInput คือ คำอธิบายที่แสดงถึงข้อมูลนำเข้าเพื่อการเชื่อมต่อกับเว็บเซอร์วิสสำหรับการเรียกใช้งาน
2. wsdOutput คือ คำอธิบายที่แสดงผลลัพธ์ที่ได้จากการเชื่อมต่อกับเว็บเซอร์วิสจากการเรียกใช้งาน

### การสร้างดัชนีเว็บเซอร์วิส

เนื่องจากในระบบสืบค้นสารสนเทศ ดัชนีถือเป็นเครื่องมืออำนวยความสะดวกในการสืบค้น จากผลลัพธ์ที่ได้จากการสกัดข้อมูลข้างต้นอาจมีข้อมูลที่ไม่สามารถนำไปใช้สร้างดัชนีได้ ดังนั้นจึงนำข้อมูลดังกล่าวเข้าสู่กระบวนการเตรียมข้อมูล (Preprocessing process) นั่นคือ การวิเคราะห์คำ (Lexical analysis), การหาคำหยุด (Stopword) และการหารากศัพท์ของคำ (Lemmatization) (Lecture International, 1993)

การวิเคราะห์คำ คือ การหาตัวแทนเอกสารโดยการแปลงข้อความทั้งหมดเพื่อจำแนกออกเป็นคำ

การหาคำหยุด คือ การจัดคำที่มีความถี่สูงออกไป เนื่องจากคำเหล่านั้นไม่มีคุณสมบัติในการบ่งชี้ จึงไม่มีประโยชน์ในการนำมาใช้เป็นดัชนี เช่น บุพบท ตัวเลข หรือสัญลักษณ์ต่างๆ เป็นต้น

การหารากศัพท์ของคำ คือ การหาคำที่เป็นรากศัพท์ของคำเดียวกัน เช่น computer, computing, computed มีรากศัพท์ร่วมกันคือ computer เป็นต้น

### การคำนวณค่าถ่วงน้ำหนักของดัชนี

ขั้นตอนนี้เป็นากำหนดค่าน้ำหนักให้กับตัวแทนเอกสารแต่ละคำ โดยทั่วไปหาได้จาก ผลคูณระหว่างค่าความถี่ของคำ (Term Frequency: tf) กับค่าความถี่ผกผันของเอกสารเว็บเซอร์วิส (Inverse Document Frequency: idf) โดยสมการที่ใช้ในการคำนวณค่าน้ำหนักความถี่ผกผัน ดังสมการที่ 1

$$idf_k = \log \left( \frac{N}{n_k} \right) \quad \text{สมการ (1)}$$

โดยที่  $idf_k$  แทนด้วยค่าความถี่ผกผันของเอกสารเว็บเซอร์วิสของดัชนีตัวที่  $k$  เมื่อ  $k \in \{1 \dots M\}$

$M$  คือ จำนวนดัชนีทั้งหมด,

$N$  แทนด้วยจำนวนเอกสารเว็บเซอร์วิสทั้งหมด,

$n_k$  แทนด้วยจำนวนเอกสารเว็บเซอร์วิสที่เจอดัชนีตัวที่  $k$  ปรากฏอยู่ และสำหรับการหาค่าถ่วง

น้ำหนักของดัชนีเว็บเซอร์วิสสามารถหาได้จากสมการที่ 2.

$$w_{i,k} = tf_k * idf_k \quad \text{สมการ (2)}$$

โดยที่  $w_{i,k}$  แทนด้วยค่าถ่วงน้ำหนักของดัชนีตัวที่  $k$  ในเอกสารเว็บเซอร์วิสลำดับที่  $i$

$tf_k$  แทนด้วยค่าความถี่ของดัชนีตัวที่  $k$  ในเอกสารเว็บเซอร์วิสลำดับที่  $i$  ซึ่ง  $i \in \{1 \dots N\}$  และ  $k \in \{1 \dots M\}$

ตารางที่ 1 แสดงตัวอย่างการหาค่าความถี่และค่าความถี่ผกผันของดัชนี 4 ตัวใน 5 เว็บเซอร์วิส ประกอบด้วยดัชนี government, scholarship, degree, german และเว็บเซอร์วิส S1, S2, S3, S4, S5

ตารางที่ 1. ตัวอย่างค่าความถี่และค่าผกผันของดัชนี

| ดัชนี/เว็บเซอร์วิส | ค่าความถี่ |    |    |    |    |   | ค่าความถี่ผกผัน |
|--------------------|------------|----|----|----|----|---|-----------------|
|                    | S1         | S2 | S3 | S4 | S5 | Q |                 |
| government         | 1          | 1  | 1  | 0  | 2  | 1 | 0.09691         |
| scholarship        | 1          | 0  | 0  | 1  | 2  | 1 | 0.22184         |
| degree             | 2          | 3  | 2  | 1  | 0  | 1 | 0.09691         |
| german             | 0          | 1  | 0  | 3  | 0  | 1 | 0.39794         |

ตารางที่ 2 แสดงตัวอย่างค่าความถ่วงน้ำหนักดัชนีของ 5 เว็บเซอร์วิสจากข้อมูลในตารางที่ 1

ตารางที่ 2. ตัวอย่างค่าถ่วงน้ำหนักของดัชนี

| ดัชนี/เว็บเซอร์วิส | ค่าถ่วงน้ำหนักของดัชนี |         |         |         |         |          |
|--------------------|------------------------|---------|---------|---------|---------|----------|
|                    | S1                     | S2      | S3      | S4      | S5      | Q        |
| government         | 0.09691                | 0.09691 | 0.09691 | 0       | 0.19382 | 0.09691  |
| scholarship        | 0.22184                | 0       | 0       | 0.22184 | 0.44369 | 0.221849 |
| degree             | 0.19382                | 0.29073 | 0.19382 | 0.09691 | 0       | 0.09691  |
| german             | 0                      | 0.39794 | 0       | 1.19382 | 0       | 0.39794  |

การหาค่าความคล้ายคลึงระหว่างเอกสารที่เป็นตัวแทนเว็บเซอร์วิสและคำสอบถามของผู้ใช้

วิธีการเปรียบเทียบเวกเตอร์ระหว่างเวกเตอร์เอกสารกับเวกเตอร์คำสอบถาม สามารถทำได้โดยการคำนวณค่าความคล้ายคลึงระหว่างเอกสารและคำสอบถาม (โกรคักดี เกสร, 2554) โดยทำการหาผลคูณภายใน (Dot Product) ระหว่างเวกเตอร์ โดยผลลัพธ์ที่ได้จะเป็นค่าตัวเลขที่บ่งบอกถึงระดับความคล้ายคลึงระหว่างเอกสารและคำสอบถาม ด้วยสมการที่ 3.

$$sim(q, d_i) = \frac{\sum_{k=1}^N (w_{q,k}) * (w_{i,k})}{\sqrt{\sum_{k=1}^N (w_{q,k})^2 * \sum_{k=1}^N (w_{i,k})^2}} \quad \text{สมการ (3)}$$

เมื่อ  $sim(q, d_i)$  แทนด้วย ค่าความคล้ายคลึงระหว่างเอกสารกับคำสอบถาม

$(w_{i,k})$  แทนด้วย ค่าถ่วงน้ำหนักของดัชนีที่  $k$  ในเอกสารลำดับที่  $i$  , โดยที่  $i \in \{1 \dots N\}$

และ  $N$  แทนด้วย จำนวนเอกสารทั้งหมด

$(w_{q,k})$  แทนด้วย ค่าถ่วงน้ำหนักของดัชนีที่  $k$  ในคำสอบถาม

สมมุติให้คำสอบถามประกอบด้วยคำสำคัญ คือ german, government, degree, scholarship จากสมการที่ 3 สามารถคำนวณหาความคล้ายคลึงระหว่างแต่ละเว็บเซอร์วิสและคำสอบถามได้ดังต่อไปนี้

$$Q = \sqrt{(0.09691 * 0.09691) + (0.22184 * 0.22184) + (0.09691 * 0.09691) + (0.39794)} = 0.4576$$

$$S1 = \frac{(0.09691*0.19382)+(0.22184*0.2218487)+(0.09691*0.19382)}{\sqrt{((0.09691*0.09691)+(0.22184*0.22184)+(0.19382*0.19382))*0.4576}} = 0.524526$$

$$S2 = \frac{(0.09691*0.09691)+(0.09691*0.29073)+(0.39794*0.39794)}{\sqrt{((0.09691*0.09691)+(0.29073*0.29073)+(0.39794*0.39794))*0.4576}} = 0.819886$$

$$S3 = \frac{(0.09691*0.09691)+(0.09691*0.19382)}{\sqrt{((0.09691*0.09691)+(0.19382*0.19382))*0.4576}} = 0.273281$$

$$S4 = \frac{(0.22184*0.2218487)+(0.09691*0.09691)+(0.39794*0.19382)}{\sqrt{((0.22184*0.22184)+(0.09691*0.09691)+(0.19382*0.19382))*0.4576}} = 0.9208581$$

$$S5 = \frac{(0.09691*0.19382)+(0.22184*0.4436975)}{\sqrt{((0.19382*0.19382)+(0.44369*0.44369))*0.4576}} = 0.508843$$

ดังนั้นจึงทำการจัดลำดับของผลลัพธ์ที่ได้ของค่าความคล้ายคลึงได้ดังตารางที่ 3

ตารางที่ 3. ลำดับของผลลัพธ์จากการสืบค้น

| ลำดับ | เว็บเซอร์วิส | ค่าความคล้ายคลึง |
|-------|--------------|------------------|
| 1     | S1           | 0.920858         |
| 2     | S2           | 0.819886         |
| 3     | S3           | 0.524526         |
| 4     | S4           | 0.508843         |
| 5     | S5           | 0.273281         |

### ผลการวิจัย

จากภาพรวมของระบบการสืบค้นเว็บเซอร์วิสเชิงความหมายด้วยดัชนีที่กล่าวถึงข้างต้น การทดสอบแนวคิดดังกล่าวเป็นการทดลองในขั้นต้นเพื่อประเมินประสิทธิภาพการสืบค้นกับชุดคำสอบถามที่กำหนดโดยผู้เชี่ยวชาญ โดยข้อมูลที่เกี่ยวข้องกับการทดลองมีดังนี้ งานวิจัยนี้จึงทำการประเมินผลการสืบค้นด้วยการวัดค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยความแม่นยำ ซึ่งแสดงได้ดังสมการที่ 4, 5 และ 6 ตามลำดับ

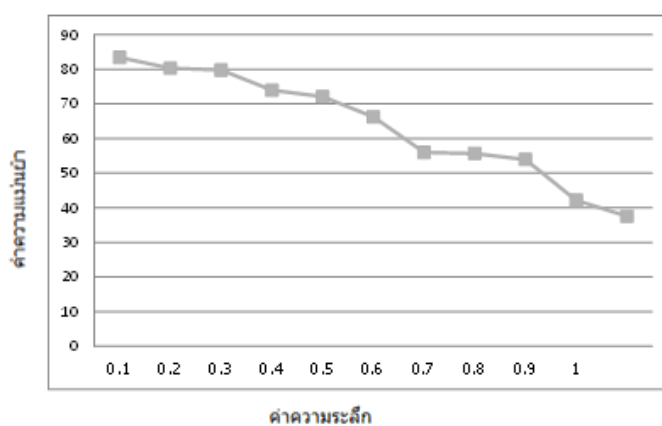
$$\text{ค่าความแม่นยำ (P)} = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารทั้งหมดที่สืบค้นออกมาได้}} \quad \text{สมการ (4)}$$

$$\text{ค่าความเที่ยงตรง (R)} = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล}} \quad \text{สมการ (5)}$$

$$\text{ค่าเฉลี่ยความแม่นยำ } (\bar{P}(r)) = \sum_{i=1}^{Nq} P_i(r) / Nq \quad \text{สมการ (6)}$$

โดยมีข้อมูลที่ใช้ในการทดสอบดังต่อไปนี้

1. เอกสารเว็บเซอร์วิสเชิงความหมาย จากโดเมนการศึกษา (Education Domain) ใน OWLS-TC version 4.0 ซึ่งในงานวิจัยนี้ได้นำเอาข้อมูลจากแหล่งดังกล่าวมาใช้ในการทดลอง สำหรับรายละเอียดเพิ่มเติมสามารถศึกษาได้จาก (Matthias et al, 2010)
2. คำสอบถาม ที่ใช้ในการทดลองนี้ทั้งสิ้น 30 คำสอบถาม
3. ดัชนีเว็บเซอร์วิสจำนวน 170 ดัชนี จากการพิจารณาเอกสารเว็บเซอร์วิสเชิงความหมายทั้งหมดจำนวน 137 เอกสาร



ภาพที่ 7 แสดงผลการทดลองการใช้ดัชนีในการสืบค้นเว็บเซอร์วิสเชิงความหมาย

จากภาพที่ 7 ชี้ให้เห็นว่าในระดับค่าความระลึกที่เพิ่มมากขึ้น ค่าความแม่นยำมีแนวโน้มลดลงอย่างต่อเนื่อง โดยมีค่าเฉลี่ยความแม่นยำเท่ากับ ร้อยละ 63.60 จากจำนวนคำสอบถามที่ใช้ในการทดลองในเบื้องต้นนี้ 30 คำสอบถาม ซึ่งในแต่ละระดับค่าความระลึกที่มากขึ้นนั้นมีค่าความแม่นยำดังแสดงในตารางที่ 4

| R | 0     | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   | 0.9   | 1.0   |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| P | 82.03 | 80.29 | 79.74 | 73.93 | 72.08 | 66.25 | 55.99 | 55.65 | 53.92 | 42.15 | 37.57 |

### สรุปและอภิปรายผล

ระบบการสืบค้นเว็บเซอร์วิสเชิงความหมายที่พัฒนาขึ้นนี้เกิดจากแนวคิดที่ต้องการใช้ประโยชน์ข้อมูลจากโครงสร้างข้อมูลทุกส่วนของเอกสารโอดับบลิวแอลเอส เพื่อประโยชน์ในการสืบค้นเว็บเซอร์วิสเชิงความหมาย โดยนำข้อมูลที่ได้ไปสร้างดัชนี พร้อมทั้งหาค่าถ่วงน้ำหนักของแต่ละดัชนีเพื่อใช้ในการเปรียบเทียบระหว่างคำสอบถามจากผู้ใช้งานกับแต่ละเอกสารเว็บเซอร์วิสเพื่อหาค่าความคล้ายคลึง โดยผลจากการทดลองในเบื้องต้นชี้ให้เห็นว่าระบบดังกล่าวมีค่าเฉลี่ยความแม่นยำอยู่ที่ร้อยละ 63.60

นอกจากนั้น ผลการทดลองยังแสดงให้เห็นว่าค่าเฉลี่ยความแม่นยำมีแนวโน้มลดลงอย่างต่อเนื่องเมื่อระดับค่าความระลึกเพิ่มมากขึ้น ซึ่งแนวโน้มดังกล่าวสอดคล้องกับ (Ricardo and Berthier, 2011). ที่กล่าวถึงระบบการสืบค้นที่ให้แนวโน้มของค่าเฉลี่ยความแม่นยำในลักษณะนี้เหมาะสมสำหรับการสืบค้นข้อมูลในรูปแบบเอกสารหรือเว็บไซต์

สำหรับการพัฒนางานวิจัยต่อไปในอนาคตจะได้มีการขยายผลการทดลองโดยมีการเพิ่มจำนวนคำ  
สอบถามในการทดลองให้มากขึ้นเพื่อพิจารณาแนวโน้มของค่าความแม่นยำ และทำการทดสอบประสิทธิภาพของ  
ระบบเทียบกับการขยายคำสอบถามด้วยวิธีการเพิ่มคำสำคัญจากผู้ใช้งานในคำสอบถาม

#### เอกสารอ้างอิง

##### ภาษาไทย

- ไกรศักดิ์ เกสร. (2554). “การค้นหาข้อมูลเชิงความหมาย: แนวคิดใหม่ของโปรแกรมการค้นหา (Search engine)และแนวทางการพัฒนาในอนาคต”. วไลยอลงกรณ์ปริทัศน์1,2 (กรกฎาคม-ธันวาคม):1-16.  
จตุพล โพธิ์คำพุก และดร.สมจิตร อาจอินทร์. (2555). “การจัดกลุ่มเว็บเซอร์วิสเพื่อการสืบค้นเชิงความหมาย”.  
วิจัย มข (บศ) 12,4 (ตุลาคม – ธันวาคม): 1-11.  
สมเกียรติ สอนนวล และคณะ. (2556). “การสืบค้นเว็บเซอร์วิสแบบสื่อความหมายโดยใช้ซอฟต์แวร์เอเจนต์”  
ใน **Proceedings of 2013 International Computer Science AndEngineering  
Conference**, 447-453. โรงแรมวินเซอร์, 4-6, กันยายน 2556. กรุงเทพฯ

##### ภาษาต่างประเทศ

- Adam Czyszczon and Aleksander Zgrzywa. (2014). “Indexing method for effective web service  
retrieval”. **International Journal of Intelligent Information and Database Systems**,  
8,3, (December): 189-208.
- Bo Zhou, Tinglei Huang, Jie Liu, and MeizhouShen. (2009). “Using Inverted Indexing to Semantic  
Web Service Discovery Search Model”. In **Proceeding of 5<sup>th</sup> IEEE International  
Conference on Wireless Communications, Networking and Mobile Computing**,  
1-4. Beijing, September 24-26, 2009.
- David Martin, Anupriya Ankolekar, Mark Burstein, Grit Demker, Daniel Elenius, Jerry Hobbs,  
Lalana Kagal, Ora Lassila, Drew McDermott, Deborah McGuinness, Shelia McIlraith,  
Massimo Paolucci, Bijan Parsia, Terry Payne, Marta Sabou, Evren Sirn, Monika Solanki,  
Naveen Srinivasan, and Katia Sycard. (2008). **OWL-S 1.2** . Accessed January 12.  
Available from <http://www.daml.org/services/owl-s/1.2/>
- Debajyoti Mukopadhyay and Archana Chougule. (2012). “A Survey on Web Service Discovery  
Approaches”, In **Proceedings of the Second International Conference on  
Computer Science, Engineering and Applications (ICCSEA 2012)**, 1001-1012,  
May 25-27, ,New Delhi, India.
- K. Tamilarasi and Dr. M. Ramakrishnan. (2015). “Indexing Traditional UDDI for Efficient Discovery  
of Web Services”. **Indian Journal of Computer Science and Engineering (IJCSE)**, 6,  
1, (February – March): 14-20.

- Lextek International. (1993). **English Stop Word List**. Accessed December 25. Available from <http://www.lextek.com/manuals/onix/stopwords1.html>.
- Marco Aiello, Christian Platzner, Florian Rosenberg, Huy Tran, Martin Vasko, and Schahram Dustdar. “Web Service Indexing for Effective Retrieval and Composition”. In **Proceedings of the IEEE International Conference on E-Commerce Technology and the 3<sup>rd</sup> IEEE International Conference on Enterprise Computing, E-Commerce, and E-Service (CEC/EEE’06)**, 1-8, San Francisco, CA, June 26-29 2006.
- Matthias Klusch, Patrick Kapahnke, Stefan Schutle, Freddy Lecue and Abraham Bernstein. (2015). “Semantic Web Service Search: A Brief Survey”. **Journal on Artificial Intelligence**, 30 ,2, (June): 139-147.
- Matthias Klusch, Mahboob Alam Khalid, Patrick Kapahnke, Benedikt Fries, and Martin Vasileski. (2010). **OWLS-TC**, Accessed January 3. Available from <http://projects.semwebcentral.org/projects/owls-tc>
- Ricardo Baeza-Yates, and Berthier Ribeiro-Neto. (2011). **Modern Information Retrieval**. 2nd ed, Boston, MA, USA: Addison Wesley Published.
- Torben Schreiter. (2007). **An introduction into Semantic Web services**. Access July 25. Available from <http://blog.schreiter.info/wp-content/uploads/2007/03/an-introduction-into-semantic-web-services-torben-schreiter.pdf>