การพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย

นางสาวนวรัตน์ คำเสียง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย
บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository(CUIR)
are the thesis authors' files submitted through the Graduate School.

A DEVELOPMENT OF A TOOL FOR COMMON RESEARCH INTERESTS ANALYSIS FOR RESEARCHERS

Miss Nawarat Kamsiang

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย
โดย	นางสาวนวรัตน์ คำเสียง
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

……. ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.วีระ เหมืองสิน)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา)

……. กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.กฤดาภัทร สีหารี)

นวรัตน์ คำเสียง : การพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย. (A DEVELOPMENT OF A TOOL FOR COMMON RESEARCH INTERESTS ANALYSIS FOR RESEARCHERS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา, 96 หน้า.

วิทยานิพนธ์นี้นำเสนอระเบียบวิธีและพัฒนาเครื่องมือเพื่อวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย ซึ่งการวิเคราะห์จะมีประโยชน์สำหรับนักวิจัยในการสร้างความร่วมมือกัน รวมทั้งสามารถช่วยในการระบุด้านของงานวิจัยและระดับของความสนใจที่นักวิจัยมีความสนใจร่วมกัน ในการนี้ผู้วิจัยใช้คำสำคัญจากผลงานวิจัยที่ได้รับการตีพิมพ์และจัดทำดัชนีโดยฐานข้อมูลไอเอสไอเพื่อสร้างออนโทโลจีโปรไฟล์งานวิจัยสำหรับนักวิจัย ซึ่งแนวคิดการสร้างออนโทโลจีข้อมูลงานวิจัย และวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยได้ประยุกต์วิธีการสร้างออนโทโลจีและการจับคู่ออนโทโลจีที่มีอยู่แล้ว โดยการเปรียบเทียบความเหมือนกันระหว่างออนโทโลจีโปรไฟล์งานวิจัยจะพิจารณาจากความเหมือนกันของสายอักขระและความเหมือนกันในเชิงภาษาของคำที่ปรากฏในสองโปรไฟล์ นอกจากนี้ผู้วิจัยได้เพิ่มแนวคิดเกี่ยวกับน้ำหนักความลึกสำหรับการจับคู่ออนโทโลจี โดยที่น้ำหนักความลึกของคู่คำที่ถูกจับคู่จะถูกกำหนดโดยระยะทางจากรากของออนโทโลจีจนถึงตำแหน่งของคู่คำที่ปรากฏในออนโทโลจี เนื่องจากคู่คำที่เหมือนกันและอยู่ลึกลงไปในระดับล่างของออนโทโลจีจะได้รับความสนใจมากกว่าคู่คำที่อยู่ในระดับบน เพราะถือว่าเป็นคู่คำที่แสดงความสนใจเฉพาะทางมากกว่า การวัดประสิทธิภาพระหว่างระเบียบวิธีที่นำเสนอกับวิธีการจับคู่ออนโทโลจีที่มีอยู่แล้วจะใช้มาตรฐานของโอเออีไอ 2011 ซึ่งผลที่ได้พบว่าแนวคิดของน้ำหนักความลึกจะให้ค่าความแม่นยำที่ดีขึ้น แต่ค่าเรียกคืนจะลดลง

ภาควิชา...วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต
สาขาวิชา..วิทยาศาสตร์คอมพิวเตอร์..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....
ปีการศึกษา..2554.....

5271429921 : MAJOR COMPUTER SCIENCE

KEYWORDS : ONTOLOGY BUILDING / ONTOLOGY MATCHING / COMMUNITY OF PRACTICE

NAWARAT KAMSIANG : A DEVELOPMENT OF A TOOL FOR COMMON RESEARCH INTERESTS ANALYSIS FOR RESEARCHERS. ADVISOR : ASSOC. PROF. TWITTIE SENIVONGSE, Ph.D., 96 pp.

This thesis presents a methodology and develops a tool to analyze common research interest for researchers. The analysis can be useful for researchers in establishing further collaboration as it can help to identify the areas and degree of interest that any two researchers share. Using keywords from the publications indexed by ISI Web of Knowledge, we build ontological research profiles for the researchers. Our methodology builds upon existing approaches to ontology building and ontology matching in which the comparison between research profiles is based on string similarity and linguistic similarity between the terms in the two profiles. In addition, we add the concept of depth weights to ontology matching. A depth weight of a pair of matched terms is determined by the distance of the terms from the root of their ontologies. We see that more attention should be paid to the matched pair that are located near the bottom of the ontologies than to the matched pair that are near the root, since they are more specialized areas of interest. A comparison between our methodology and an existing ontology matching approach, using OAEI 2011 benchmark, shows that the concept of depth weights gives better precision but lower recall.

Department : Computer Engineering…….. Student's Signature .....

Field of Study : Computer Science…….. Advisor's Signature .....

Academic Year : 2011……

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบพระคุณ รองศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่เสียสละเวลาช่วยเหลือให้คำปรึกษา คำแนะนำและข้อคิดเห็นที่มีประโยชน์ทำให้การจัดทำวิทยานิพนธ์ให้สำเร็จลุล่วงไปด้วยดี ขอขอบพระคุณคณาจารย์ทุกท่านที่แนะนำสั่งสอน และให้ความรู้แก่ข้าพเจ้าตลอดระยะเวลาการศึกษา

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.วีระ เหมืองสิน ประธานกรรมการสอบวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ ดร.กฤดาภัทร สีหารี กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำต่างๆ ทำให้วิทยานิพนธ์นี้มีความถูกต้องและสมบูรณ์มากขึ้น

ขอขอบคุณมหาวิทยาลัยศรีปทุม วิทยาเขตชลบุรี ที่สนับสนุนในด้านค่าใช้จ่ายในการศึกษาจนสำเร็จลุล่วงได้

ขอขอบพระคุณ คุณพ่อ คุณแม่ ที่ให้ความรัก ความห่วงใย คอยให้กำลังใจ พร้อมทั้งให้ความช่วยเหลือ และสนับสนุนในด้านค่าใช้จ่ายในการศึกษาจนสำเร็จลุล่วงได้

ขอขอบคุณพี่ ๆ ที่สำนักงานบริการคอมพิวเตอร์ มหาวิทยาลัยศรีปทุม วิทยาเขตชลบุรี และเพื่อนๆ ทุกคน สำหรับคำปรึกษาที่ดีในทุกๆ ด้าน รวมทั้งกำลังใจและความช่วยเหลืออื่นๆ ที่มอบให้มาโดยตลอด

งานวิทยานิพนธ์นี้ ข้าพเจ้าหวังเป็นอย่างยิ่งว่าจะเป็นประโยชน์ต่อผู้ศึกษา ค้นคว้าและสนใจ หากผิดพลาดประการใด ข้าพเจ้าขอน้อมรับไว้เพื่อนำไปใช้ในการปรับปรุงให้วิทยานิพนธ์นี้มีความสมบูรณ์ยิ่งขึ้น สำหรับความดีที่ได้รับจากวิทยานิพนธ์นี้ข้าพเจ้าขอมอบให้แก่ผู้มีพระคุณทุกท่าน

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฎ

บทที่

1 บทนำ..... **1**

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 2

1.3 ขอบเขตของการวิจัย..... 3

1.4 ขั้นตอนการวิจัย..... 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ..... 4

1.6 ผลงานตีพิมพ์..... 4

2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... **5**

2.1 แนวคิดและทฤษฎี..... 5

2.1.1 ออนโทโลจี (Ontology)..... 5

2.1.2 เวิร์ดเน็ต (Wordnet)..... 6

2.1.3 การวัดค่าความคล้ายคลึงกันของสายอักขระ (String Similarity)..... 7

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 9

2.2.1 การสร้างออนโทโลจี (Ontology Building)..... 9

2.2.2 การจับคู่ออนโทโลจี (Ontology Matching)..... 11

2.2.3 การหาความสัมพันธ์แบบ Hyponym ในเวิร์ดเน็ต..... 16

2.2.4 การกำหนดระยะห่างสำหรับการจับคู่เว็บเซอร์วิซเชิงความหมาย..... 16

2.2.5 การวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยเรื่องเดียวกัน..... 17

บทที่ หน้า

3 ระเบียบวิธีวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย..... 21
3.1 กำหนดแหล่งข้อมูลและขอบเขตข้อมูลของนักวิจัยที่จะนำมาสร้างออนโทโลจี..... 21
3.2 กระบวนการสร้างออนโทโลจี..... 23
3.3 กระบวนการจับคู่ความคล้ายคลึงกันระหว่างออนโทโลจีนักวิจัย..... 28
4 การพัฒนาเครื่องมือ..... 50
4.1 โครงสร้างของเครื่องมือ..... 50
4.1.1 ส่วนจัดเก็บข้อมูลนักวิจัย..... 51
4.1.2 ส่วนสร้างออนโทโลจีนักวิจัย..... 52
4.1.3 ส่วนจับคู่ออนโทโลจีนักวิจัย..... 55
4.2 สภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ..... 58
4.2.1 ฮาร์ดแวร์..... 58
4.2.2 ซอฟต์แวร์..... 59
5 ผลการทดสอบ..... 60
5.1 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่เหมือนกัน..... 60
5.2 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่ใกล้เคียงกัน..... 72
5.3 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่แตกต่างกัน..... 76
5.4 การวัดประสิทธิภาพของการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย..... 81
5.5 การทดสอบประโยชน์ของแนวคิด MLMA+ with Depth Weights..... 84
6 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 90
6.1 สรุปผลการวิจัย..... 90
6.2 อภิปรายผลการวิจัย..... 91
6.3 ข้อเสนอแนะ..... 92
รายการอ้างอิง..... 93
ประวัติผู้เขียนวิทยานิพนธ์..... 96

สารบัญตาราง

ตารางที่		หน้า
2.1	ค่า Score Value สำหรับแต่ละสเตทเพื่อนบ้าน……	15
2.2	ตัวอย่างข้อมูลส่วนตัวที่ประกอบด้วยชุดของแท็กหรือคำสำคัญ……..	18
3.1	ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+……...	35
3.2	ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ with Depth Weights……	45
3.3	การเปรียบเทียบ Matched Pairs ระหว่างแนวคิด MLMA+ และ MLMA+ with Depth Weights……..……..	48
3.4	ผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกัน……	49
5.1	จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA……	60
5.2	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA……	63
5.3	จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Sudsang, A……..	72
5.4	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A……..	74
5.5	จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ruppin, E……	77
5.6	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E……	78
5.7	ผลการวัดประสิทธิภาพของการจับคู่ออนโทโลจี……..	83
5.8	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Einwohner, RL บนพื้นฐาน MLMA+……..	86
5.9	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Einwohner, RL บนพื้นฐาน MLMA+ with Depth Weights……	86
5.10	ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Howard, JH บนพื้นฐาน MLMA+……..	87

ตารางที่ หน้า

5.11 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Howard, JH บนพื้นฐาน MLMA+ with Depth Weights…….. 88

5.12 เปรียบเทียบผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกัน บนพื้นฐาน MLMA+ และ MLMA+ with Depth Weights…….... 88

สารบัญภาพ

ภาพที่		หน้า
2.1	ตัวอย่างข้อมูลเค้าร่างอธิบายรายละเอียดคลาสคอมพิวเตอร์……..	6
2.2	ตัวอย่างข้อมูลอินสแตนซ์อธิบายรายละเอียดคลาสคอมพิวเตอร์……	6
2.3	ชุดคำศัพท์ที่ได้จาก synset_id ของคำว่า computer	7
2.4	การสร้างออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุทิศทาง……	10
2.5	ตัวอย่างคอมพิวเตอร์ออนโทโลจี……...	12
2.6	เมตริกซ์ L และเมตริกซ์ $Map_{0\text{-}1}$ที่ได้จากการเปลี่ยน O_1 และ O_2……..	12
2.7	ตัวอย่างวิธีการคำนวณค่า Score Value ของสเตท St_{n1}……..	15
3.1	ข้อมูลนักวิจัยในรูปแบบเท็กซ์ไฟล์ที่ได้จากฐานข้อมูลไอเอสไอ……..	22
3.2	ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Kijsirikul, B……...	24
3.3	ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Ratanamahatana, CA……	24
3.4	ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Sudsang, A……..	27
3.5	ออนโทโลจี S……	28
3.6	ออนโทโลจี T……	29
3.7	เมตริกซ์ L ที่ได้จากการเปลี่ยนออนโทโลจี S และ T บนพื้นฐาน MLMA+……..	31
3.8	เมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+……..	32
3.9	สเตทเพื่อนบ้านของเมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+…….....	34
3.10	เมตริกซ์ L ที่ได้จากการเปลี่ยนออนโทโลจี S และ T บนพื้นฐาน MLMA+ with Depth Weights……..……	43
3.11	เมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+ with Depth Weights…….....	44
4.1	แผนภาพแสดงส่วนประกอบโครงสร้างของเครื่องมือ……..	50
4.2	หน้าจอสำหรับการอัพโหลดเท็กซ์ไฟล์……..……	51
4.3	หน้าจอสำหรับจัดเก็บข้อมูลนักวิจัย……..	52
4.4	ส่วนต่อประสานกับผู้ใช้สำหรับสร้างและจับคู่ออนโทโลจี……..	53
4.5	ออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B เมื่อใช้คำสำคัญเริ่มต้น 5 คำ…….	54
4.6	ออนโทโลจีโปรไฟล์งานวิจัยของ Ratanamahatana, CA เมื่อใช้คำสำคัญเริ่มต้น 5 คำ……..……	54
4.7	ออนโทโลจีโปรไฟล์งานวิจัยของ Sudsang, A เมื่อใช้คำสำคัญเริ่มต้น 5 คำ…...	55

ภาพที่		หน้า
4.8	ออนโทโลจีโปรไฟล์งานวิจัยของ Ruppin, E เมื่อใช้คำสำคัญเริ่มต้น 5 คำ……..	55
4.9	เมตริกซ์ L และค่าสัมประสิทธิ์ความคล้ายคลึงกันระหว่างเทอมในออนโทโลจีโปรไฟลงานวิจัย……	56
4.10	ช่องสำหรับกรอกค่าขีดแบ่งเริ่มต้นความคล้าย…….....	56
4.11	เมตริกซ์ $Map_{0\text{-}1}$ ที่ได้จากการเปรียบเทียบค่าในเมตริกซ์ L กับค่าขีดแบ่งเริ่มต้นความคล้าย…….....	57
4.12	ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านระหว่างออนโทโลจีโปรไฟล์งานวิจัย 2 โปรไฟล์……	57
4.13	ค่าความสอดคล้องกันของนักวิจัย……	58
5.1	เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA……	62
5.2	เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA……	71
5.3	เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Sudsang, A…….....	73
5.4	เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A……..	76
5.5	เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ruppin, E……	77
5.6	เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E……	81
5.7	ออนโทโลจีโปรไฟล์งานวิจัยของ Einwohner, RL เมื่อใช้คำสำคัญเริ่มต้น 5 คำ..	85
5.8	ออนโทโลจีโปรไฟล์งานวิจัยของ Howard, JH เมื่อใช้คำสำคัญเริ่มต้น 5 คำ…..	85

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันองค์ความรู้มักจะถูกจัดเก็บไว้ในรูปแบบต่าง ๆ ได้แก่ เอกสาร เอกสารอิเล็ก-ทรอนิกส์ สื่อทางอิเล็กทรอนิกส์อื่น ๆ เช่น เทป วิดีโอ ซีดี และที่สำคัญคือการเผยแพร่องค์ความรู้ผ่านเว็บไซต์ แต่องค์ความรู้โดยส่วนใหญ่จะถูกจัดเก็บอยู่ในสมองของมนุษย์ในรูปแบบของประสบการณ์และความจำ แต่ในยุคของเทคโนโลยีสารสนเทศและการสื่อสารทำให้องค์ความรู้ทางด้านต่าง ๆ พัฒนาความก้าวหน้าไปอย่างรวดเร็ว เราจำเป็นต้องมีความรู้ที่หลากหลายมากขึ้น เนื่องจากปัญหาที่ประสบจะมีความซับซ้อนมากขึ้น จึงจำเป็นต้องก้าวสู่การทำงานที่มีการเชื่อมโยงเป็นเครือข่ายระหว่างผู้เชี่ยวชาญที่เกิดขึ้นด้วยความเต็มใจที่จะแลกเปลี่ยนความรู้ซึ่งกันและกัน

จากปัญหาดังกล่าวทำให้เกิดชุมชนนักปฏิบัติ (Community of Practice) [1-2] ซึ่งเป็นกลุ่มบุคคลที่แต่ละคนมีความสนใจในเรื่องเดียวกัน ต้องการแลกเปลี่ยนความรู้และประสบการณ์ซึ่งกันและกัน อาจจะพบกันด้วยตัวจริง หรือผ่านเครือข่ายอินทราเน็ต อินเทอร์เน็ต รวมถึงมีวิธีปฏิบัติคล้ายกัน ใช้เครื่องมือในการทำงานและภาษาเดียวกัน มีเป้าหมายในการทำงานร่วมกัน มีความมุ่งมั่นร่วมกันที่จะพัฒนาวิธีการทำงานให้มีประสิทธิภาพมากขึ้น และหากองค์กรสามารถเชื่อมโยงชุมชนนักปฏิบัติเข้าด้วยกันได้ จะทำให้เกิดเครือข่ายทางสังคม (Social Network) ที่มีการปฏิบัติงานร่วมกันจากหลาย ๆ ฟังก์ชันงาน การรวมกลุ่มกันของชุมชนนักปฏิบัติจึงมีประโยชน์ต่อการทำงานในด้านต่าง ๆ โดยเฉพาะงานด้านวิจัย ช่วยให้นักวิจัยมีการพัฒนาความรู้ใหม่ ๆ ที่สามารถนำไปประยุกต์ใช้ในการทำงานได้ ทำให้รู้จักนักวิจัยอื่น ๆ ที่มีความสนใจหลายสิ่งหลายอย่างที่คล้ายคลึงกัน รวมทั้งอาจประสบปัญหาที่คล้ายกันเมื่อได้แลกเปลี่ยน ประสบการณ์ซึ่งกันและกันจะทำให้สามารถค้นพบวิธีแก้ปัญหาได้ และได้พัฒนาตนเองอย่างไม่มีที่สิ้นสุด

ในวงการวิจัยได้มีการนำบิบลิโอเมตริกส์ (Bibliometrics) [3] มาช่วยในการวิเคราะห์เชิงคณิตศาสตร์และสถิติเพื่อเชื่อมโยง เปรียบเทียบ สิ่งพิมพ์หรือวรรณกรรมในสาขาวิชาวิทยาศาสตร์หรือสาขาอื่น ๆ เพื่อประเมินคุณภาพ ภาพรวมของนักวิจัย งานวิจัย หรือความสัมพันธ์ระหว่างนักวิจัยและงานวิจัย ซึ่งรูปแบบของการวิเคราะห์ดัชนีวรรณกรรมสามารถแบ่งออกเป็น 2 ประเภท ได้แก่ ดัชนีเชิงปริมาณ ประกอบด้วย จำนวนบทความ จำนวนบทความที่ถูกอ้างอิง และจำนวนบทความที่มีผู้วิจัยร่วม เป็นต้น และดัชนีเชิงสัมพันธ์ ประกอบด้วย ดัชนี ผลงานตีพิมพ์ร่วมกัน ดัชนีการแลกเปลี่ยนความรู้ และดัชนีอ้างอิงบทความร่วมกัน เป็นต้น นอกจากนี้ยังมีการวิเคราะห์ดัชนีวรรณกรรมในแนวทางอื่นที่มีลักษณะใกล้เคียงกันคือการวิเคราะห์โปรไฟล์งานวิจัย โดย

สามารถสร้างโปรไฟล์งานวิจัยจากการเก็บรวบรวมข้อมูลหรือการทำเหมืองข้อมูลจากแหล่งข้อมูลทางอิเล็กทรอนิกส์ เช่น เว็บไซต์ สิ่งตีพิมพ์ บล็อก ข้อมูลส่วนบุคคล และเอกสารโครงการวิจัย เป็นต้น และจากวิธีการที่กล่าวมาจะทำให้เกิดการค้นพบความเชี่ยวชาญของนักวิจัยและความสอดคล้องกันเชิงความหมายระหว่างโปรไฟล์นักวิจัย

ผู้วิจัยจึงมีแนวคิดที่จะนำเสนอระเบียบวิธีและพัฒนาเครื่องมือเพื่อวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย เพื่อวิเคราะห์หรือจับคู่นักวิจัยที่สนใจในงานวิจัยที่เป็นเรื่องเรื่องเดียวกัน หรือแนวทางเดียวกัน ซึ่งจะช่วยให้นักวิจัยสามารถทำงานวิจัยร่วมกัน โดยการนำนำออนโทโลจี (Ontology) เข้ามาช่วยในการดำเนินการ เนื่องจากแนวคิดเรื่องออนโทโลจีถูกนำมาใช้งานอย่างแพร่หลายในงานวิจัยด้านเว็บเชิงความหมาย (Semantic Web) การจัดการองค์ความรู้ (Knowledge Management) และการค้นคืนสารสนเทศ (Information Retrieval) เป็นต้น การพัฒนาเครื่องมือดังกล่าวได้นำแนวความคิดของ An และคณะ [4] มาสร้างออนโทโลจีความสนใจด้านงานวิจัยของนักวิจัยในรูปของกราฟ โดยวิเคราะห์จากคำสำคัญ (Keyword) ซึ่งเป็นข้อมูลการตีพิมพ์ผลงานของนักวิจัยจากฐานข้อมูลไอเอสไอ จากนั้นนำออนโทโลจี 2 ออนโทโลจีมาหาค่าความคล้ายคลึงกัน โดยใช้วิธีของ Alasoud และคณะ [5-6] ซึ่งใช้การวัดค่าความคล้ายคลึงกันร่วมกัน 2 วิธี ได้แก่ การหาค่าความเหมือนกันของสายอักขระ (String Similarity) และการหาค่าความเหมือนกันเชิงภาษา (Linguistic Similarity) ของเวิร์ดเน็ต (Wordnet) ในขั้นตอนนี้ผู้วิจัยได้เพิ่มการพิจารณาน้ำหนักความลึก (Depth Weight) เข้าไปในแต่ละระดับของออนโทโลจีนักวิจัยด้วย เพื่อแสดงให้เห็นว่าคำศัพท์ที่อยู่ลึกลงไปในระดับล่างของออนโทโลจีจะเป็นคำศัพท์ที่เฉพาะเจาะจง และมีความสัมพันธ์ใกล้เคียงกันมากกว่าคำศัพท์ที่อยู่ ระดับบนของออนโทโลจี [7]

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อหาค่าความสอดคล้องกันของนักวิจัยว่ามีความสนใจร่วมกันในงานวิจัยด้านใดและมากน้อยเพียงใด

1.2.2 เพื่อพัฒนาเครื่องมือที่สนับสนุนการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย

1.3 ขอบเขตของการวิจัย

1.3.1 สกัดข้อมูลที่เกี่ยวข้องกับนักวิจัย เช่น ชื่อผู้วิจัย คำสำคัญ หมวดหมู่สาขาวิชา (Subject Area) และปีที่บทความได้รับการตีพิมพ์จาก ISI Web of Knowledge แล้วนำมาสร้างออนโทโลจีความสนใจด้านงานวิจัยของนักวิจัย

1.3.2 เปรียบเทียบออนโทโลจีของนักวิจัยเพื่อหาค่าความสอดคล้องกันของความสนใจร่วมกันในด้านงานวิจัย

1.3.3 พัฒนาเครื่องมือที่สามารถสร้างและแสดงออนโทโลจีของนักวิจัย และหาค่าความสอดคล้องกันของนักวิจัย เครื่องมือสามารถรองรับการปรับพารามิเตอร์ค่าขีดแบ่งเริ่มต้นความคล้ายตามความเหมาะสมของคำสำคัญ รวมถึงเงื่อนไขที่เกี่ยวข้องกับคำสำคัญที่เลือกมาใช้สร้างออนโทโลจี เช่น จำนวนงานวิจัย ช่วงเวลาที่ตีพิมพ์ หรือจำนวนคำสำคัญที่พิจารณาสำหรับนักวิจัยแต่ละคน เป็นต้น

1.3.4 ทดลองจับคู่นักวิจัยที่สนใจในหมวดหมู่สาขาวิชาที่เหมือนกัน ใกล้เคียงกัน และแตกต่างกัน แล้วทำการเปรียบเทียบค่า Score Value เมื่อคำสำคัญที่ใช้สร้างออนโทโลจีมีจำนวนต่าง ๆ

1.3.5 วัดความสามารถของเครื่องมือในด้านความแม่นยำ และสมรรถนะในการประมวลผล เมื่อคำสำคัญมีจำนวนต่าง ๆ

1.4 ขั้นตอนการวิจัย

1.4.1 ศึกษาโครงสร้างของออนโทโลจี

1.4.2 ศึกษาอัลกอริทึม หรือวิธีการสำหรับสร้างออนโทโลจีนักวิจัย

1.4.3 สกัดข้อมูลนักวิจัยจากฐานข้อมูลงานวิจัย ISI Web of Knowledge

1.4.4 สร้างออนโทโลจีและจับคู่ออนโทโลจีนักวิจัย พร้อมทั้งแสดงค่าความสอดคล้องกันในด้านงานวิจัยสำหรับนักวิจัย

1.4.5 พัฒนาเครื่องมือเพื่อวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย

1.4.6 ทดสอบและปรับปรุงเครื่องมือ

1.4.7 วิเคราะห์และสรุปผล พร้อมข้อเสนอแนะ

1.4.8 จัดทำวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

ได้ระเบียบวิธีและเครื่องมือซึ่งสามารถวิเคราะห์ หรือจับคู่นักวิจัยที่สนใจในงานวิจัยที่เป็นเรื่องเดียวกัน หรือแนวทางเดียวกัน ซึ่งจะช่วยให้นักวิจัยสามารถทำงานวิจัยร่วมกันได้

1.6 ผลงานตีพิมพ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้นำเสนอในการประชุมวิชาการ ดังนี้

1. The 8th International Conference on Computing and Information Technology (IC^{2}IT 2012), 9 – 10 May 2012, Dusit Thani Hotel, Pattaya City, Thailand ในบทความเรื่อง An Ontological Analysis of Common Research Interest for Researchers โดยผู้แต่งคือ Nawarat Kamsiang และ Twittie Senivongse

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

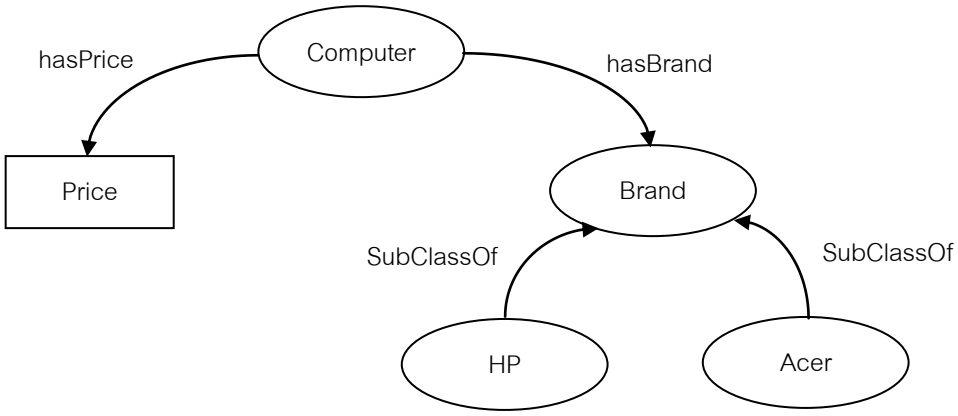
2.1 แนวคิดและทฤษฎี

2.1.1 ออนโทโลจี (Ontology)

ออนโทโลจี [8-10] ถูกนำมาใช้เพื่อบรรยายแนวคิดของโดเมนหรือขอบเขตความสนใจของสิ่งต่าง ๆ ที่อยู่ภายในโดเมนและความสัมพันธ์ระหว่างสิ่งเหล่านั้น โดยแสดงออกมาในรูปแบบของคำสามัญ (Common Word) เพื่อให้มนุษย์และคอมพิวเตอร์มีความเข้าใจในโดเมนหรือขอบเขตความสนใจของสิ่งต่าง ๆ นอกจากนี้ออนโทโลจียังมีความสามารถในการใช้ข้อมูลร่วมกัน (Share) สามารถนำข้อมูลกลับมาใช้ได้ (Reuse) รวมทั้งมีความสามารถในการถ่ายทอดคุณสมบัติ (Inheritance) การนำออนโทโลจีมาใช้งานจึงเป็นทางเลือกหนึ่งในการใช้ข้อมูลร่วมกันและแยกองค์ความรู้ออกจากฐานข้อมูล ซึ่งในสาขาเทคโนโลยีสารสนเทศจะบรรยายแนวคิดในรูปแบบของสัญลักษณ์ (Notation) ยกตัวอย่างเช่น คลาส (Class) ความสัมพันธ์ระหว่างคลาส (Relation) คุณสมบัติของคลาส (Property) กฎ (Role) และอินสแตนซ์ (Instance) ปัจจุบันออนโทโลจีได้ถูกนำมาประยุกต์ใช้งานอย่างแพร่หลาย เช่น เว็บเชิงความหมาย (Semantic Web) การจัดการองค์ความรู้ (Knowledge Management) ธุรกิจอิเล็กทรอนิกส์ (e-Business) พาณิชย์อิเล็กทรอนิกส์ (e-Commerce) และการค้นคืนสารสนเทศ (Information Retrieval) โดยเฉพาะทางด้านปัญญาประดิษฐ์ (Artificial Intelligence)

ส่วนประกอบหลักของออนโทโลจีประกอบด้วยการกำหนดข้อมูลเค้าร่างสำหรับการอธิบายข้อมูลเชิงความหมาย (Semantic Data) และข้อมูลอินสแตนซ์ (Instance Data) ดังนี้

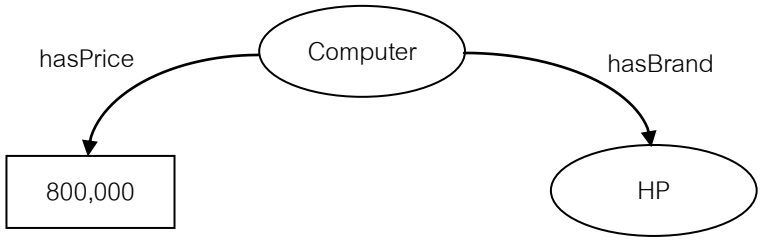
1. กำหนดข้อมูลเค้าร่างสำหรับการอธิบายข้อมูลเชิงความหมาย (Semantic Data) คือ การกำหนดออนโทโลจีระดับบนเพื่อเป็นโมเดลแสดงโครงสร้างการอธิบายข้อมูลเชิงความหมาย โดยกำหนดคลาส (Class) คุณสมบัติของคลาส (Property) และเงื่อนไข (Restriction) ดังภาพที่ 2.1



ภาพที่ 2.1 ตัวอย่างข้อมูลเค้าร่างอธิบายรายละเอียดคลาสคอมพิวเตอร์

จากภาพที่ 2.1 เป็นการแสดงตัวอย่างข้อมูลเค้าร่างอธิบายรายละเอียดคลาสคอมพิวเตอร์ โดยมีการประกาศคลาส Computer ซึ่งมีความสัมพันธ์ hasPrice กับค่า Price รวมทั้งมีความสัมพันธ์ hasBrand กับคลาส Brand ที่มีซับคลาส (Subclass) 2 คลาส คือ คลาส HP และ คลาส Acer

2. ข้อมูลอินสแตนซ์ (Instance Data) คือ การอธิบายรายละเอียดของข้อมูลซึ่งใช้ข้อมูลเค้าร่างเป็นแม่แบบในการอธิบาย ดังภาพที่ 2.2



ภาพที่ 2.2 ตัวอย่างข้อมูลอินสแตนซ์อธิบายรายละเอียดคลาสคอมพิวเตอร์

2.1.2 เวิร์ดเน็ต (Wordnet)

เวิร์ดเน็ต [11-12] เป็นฐานข้อมูลคำศัพท์ภาษาอังกฤษที่รวบรวมคำศัพท์ไว้ทั้งหมด 206,978 คำ ถูกพัฒนาโดยมหาวิทยาลัยพรินซ์ตัน (Princeton University) มีวัตถุประสงค์เพื่อ

ผสมผสานระหว่างพจนานุกรมคำศัพท์ (Dictionary) กับพจนานุกรมอภิธานศัพท์ คำศัพท์พ้องความหมาย คำศัพท์ที่มีความหมายตรงข้าม (Thesaurus) เข้าด้วยกัน การเก็บข้อมูลคำศัพท์ในเวิร์ดเน็ตจะแบ่งตามชนิดของคำ ได้แก่ คำนาม (Noun) คำกริยา (Verb) คำวิเศษณ์ (Adverb) และคำคุณศัพท์ (Adjective) และจับกลุ่มคำศัพท์ที่มีลักษณะความหมายคล้ายคลึงเข้าด้วยกัน เรียกว่า Synonym Set หรือ Synset ซึ่ง Synset จะเชื่อมโยงไปยัง Synset อื่น ๆ ด้วยโครงสร้างแบบลำดับชั้น (Hierarchical Structure) รวมทั้งกำหนดหมายเลขเฉพาะสำหรับคำศัพท์แต่ละคำ เรียกว่า synset_id ดังภาพที่ 2.3

	synset_id	w_num	word	ss_type	sense_number	tag_count
☐	103082979	1	computer	n	1	6
☐	103082979	2	computing machine	n	1	0
☐	103082979	3	computing device	n	1	0
☐	103082979	4	data processor	n	1	0
☐	103082979	5	electronic computer	n	1	0
☐	103082979	6	information processing system	n	1	0

	synset_id	w_num	word	ss_type	sense_number	tag_count
☐	109887034	1	calculator	n	1	0
☐	109887034	2	reckoner	n	1	0
☐	109887034	3	figurer	n	1	0
☐	109887034	4	estimator	n	1	0
☐	109887034	5	computer	n	2	0

ภาพที่ 2.3 ชุดคำศัพท์ที่ได้จาก synset_id ของคำว่า computer

นอกจากนี้เวิร์ดเน็ตยังมีความสัมพันธ์แบบ Is-A หรือเรียกว่า Hypernym และ Hyponym ยกตัวอย่างเช่น (100004475,100004258) ซึ่งในเวิร์ดเน็ต synset_id 10004475 คือ organism และ synset_id 100004258 คือ living_thing และจากตัวอย่างดังกล่าวสามารถอธิบายได้ว่า living_thing เป็น hypernym ของ organism และในทางกลับกัน organism เป็น hyponym ของ living_thing

2.1.3 การวัดค่าความคล้ายคลึงกันของสายอักขระ (String Similarity)

การวัดค่าความคล้ายคลึงกันระหว่างสายอักขระทำให้เราทราบว่าสายอักขระ 2 สาย มีความคล้ายคลึงกันมากน้อยเพียงใด วิธีการวัดค่าความคล้ายคลึงมีดังเช่น

1. Levenshtein Distance [13] วิธีนี้วัดความต่างกันของสายอักขระจากจำนวนการเพิ่ม การลบ หรือการแทนที่ ที่น้อยที่สุดที่ต้องกระทำกับตัวอักษรทีละตัว เพื่อเปลี่ยนสาย

อักขระหนึ่งให้เป็นอีกสายอักขระหนึ่ง สามารถแสดงการคำนวณค่า Levenshtein Distance ระหว่าง “survey” และ “surgery” ดังนี้

- กำหนดให้เมตริกซ์มีขนาด nxm
- เปรียบเทียบตัวอักษรระหว่างแถวและคอลัมน์ทีละตัวอักษรจนครบทุกตัวอักษร
- ถ้าตัวอักษรที่เปรียบเทียบระหว่างแถวและคอลัมน์เหมือนกันให้มีค่าเท่ากับ 0 หากตัวอักษรไม่เหมือนกันให้มีค่าเท่ากับ 1
- นำค่า 0 หรือ 1 ที่ได้จากการเปรียบเทียบความเหมือนกันของตัวอักษรมาบวกกับค่าของตัวอักษรที่อยู่ด้านบน ด้านบนซ้าย และด้านซ้ายของตัวอักษรที่กำลังเปรียบเทียบ แล้วเลือกค่าผลบวกที่น้อยที่สุด และนำมาใส่ในช่องของของตัวอักษรดังกล่าว
- ค่า Levenshtein Distance ที่ได้จะอยู่ในช่องที่ (n, m)

		s	u	r	g	e	r	y
	0	1	2	3	4	5	6	7
s	1	0	1	2	3	4	5	6
u	2	1	0	1	2	3	4	5
r	3	2	1	0	1	2	3	4
v	4	3	2	1	1	2	3	4
e	5	4	3	2	2	1	2	3
y	6	5	4	3	3	2	2	2

จากการคำนวณค่า Levenshtein Distance ระหว่าง “survey” และ “surgery” ผลลัพธ์ที่ได้คือ 2 จากนั้นนำค่า Levenshtein Distance ที่ได้มาคำนวณหาค่าความเหมือนกันระหว่างสายอักขระ 2 สาย มีสูตรการคำนวณดังสมการที่ (2.1) [14]

$$\mathrm{sim}_{\mathrm{edit}}(x, y) = \frac{1}{(1 + \mathrm{editDist}(x, y))} \tag{2.1}$$

โดยที่ editDist(x, y) คือ ค่า Levenshtein Distance

2. Jaro Distance [15, 16] วิธีนี้วัดความต่างกันของสายอักขระจากจำนวนตัวอักษรร่วม (Common Character) และจำนวนการเพิ่ม การลบ หรือการย้ายที่ตัวอักษร โดย

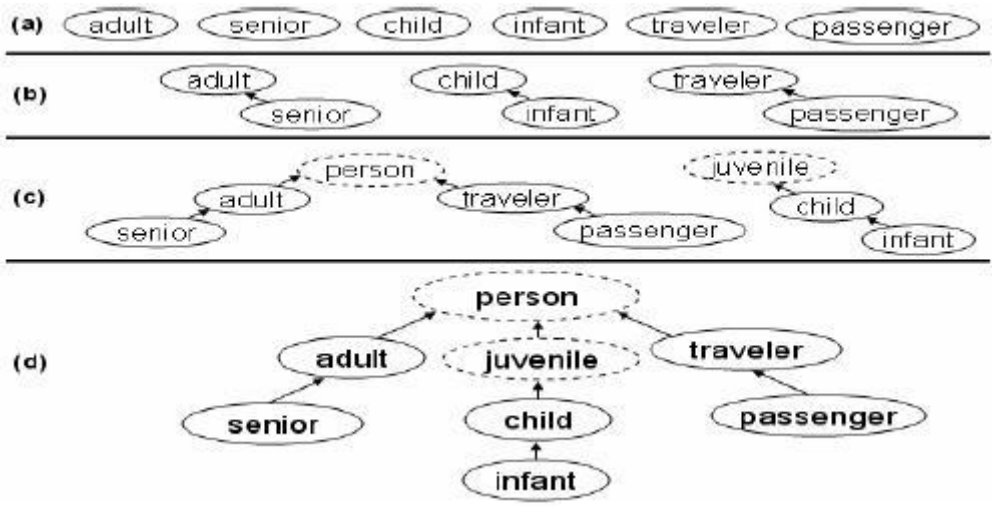
ตัวอักษรร่วมในที่นี้คือตัวอักษรที่เหมือนกันและอยู่ในตำแหน่งใกล้เคียงกัน (ไม่ไกลกว่าครึ่งหนึ่งของความยาวของสายอักขระสายที่สั้น) สามารถแสดงการคำนวณค่า Jaro Distance ดังนี้

- กำหนดสายอักขระ จำนวน 2 สายอักขระ ได้แก่ “tanya” และ “tonya”
- คำนวณค่า Jaro Distance โดยใช้สูตร $s_j = \frac{1}{3}\left(\frac{c}{m}+\frac{c}{n}+\frac{c-t}{n}\right)$
- ค่า c คือ จำนวนของตัวอักษรร่วม หรือจำนวนตัวอักษรที่เหมือนกันระหว่างสายอักขระ 2 สายอักขระ ได้แก่ t,n,y,a มีค่าเท่ากับ 4
- ค่า m คือ จำนวนตัวอักษรของสายอักขระ “tanya” มีค่าเท่ากับ 5
- ค่า n คือ จำนวนตัวอักษรของสายอักขระ “tonya” มีค่าเท่ากับ 5
- ค่า m และ n สามารถเป็นความยาวได้ทั้งสายอักขระที่สั้น หรือสายอักขระที่ยาว
- ค่า t คือ จำนวนตัวอักษรร่วมที่มีการสับเปลี่ยนที่กันของสายอักขระ 2 สายอักขระ มีค่าเท่ากับ 0
- จากการแทนค่าลงในสูตรข้างต้น จะได้
- $s_j = \frac{1}{3}\left(\frac{4}{5}+\frac{4}{5}+\frac{4-0}{4}\right) = \frac{52}{60} = 0.867$

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

2.2.1 การสร้างออนโทโลจี (Ontology Building)

งานวิจัยของ An และคณะ [4] ได้นำเสนอวิธีการสร้างออนโทโลจีแบบอัตโนมัติโดยใช้ข้อมูลจากเว็บเพจที่ไม่สามารถเข้าถึงข้อมูลได้โดยใช้โปรแกรมค้นหา (Search Engine) ซึ่งเว็บเพจดังกล่าวจะถูกเรียกว่า Deep Web หรือ Hidden Web โดยส่วนใหญ่จะสามารถเข้าถึงข้อมูลของ Deep Web ได้โดยกรอกข้อคำถาม (Query) ผ่านฟอร์มเอชทีเอ็มแอลที่เว็บไซต์เหล่านั้นจัดทำขึ้น ซึ่งการสร้างออนโทโลจีจาก Deep Web ผู้วิจัยจะทำการสกัดคำศัพท์ หรือแอททริบิวต์ (Attribute) จากDeep Web และเมื่อได้แอททริบิวต์มาเรียบร้อยแล้ว จะลบแอททริบิวต์ที่เหมือนกันออกไป จากนั้นใช้ ODO Algorithm ที่ผู้วิจัยคิดค้นขึ้นโดยอาศัยพื้นฐานเวิร์ดเน็ต และฟังก์ชัน Hypernym หรือซูเปอร์คลาส (Superclass) ที่มีอยู่ในเวิร์ดเน็ตมาสร้างออนโทโลจี ซึ่งอัลกอริทึมในเฟสที่ 1 จะได้ Schema Fragments (SFs) อัลกอริทึมในเฟสที่ 2 จะเป็นเฟสของการนำ Schema Fragments (SFs) มารวมกันเพื่อให้ได้ออนโทโลจี ซึ่งออนโทโลจีที่สร้างได้นั้นจะอยู่ในรูปของกราฟอวัฏจักรระบุทิศทาง (Directed Acyclic Graphs) ดังภาพที่ 2.4



ภาพที่ 2.4 การสร้างออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุทิศทาง [4]

ภาพที่ 2.4(a) คือ แอททริบิวต์ที่สกัดได้จาก Deep Web ประกอบด้วย 6 แอททริบิวต์ ได้แก่ adult, senior, child, infant, traveler และ passenger ภาพที่ 2.4(b) แสดงการสร้าง Schema Fragments (SFs) ซึ่งจะใช้ฟังก์ชัน Hypernym ที่มีอยู่ในเวิร์ดเน็ตเข้ามาช่วยดำเนินการ และราก (Root) ของ Schema Fragments (SFs) จะเป็นคำศัพท์ที่อยู่ในแอททริบิวต์เท่านั้น ภาพที่ 2.4(c) แสดงวิธีการเพิ่มรากสำหรับ Schema Fragments (SFs) ซึ่งรากที่เพิ่มเข้าไปใหม่จะต้องเป็น Hypernym ของรากในภาพที่ 2.4(b) ภาพที่ 2.4(d) แสดงออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุระบุทิศทาง

ในกรณีที่คำศัพท์ที่มีอยู่ในเวิร์ดเน็ตมี synset_id มากกว่า 1 synset_id ขึ้นไป จะเลือกใช้คำศัพท์ที่สามารถสร้างออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุทิศทางที่มีลำดับชั้น (Hierarchy) น้อยที่สุด และมีค่า Tag Count มากที่สุด ซึ่งค่าดังกล่าวจะแสดงให้เห็นว่าคำศัพท์นั้น ๆ ถูกนำมาใช้มากน้อยเพียงใดและถูกแท็กในข้อความต่าง ๆ ที่สอดคล้องกันเชิงความหมาย จากรูปที่ 2.4(c) คำศัพท์ child มี synset_id จำนวน 4 synset_id ได้แก่ 109917593, 109918248, 109918554 และ 109918762 ซึ่ง synset_id ของคำศัพท์ child ที่เป็น Hypernym ของคำศัพท์ infant ได้แก่ 109918248 แต่ synset_id ดังกล่าวจะสร้างออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุทิศทางที่มีลำดับชั้นมากกว่า synset_id 109917593 ดังนี้

synset_id 109918248 ได้แก่ คำศัพท์ child มีค่า Tag Count เท่ากับ 65 สามารถหา Hypernym ได้ตามลำดับชั้น คือ child -> offspring -> relative -> person

synset_id 109917593 ได้แก่ คำศัพท์ child มีค่า Tag Count เท่ากับ 148 สามารถหา Hypernym ได้ตามลำดับชั้น คือ child -> juvenile -> person

synset_id 109918554 ได้แก่ คำศัพท์ child มีค่า Tag Count เท่ากับ 1 สามารถหา Hypernym ได้ตามลำดับชั้น คือ child -> person

synset_id 109918762 ได้แก่ คำศัพท์ child มีค่า Tag Count เท่ากับ 0 สามารถหา Hypernym ได้ตามลำดับชั้น คือ child -> descendant -> relative -> person

ดังนั้นจึงเลือกใช้คำศัพท์ child ที่มี synset_id 109917593 ในการสร้างออนโทโลจี เนื่องจากมีค่า Tag Count มากกว่า synset_id 109918248 ซึ่งเป็น Hypernym ของคำศัพท์ infant โดยตรง เพื่อให้สามารถสร้างออนโทโลจีในรูปแบบกราฟอวัฏจักรระบุทิศทางที่มีลำดับชั้นน้อยที่สุด

ผู้วิจัยเลือกใช้แนวคิด ODO Algorithm ของงานวิจัยนี้มาสร้างออนโทโลจีความสนใจในด้านงานวิจัยสำหรับนักวิจัยในรูปของกราฟ โดยใช้คำสำคัญซึ่งเป็นข้อมูลการตีพิมพ์ผลงานของนักวิจัยจากฐานข้อมูลไอเอสไอ

2.2.2 การจับคู่ออนโทโลจี (Ontology Matching)

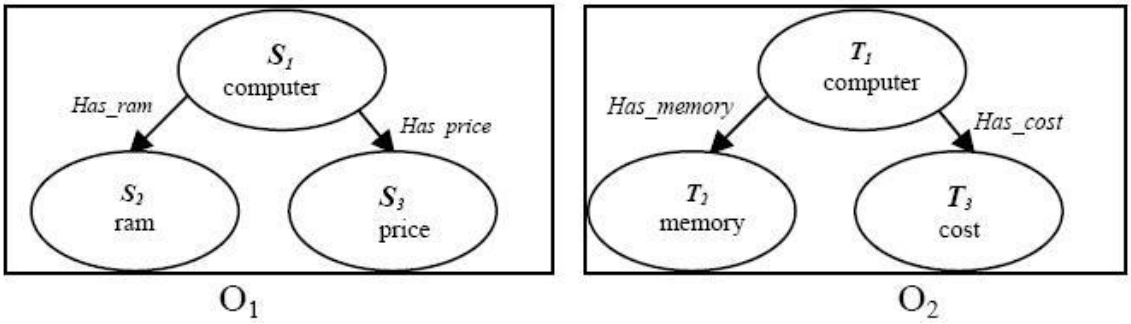
งานวิจัยของ Alasoud และคณะ [5-6] เสนอการจับคู่ออนโทโลจีโดยใช้อัลกอริทึม MLMA+ (Multi-level Matching Algorithm with the Neighbor Search Algorithm) โดยได้กำหนดนิยามและเทอมสำหรับการจับคู่ออนโทโลจี ดังนี้

1. กำหนดความสัมพันธ์ระหว่างเอนทิตี (Entity-relationship) กำหนดให้ S คือ ออนโทโลจีตั้งต้น และ T คือออนโทโลจีเป้าหมาย กำหนดให้ $E^S = \{s_1, s_2, \ldots, s_n\}$ และ $E^T = \{t_1, t_2, \ldots, t_m\}$ คือเซตของเอนทิตีใน S และ T
2. กำหนดเมตริกซ์ความคล้ายคลึง (Similarity Matrix) กำหนดให้ $L(l_{ij})$ ประกอบด้วยค่าที่อยู่ในช่วง [0, 1] ซึ่งเป็นระดับความคล้ายคลึงกันระหว่าง s_i และ t_j
3. เมตริกซ์จับคู่ (Matching Matrix) กำหนดให้ $Map_{0\text{-}1}$ คือ เมตริกซ์ที่มีขนาด nxm และ $r_{ij} \in \{0, 1\}$ ถ้า $r_{ij} = 1$ หมายความว่า s_i และ t_j สามารถจับคู่กันได้ ถ้า $r_{ij} = 0$ หมายความว่า s_i และ t_j ไม่สามารถจับคู่กันได้
4. สเปซจับคู่ (Matching Space) สเปซจับคู่ประกอบด้วยการกำหนดเมตริกซ์จับคู่ที่เป็นไปได้ทั้งหมด ซึ่งจะถูกเรียกว่า mapping space ยกตัวอย่างเช่น ถ้า เมตริกซ์

Map_{0-1} มีขนาด 2 $\times$ 2 จะมีสเปซจับคู่ $2^{2 \times 2}$ สเตท (State) หรือเท่ากับ 16 สเตท ดังนี้

$$\left(\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ldots,\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\right)$$

นอกจากนี้ได้นำเสนอเทคนิคการจับคู่ออนโทโลจี โดยผู้วิจัยได้แสดงตัวอย่างการนำคอมพิวเตอร์ออนโทโลจี 2 ออนโทโลจีมาใช้สำหรับจับคู่ออนโทโลจี และกำหนดออนโทโลจีดังกล่าวให้เป็น O_1 และ O_2 เช่น ดังภาพที่ 2.5



ภาพที่ 2.5 ตัวอย่างคอมพิวเตอร์ออนโทโลจี [6]

จากนั้นทำการเปลี่ยน O_1 และ O_2 ให้อยู่ในรูปของเมตริกซ์ L และเมตริกซ์ Map_{0-1} ดังภาพที่ 2.6

$$L = \begin{bmatrix}1.0 & 0.4 & 0.265\\ 0.463 & 0.534 & 0.083\\ 0.363 & 0.158 & 0.5\end{bmatrix} \qquad \text{Map}_{0-1} = \begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

ภาพที่ 2.6 เมตริกซ์ L และเมตริกซ์ Map_{0-1}ที่ได้จากการเปลี่ยน O_1 และ O_2 [6]

ตัวเลขที่อยู่ภายในเมตริกซ์ L คือ ค่าสัมประสิทธิ์ความคล้ายคลึง (Similarity Coefficient) ระหว่างคำใน O_1 และ O_2 ยกตัวอย่าง เช่น ค่า 1.0 ในแถวที่ 1 หลักที่ 1 เป็นค่าสัมประสิทธิ์ความ

คล้ายคลึงระหว่าง S_1 จาก O_1 และ T_1 จาก O_2 และค่า 0.4 ในแถวที่ 1 หลักที่ 2 เป็นค่าสัมประสิทธิ์ความคล้ายคลึงระหว่าง S_1 จาก O_1 และ T_2 จาก O_2 ค่าสัมประสิทธิ์ความคล้ายคลึงได้มาจากการคำนวณค่าเฉลี่ยระหว่าง Levenshtein Distance และ Linguistic Similarity (Wordnet) [17] ซึ่งเป็นวิธีการคำนวณค่าความคล้ายคลึงกันระหว่างคำโดยใช้เวิร์ดเน็ต โดยฟังก์ชันการทำงานที่พัฒนาโดยใช้โปรแกรมเชิงวัตถุของภาษาเพิร์ล (Perl) และภายในมอดูล (Module) จะมีฟังก์ชันสำหรับการวัดค่าความสัมพันธ์กันเชิงความหมาย ซึ่งประกอบด้วยวิธีการของ Leacock-Chodorow (1998), Jiang-Conrath (1997), Resnik (1995), Lin (1998), Hirst-St-Onge (1998), Wu & Palmer (1994), Banerjee-Pedersen (2002) และ Patwardhan (2003) [18-19]

ผู้วิจัยได้ทำการกำหนดค่าขีดแบ่งเริ่มต้นความคล้าย (Threshold) ให้มีค่ามากกว่าเท่ากับ 0.45 โดยค่าขีดแบ่งเริ่มต้นความคล้ายจะถูกนำมาใช้ในการเปลี่ยนเมตริกซ์ L ให้เป็นเมตริกซ์ $Map_{0\text{-}1}$ หากค่าสัมประสิทธิ์ความคล้ายคลึงในเมตริกซ์ L มีค่ามากกว่าหรือเท่ากับ 0.45 จะถูกเปลี่ยนเป็นเลข 1 ในเมตริกซ์ $Map_{0\text{-}1}$ ในกรณีที่มีค่าน้อยกว่า 0.45 จะถูกเปลี่ยนเป็นเลข 0 ในเมตริกซ์ $Map_{0\text{-}1}$ ยกตัวอย่างเช่น ค่า 1.0 ในแถวที่ 1 หลักที่ 1 ของเมตริกซ์ L มีค่ามากกว่าเท่ากับ 0.45 จึงถูกเปลี่ยนให้เป็นเลข 1 ในเมตริกซ์ $Map_{0\text{-}1}$ และค่า 0.4 ในแถวที่ 1 หลักที่ 2 ของเมตริกซ์ L มีค่าน้อยกว่า 0.45 จึงถูกเปลี่ยนให้เป็นเลข 0 ในเมตริกซ์ $Map_{0\text{-}1}$

จากเมตริกซ์ $Map_{0\text{-}1}$ เมื่อพิจารณาค่าภายในที่เป็น 1 จะทำให้ได้สเตท (State) St_0 โดย $St_0 = \{(s_1 , t_1), (s_2 , t_1), (s_2 , t_2), (s_3 , t_3)\}$ ผู้วิจัยจะใช้เป็นสเตทเริ่มต้นใน Neighbor Search Algorithm ซึ่งมีหลักการทำงาน ดังนี้

1. กำหนดให้ $St_f = St_0$
2. กำหนดให้ St คือ สเปซจับคู่ ซึ่งคือเซตของเมตริกซ์ $Map_{0\text{-}1}$ หรือ St_n ซึ่งเป็นสเตทเพื่อนบ้านของ St_0 ที่เป็นไปได้ทั้งหมด และมีขนาด $n \times m$ โดยที่ $St_n \in St$ และจำนวนสมาชิกที่อยู่ภายในเมตริกซ์ต้องมีจำนวน k-1 หรือ k+1 ของสเตท St_0 เช่น

Map_{0-1} ของ $\mathrm{St}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ดังนั้นตัวอย่างที่เป็นไปได้ของสเตทเพื่อนบ้าน St_n คือ

Map_{0-1} ของ $\mathrm{St}_n = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ หรือ $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ เป็นต้น

3. เลือก St_n 1 ตัว เพื่อเปรียบเทียบค่า Score Value v ระหว่าง St_n และ St_f โดยสามารถคำนวณค่า Score Value ของโครงสร้างเพื่อนบ้านโดยใช้สูตรการคำนวณ [6] ดังสมการที่ (2.2)

$$v = (Map_{0-1} \cdot L)/k = \sum_{i=1}^{n}\sum_{j=1}^{m} Map_{0-1}(i,j) \cdot L(i,j) \Big/ \sum_{i=1}^{n}\sum_{j=1}^{m} Map_{0-1}(i,j) \; ; \; v \geq th \qquad (2.2)$$

เมื่อ v คือ ค่า Score Value ของสเตทเพื่อนบ้าน (Neighbor State)

k คือ จำนวนของ Matched Pairs โดยที่ $k \geq \min(n, m)$

th คือ ค่าขีดแบ่งเริ่มต้นความคล้าย

n คือ จำนวนเอนทิตี (Entity) ของออนโทโลจี S

m คือ จำนวนเอนทิตี (Entity) ของออนโทโลจี T

4. เลือก St_n ตัวถัดไปจาก St เพื่อคำนวณค่า Score Value ในข้อ 3. ทำซ้ำจนกว่าจะคำนวณค่า Score Value ให้กับทุก St_n

5. ค่า St_f สุดท้ายจะแทนสเตทซึ่งแสดงความคล้ายกันของออนโทโลจี 2 ออนโทโลจี

Neighbor Search Algorithm เป็นอัลกอริทึมสำหรับหาสเตทเพื่อนบ้านอื่น ๆ ของสเตท St_0 พร้อมทั้งคำนวณค่า Score Value ของสเตทเพื่อนบ้านที่หาได้ ดังตารางที่ 2.1

ตารางที่ 2.1 ค่า Score Value สำหรับแต่ละสเตทเพื่อนบ้าน [6]

Neighbor number	Matched pairs	Score Value
St_{n1}	$\{(s_2 , t_1), (s_2 , t_2), (s_3, t_3)\}$	0.499
St_{n2}	$\{(s_1 , t_1), (s_1 , t_2), (s_2 , t_1), (s_2 , t_2), (s_3, t_3)\}$	0.5794
St_{n3}	$\{(s_1 , t_1), (s_1 , t_3), (s_2 , t_1), (s_2 , t_2), (s_3, t_3)\}$	0.5524
St_{n4}	$\{(s_1 , t_1), (s_2 , t_2), (s_3, t_3)\}$	**0.678**
St_{n5}	$\{(s_1 , t_1), (s_2 , t_1), (s_3, t_3)\}$	0.6543
St_{n6}	$\{(s_1 , t_1), (s_2 , t_1), (s_2, t_2), (s_2 , t_3), (s_3, t_3)\}$	0.516
St_{n7}	$\{(s_1 , t_1), (s_2 , t_1), (s_2, t_2), (s_3 , t_1), (s_3, t_3)\}$	0.572
St_{n8}	$\{(s_1 , t_1), (s_2 , t_1), (s_2, t_2), (s_3 , t_2), (s_3, t_3)\}$	0.531
St_{n9}	$\{(s_1 , t_1), (s_2 , t_1), (s_2, t_2)\}$	0.6656

ในที่นี้ขอยกตัวอย่างวิธีการคำนวณค่า Score Value ของสเตท St_{n1} จากตารางที่ 2.1 ซึ่งสามารถคำนวณค่า Score Value ดังภาพที่ 2.7

$$St_{n1} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1.0 & 0.4 & 0.265 \\ 0.463 & 0.534 & 0.083 \\ 0.363 & 0.158 & 0.5 \end{bmatrix} \Big/ 3 = 0.499$$

ภาพที่ 2.7 ตัวอย่างวิธีการคำนวณค่า Score Value ของสเตท St_{n1} [6]

จากตารางที่ 2.1 สามารถสรุปได้ว่าสเตท St_{n4} มีค่า Score Value มากที่สุด คือ 0.678 ซึ่งทำให้ได้ค่า St_f สุดท้ายที่เป็นสเตทที่แสดงความคล้ายกันของออนโทโลจี 2 ออนโทโลจี ดังนั้นสามารถวิเคราะห์ได้ว่าออนโทโลจี 2 ออนโทโลจี มีค่าความคล้ายกัน คือ 0.678 โดยที่เอนทิตีที่คล้ายกันจาก 2 ออนโทโลจีนี้คือ computer กับ computer, ram กับ memory และ price กับ cost

ผู้วิจัยเลือกใช้เทคนิคการจับคู่ออนโทโลจีของงานวิจัยนี้ เนื่องจากรูปแบบในการจับคู่ออนโทโลจีจะสอดคล้องกับออนโทโลจีนักวิจัยตามแนวคิดของ ODO Algorithm รวมถึง Neighbor Search Algorithm ซึ่งเป็นอัลกอริทึมสำหรับหาคู่คำที่มีความคล้ายคลึงกัน พร้อมทั้งคำนวณค่า Score Valure ระหว่างออนโทโลจี 2 ออนโทโลจี ทำให้สามารถวิเคราะห์ได้ว่านักวิจัยมีความสนใจร่วมกันในด้านงานวิจัยมากน้อยเพียงใด

2.2.3 การหาความสัมพันธ์แบบ Hyponym ในเวิร์ดเน็ต

งานวิจัยของ Hearst [20] ได้นำเสนอวิธีการหาความสัมพันธ์แบบ hyponym(NP_0, NP_1) ภายในเวิร์ดเน็ต โดยใช้รูปแบบ Lexico-Syntactic โดย NP_0 จะเป็น hyponym ของ NP_1 และในทางกลับกัน NP_1 จะเป็น hypernym ของ NP_0 ดังนี้

1. NP {, NP} * {,} or other NP

   เช่น Bruises, …, broken bones or other injuries …

   จะได้ hyponym(“bruise”, “injury”), hyponym(“broken bone”, “injury”)

2. NP {, NP} * {,} and other NP

   เช่น … temples, treasuries, and other important civic buildings.

   จะได้ hyponym(“temple”, “civic building”), hyponym(“treasury”, “civic building”)

3. NP {,} including {NP ,} * {or | and} NP

   เช่น All common-law countries, including Canada and England …จะได้ hyponym(“Canada”, “common-law country”), hyponym(“England”, “common-law country”)

4. NP {,} especially {NP ,} * {or | and} NP

   เช่น … most European countries, especially France, England, and Spain.

   จะได้ hyponym(“France”, “European country”), hyponym(“England”, “European country”) , hyponym(“Spain”, “European country”)

2.2.4 การกำหนดระยะห่างสำหรับการจับคู่เว็บเซอร์วิซเชิงความหมาย

งานวิจัยของ Yang และคณะ [7] นำเสนอแนวคิดในการกำหนดระยะห่าง (Distance) หรือน้ำหนัก (Weight) เพื่อจับคู่ในเว็บเซอร์วิซเชิงความหมาย (Semantic Web Services) โดยให้

แนวคิดที่ว่าเซอร์วิซที่อยู่ในระดับล่างของออนโทโลจีควรจะมีความสัมพันธ์ที่ใกล้เคียงกันมากกว่าเซอร์วิซที่อยู่ในระดับบนของออนโทโลจี หรือเซอร์วิซที่อยู่ในระดับลึกของต้นไม้ (Tree) ควรจะมีความสัมพันธ์ที่ใกล้เคียงกันมากกว่าเซอร์วิซที่อยู่ในระดับที่สูงกว่าของต้นไม้ ซึ่งสูตรการการคำนวณน้ำหนักสำหรับใช้ในการจับคู่เว็บเซอร์วิซเชิงความหมายเป็นดังสมการที่ (2.3)

$$L_{ws} = L_w (S_1) + L_w (S_2) \quad (2.3)$$

โดยที่ $L_w (S_1)$ คือ ระยะห่างของเซอร์วิซที่สนใจตัวที่ 1 จากรากของออนโทโลจี

$L_w (S_2)$ คือ ระยะห่างของเซอร์วิซที่สนใจตัวที่ 2 จากรากของออนโทโลจี

ผู้วิจัยเลือกใช้แนวคิดของงานวิจัยนี้ ในขั้นตอนการจับคู่ออนโทโลจีในการกำหนดน้ำหนักสำหรับแต่ละระดับความลึกของออนโทโลจี เพื่อแสดงให้เห็นว่าคำที่อยู่ระดับบนเป็นคำทั่วไป (Generic) จากเวิร์ดเน็ต ถึงแม้จะเหมือนกันแต่น่าจะให้ความสำคัญน้อยกว่าคำที่อยู่ระดับล่าง ซึ่งคำที่อยู่ระดับล่างจะเป็นคำที่จำเพาะเจาะจง (Specific) มากกว่า หากยังเหมือนกันอีก แสดงว่าคู่คำนั้นเหมือนกันอย่างมาก

2.2.5 การวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยเรื่องเดียวกัน

งานวิจัยของ Diederic และ Iofciu [21] ได้นำเสนอระบบสำหรับสร้างข้อมูลส่วนตัว (User Profile) ของผู้ใช้งาน และค้นหากลุ่มบุคคลที่มีความสนใจร่วมกันจากข้อมูลส่วนตัวบนพื้นฐานของโฟล์กโซโนมี (Folksonomy) ในการนี้ผู้วิจัยได้ใช้ซับเซต (Subset) ของชุดข้อมูล DBLP ซึ่งเป็นชุดข้อมูลที่รวบรวมรายละเอียดเกี่ยวกับนักวิจัย และการตีพิมพ์บทความของนักวิจัย เป็นต้น มาจัดทำแท็ก (Tag) ข้อความ หรือกำหนดคำสำคัญสำหรับโฟล์กโซโนมี ดังตัวอย่างในตารางที่ 2.2

ตารางที่ 2.2 ตัวอย่างข้อมูลส่วนตัวที่ประกอบด้วยชุดของแท็กหรือคำสำคัญ [21]

Publication Title	Tags (Keyword)
Magpie: supporting browsing and navigation on the semantic web	named entity recognition (NER), semantic web, semantic web services, …
Bootstrapping ontology alignment methods with APFEL	alignment, mapping, ontology, …
Swoogle: a search and metadata engine for the semantic web	rank, search, semantic web, …

ผู้ใช้งานสามารถสร้างข้อมูลส่วนตัวโดยการเลือกแท็กข้อความจากแท็กข้อความที่กำหนด ซึ่งการเลือกแท็กข้อความผู้ใช้งานสามารถค้นหาและเลือกหัวเรื่องงานวิจัยที่ตนเองสนใจ เมื่อผู้ใช้งานเลือกหัวเรื่องงานวิจัยที่ตนเองสนใจเสร็จเรียบร้อยแล้ว ระบบจะทำการสร้างข้อมูล ส่วนตัวของผู้ใช้งาน และนำเข้าระบบสำหรับแนะนำข้อมูลส่วนตัว พร้อมทั้งระบุบุคคลที่มี ความสนใจในงานวิจัยเรื่องเดียวกันกับผู้ใช้งาน

งานวิจัยของ Sriharee และ Punnarut [22] ได้นำเสนอโปรแกรมประยุกต์บนเว็บเชิงความหมาย (Semantic Web Application) เพื่อแสดงให้เห็นเครือข่ายทางสังคมทางด้านวิชาการระหว่างอาจารย์และนักศึกษาภายในมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ซึ่งรายละเอียดข้อมูลได้มาจากการสกัดข้อมูลจากเว็บไซต์ของมหาวิทยาลัย มีขั้นตอนการดำเนิน งาน ดังนี้

1. จัดหมวดหมู่แหล่งข้อมูลเพื่อเก็บรวบรวมข้อมูล (Classifying sources of information for gathering the data) จัดหมวดหมู่เว็บเพจในโดเมนที่แตกต่างกันและเก็บข้อมูลที่จัดหมวดหมู่เรียบร้อยแล้ว
2. กำหนดเมตาดาตาและออนโทโลจีเพื่ออธิบายข้อมูล (Defining matadata and ontology to represent campus-based resources) สร้างออนโทโลจีเพื่อใช้ในการอธิบายข้อมูล โดยพิจารณาข้อมูลที่สามารถใช้ประโยชน์ได้จากเว็บไซต์ของมหาวิทยาลัย
3. สร้างรายละเอียดข้อมูล (Creating campus-based resource description) ทำการสกัดข้อมูลจากเว็บเพจที่ได้จัดหมวดหมู่สำหรับโดเมนที่แตกต่างกัน ซึ่งการสกัดข้อมูล

จะถูกเชื่อมความสัมพันธ์ระหว่างออนโทโลจีที่มีแนวความคิดเหมือนกัน ซึ่งผู้เชี่ยวชาญด้านโดเมนจะกำหนดความสัมพันธ์ระหว่างออนโทโลจีให้มีความเหมาะสม

4. การวิเคราะห์ความสัมพันธ์ในเครือข่ายสังคม (Analyzing relations defined in the social network of Semantic Campus) ใช้อัลกอริทึมสำหรับวิเคราะห์ความสัมพันธ์ในเครือข่ายสังคม
5. การสอบถามข้อมูลของมหาวิทยาลัยเชิงความหมาย (Querying information of Semantic Campus) พัฒนาเพื่อให้ผู้ใช้งานสามารถกำหนดคอนเซ็ปต์สำหรับการซักถามรายละเอียดข้อมูลที่ถูกกำหนดในมหาวิทยาลัยเชิงความหมาย (Semantic Campus) ซึ่งจะสามารถเชื่อมโยงรายละเอียดข้อมูลจากมหาวิทยาลัยเชิงความหมายไปยังข้อมูลที่เป็นปัจจุบันที่มีอยู่ในเว็บไซต์

ในแต่ละขั้นตอนการดำเนินงานดังกล่าวสามารถย้อนกลับไปดำเนินการใหม่ได้ทั้งกระบวนการ หรือย้อนกลับไปดำเนินการที่กระบวนการใดกระบวนการหนึ่งเมื่อข้อมูลมีการเปลี่ยนแปลง หากการสกัดรายละเอียดข้อมูลไม่เพียงพอสำหรับการค้นหาความเชื่อมโยงเครือข่ายทางสังคม อาจจะทำให้ประสิทธิภาพในการวิเคราะห์ความเกี่ยวข้องกันลดลง

ในปี 2010 Punnarut และ Sriharee [23] ได้นำเสนอการค้นหาความเชี่ยวชาญของนักวิจัยโดยใช้เหมืองข้อมูล (Data Mining) ร่วมกับความสามารถในการจัดหมวดหมู่ของออนโทโลจี มีขั้นตอนดำเนินงานดังนี้

1. เก็บรวบรวมเซตข้อมูล (Collecting the data set) เซตข้อมูลที่เก็บรวบรวมประกอบด้วย รายละเอียดของงานวิจัยจากการประชุมวิชาการ JCSSE และ NCSEC โดยเฉพาะปีที่ตีพิมพ์งานวิจัย รวมทั้งรายละเอียดของนักวิจัยที่มีการบันทึกข้อมูลไว้ที่สำนักงานคณะกรรมการวิจัยแห่งชาติ
2. สร้างออนโทโลจีหมวดหมู่ทักษะความชำนาญ (Creating skill classification ontology) อันประกอบด้วย CSS Ontology ซึ่งแบ่งหมวดหมู่งานวิจัยตามการแบ่งหมวดหมู่งานวิจัยของ ACM ในปี 1998 และ Support Ontology ซึ่งเป็นโมเดล (Model) ที่เป็นส่วนเพิ่มเติมของ CSS Ontology โดยการสกัดเท็กซ์ (Text) จากชื่อเรื่องของงานวิจัย ซึ่งขั้นตอนนี้จะถูก ใช้ในการค้นคืนความเชี่ยวชาญของนักวิจัย

3. วิเคราะห์ความเชี่ยวชาญของนักวิจัย (Expertise analysis of the researcher) ใช้การจัดหมวดหมู่ทักษะความชำนาญตามออนโทโลจีเพื่อวิเคราะห์หาความสัมพันธ์ของนักวิจัย
4. สร้างข้อมูลส่วนตัวของนักวิจัย (Creating researcher and research profile) จัดเตรียมการสร้างข้อมูลส่วนตัวของนักวิจัยโดยใช้อาวล์ (Web Ontology Language : OWL)
5. พัฒนาระบบซึ่งประกอบด้วยกระบวนการจับคู่ความคล้ายคลึงกันและจัดลำดับสำหรับการค้นคืนความเชี่ยวชาญของนักวิจัย

บทที่ 3

ระเบียบวิธีวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย

ผู้วิจัยมีแนวคิดที่จะพัฒนาระเบียบวิธีและเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยที่เป็นเรื่องเดียวกัน หรือแนวทางเดียวกันสำหรับนักวิจัยซึ่งจะช่วยให้นักวิจัยสามารถใช้เป็นจุดเริ่มต้นในการทำความรู้จักกันและสร้างความร่วมมือระหว่างกัน ซึ่งสามารถอธิบายแนวคิดและวิธีดำเนินการวิจัย ดังนี้

3.1 กำหนดแหล่งข้อมูลและขอบเขตข้อมูลของนักวิจัยที่จะนำมาสร้างออนโทโลจี

ผู้วิจัยใช้รายละเอียดข้อมูลนักวิจัยจากฐานข้อมูลไอเอสไอ [24] และจัดเก็บเฉพาะชื่อนักวิจัยที่ให้ความสนใจ คำสำคัญของงานวิจัย หมวดหมู่สาขาวิชา (Subject Area) และปีที่ผลงานวิจัยได้รับการตีพิมพ์ เพื่อนำมาสร้างออนโทโลจีนักวิจัย ดังนี้

1. งานวิจัยนี้ใช้ข้อมูลนักวิจัยจากฐานข้อมูลไอเอสไอ ซึ่งจะมีฟังก์ชันสำหรับส่งออก (Export) ข้อมูลนักวิจัยในรูปแบบเท็กซ์ไฟล์ (Text File) ดังเช่นภาพที่ 3.1 AU คือ ชื่อผู้วิจัย DE คือ คำสำคัญของงานวิจัย PY คือ ปีที่ผลงานวิจัยได้รับการตีพิมพ์ และ SC คือ หมวดหมู่สาขาวิชา

```
PT J
AU Chatpatanasiri, R
   Kijsirikul, B
AF Chatpatanasiri, Ratthachat
   Kijsirikul, Boonserm
TI A unified semi-supervised dimensionality reduction framework for
   manifold learning
SO NEUROCOMPUTING
LA English
DT Article
DE Semi-supervised learning; Transductive learning; Spectral methods;
   Dimensionality reduction; Manifold learning
ID DISCRIMINANT-ANALYSIS; EXTRACTION
AB We present a general framework of semi-supervised dimensionality
   reduction for manifold learning which naturally generalizes existing
   supervised and unsupervised learning frameworks which apply the spectral
   decomposition. Algorithms derived under our framework are able to employ
   both labeled and unlabeled examples and are able to handle complex
   problems where data form separate clusters of manifolds. Our framework
   offers simple views, explains relationships among existing frameworks
   and provides further extensions which can improve existing algorithms.
   Furthermore, a new semi-supervised kernelization framework called "KPCA
   trick" is proposed to handle non-linear problems. Crown Copyright (C)
   2010 Published by Elsevier B.V. All rights reserved.
C1 [Chatpatanasiri, R; Kijsirikul, B] Chulalongkorn Univ, Dept Comp Engn, Bangkok 10330, Thailand.
RP Chatpatanasiri, R (reprint author), Chulalongkorn Univ, Dept Comp Engn, Bangkok 10330, Thailand
EM ratthachat.c@gmail.com
   boonserm.k@chula.ac.th
FU Thailand Research Fund; Chulalongkorn University
FX This work is supported by the Thailand Research Fund and the 90th
   Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot
   Endowment).
NR 32
TC 2
Z9 3
PU ELSEVIER SCIENCE BV
PI AMSTERDAM
PA PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS
SN 0925-2312
J9 NEUROCOMPUTING
JI Neurocomputing
PD JUN
PY 2010
VL 73
IS 10-12
SI SI
BP 1631
EP 1640
DI 10.1016/j.neucom.2009.10.024
PG 10
WC Computer Science, Artificial Intelligence
SC Computer Science
GA 615KL
UT WOS:000279134100013
ER
```

ภาพที่ 3.1 ข้อมูลนักวิจัยในรูปแบบเท็กซ์ไฟล์ที่ได้จากฐานข้อมูลไอเอสไอ

2. จัดเก็บชื่อนักวิจัย คำสำคัญของงานวิจัย หมวดหมู่สาขาวิชา และปีที่ผลงานวิจัยได้รับการตีพิมพ์ ลงในฐานข้อมูลมายซีควล (MySQL) เพื่อนำข้อมูลดังกล่าวไปใช้ในการดำเนินงานในขั้นตอนที่ 3.2

3.2 กระบวนการสร้างออนโทโลจีนักวิจัย

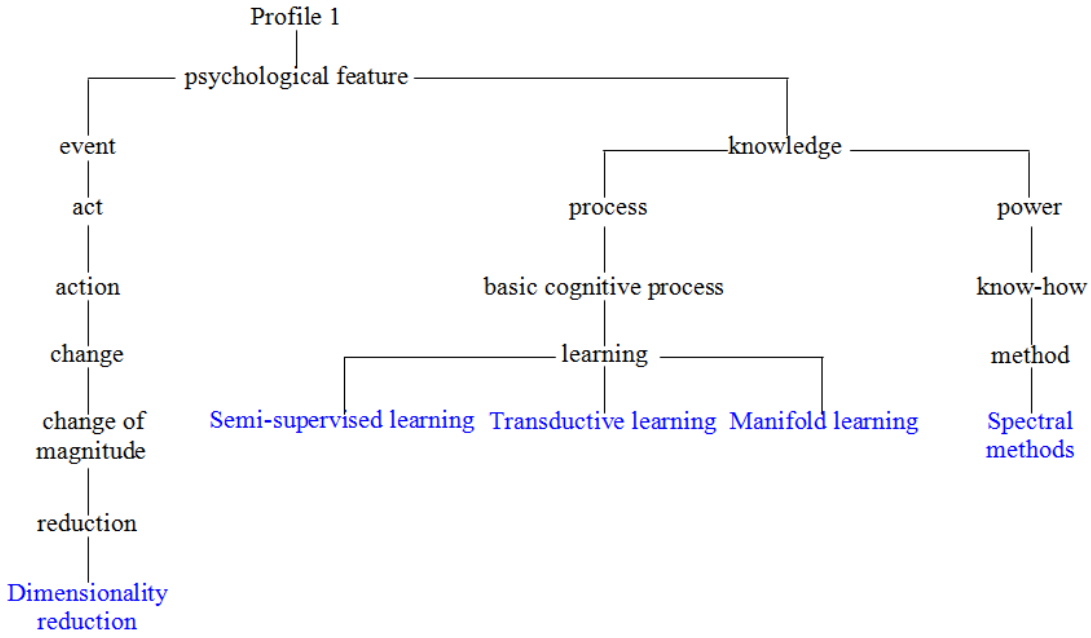
ผู้วิจัยมีแนวคิดที่จะประยุกต์การสร้างออนโทโลจีนักวิจัยจากแนวคิดงานวิจัยของ An และคณะ [4] ดังที่กล่าวในหัวข้อที่ 2.2.1 โดยจะนำคำสำคัญหรือแอททริบิวต์ที่ได้จากขั้นตอนที่ 3.1 มาหาความสัมพันธ์แบบ Hypernym หรือความสัมพันธ์แบบแม่-ลูก (Parent-Child Relation) จากเวิร์ดเน็ต จากนั้นนำแต่ละสายความสัมพันธ์แบบแม่-ลูก ของแต่ละคำสำคัญมาเชื่อมต่อกัน เพื่อสร้างออนโทโลจีนักวิจัยในรูปของกราฟอวัฏจักรระบุทิศทาง

เพื่อยกตัวอย่างในการนี้ได้ใช้คำสำคัญของงานวิจัยในหมวดหมู่สาขาวิชา Computer Science จากฐานข้อมูลไอเอสไอของผู้แต่ง จำนวน 2 คน โดยพิจารณาจากคำสำคัญที่มีความถี่สูงสุด 5 ลำดับแรก คำสำคัญที่ใช้ในการสร้างออนโทโลจีของนักวิจัยแต่ละคน มีรายละเอียดดังนี้

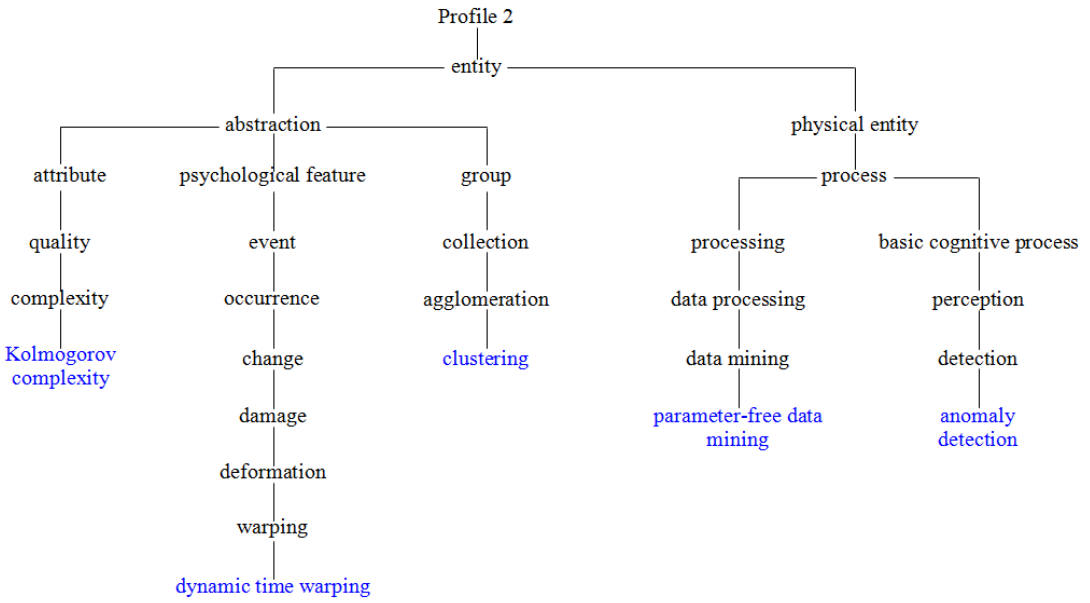
Kijsirikul, B มีคำสำคัญที่ได้จากงานวิจัย คือ Dimensionality reduction, Semi-supervised learning, Transductive learning, Spectral methods และ Manifold learning

Ratanamahatana, CA มีคำสำคัญที่ได้จากงานวิจัย คือ Kolmogorov complexity, dynamic time warping, parameter-free data mining, anomaly detection และ clustering

จากคำสำคัญที่ได้จากงานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA สามารถสร้างออนโทโลจีนักวิจัยได้ ดังภาพที่ 3.2 และ 3.3 ขั้นตอนการทำงานของอัลกอริทึมที่ใช้สร้างออนโทโลจีนักวิจัยดังกล่าวมีกระบวนการทำงาน ดังนี้



ภาพที่ 3.2 ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Kijsirikul, B



ภาพที่ 3.3 ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Ratanamahatana, CA

ข้อมูลนำเข้า : คำสำคัญ

ข้อมูลส่งออก : ออนโทโลจีโปรไฟล์งานวิจัย

เริ่มต้น

1. เลือกคำสำคัญเริ่มต้นตั้งแต่ 2 คำ ขึ้นไป เพื่อสร้างออนโทโลจีนักวิจัย
2. หา hypernym ในเวิร์ดเน็ต : นำคำสำคัญมาหาความสัมพันธ์แบบ hypernym ภายในเวิร์ดเน็ต ในกรณีที่คำสำคัญนั้น ๆ มีหลาย hypernym เลือก hypernym ที่มีค่า tag count มากที่สุด ซึ่งค่า tag count จะแสดงให้เห็นว่า hypernym นั้นถูกนำมาใช้บ่อยที่สุด และถูกแท็กในเท็กซ์ต่าง ๆ ที่สอดคล้องเชิงความหมาย จากภาพที่ 3.3 คำสำคัญเริ่มต้น คือ clustering ซึ่งจะมี agglomeration เป็น hypernym ในกรณีที่คำสำคัญไม่มีในเวิร์ดเน็ตแต่อาจเป็นคำนามพหูพจน์ เช่น คำที่ลงท้ายด้วย “ches”, “shes”, “sses”, “ies”, “ses”, “xes”, “zes” หรือ “s” เปลี่ยนเป็นคำนามเอกพจน์ จากนั้นนำคำดังกล่าวไปหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ตอีกครั้ง หากเป็นไปได้คำสำคัญเริ่มต้นจะสามารถหาความสัมพันธ์แบบ hypernym ได้ ในกรณีที่ไม่สามารถหาความสัมพันธ์ดังกล่าวได้ ให้ดำเนินการตามขั้นตอนที่ 3
3. กำหนด hypernym : ในกรณีที่ไม่สามารถหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ตได้ ดำเนินการตามขั้นตอนดังนี้
   3.1 ในกรณีที่คำสำคัญเป็นคำเดี่ยวหรือคำย่อ ให้กำหนดหมวดหมู่สาขาวิชา (Subject area) ของนักวิจัยเป็น hypernym แต่บางหมวดหมู่สาขาวิชาในฐานข้อมูลไอเอสไอจะมี “&” คั่นกลางระหว่างคำ ดังนั้นจะใช้คำที่อยู่ด้านหน้าและด้านหลัง เป็น hypernym เช่น “Science & Technology” จะกำหนดให้ “Science” และ “Technology” เป็น hypernym ของคำสำคัญนั้น ๆ
   3.2 ในกรณีที่คำสำคัญประกอบด้วยนามวลี (Noun Phrase) จะใช้แนวคิดของ Hearst [20] ซึ่งจะพิจารณารูปแบบของประโยคเพื่อค้นพบความสัมพันธ์แบบ hypernym แต่ในที่นี้จะพิจารณารูปแบบของคำเพื่อหาความสัมพันธ์แบบ hypernym โดยการตัดคำที่อยู่ด้านหน้า และนำนามวลีที่เหลือไปหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ต และทำซ้ำจนกระทั่งสามารถหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ตได้ จากภาพที่ 3.2 คำสำคัญ

"Dimensionality reduction" เมื่อดำเนินการตัดคำที่อยู่ด้านหน้า คือ Dimensionality และนำนามวลีที่เหลือ คือ reduction ไปหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ต ปรากฏว่าค้นพบความสัมพันธ์ดังกล่าวในเวิร์ดเน็ต จึงกำหนดให้ "reduction" เป็น hypernym ของ "Dimensionality reduction" และจากภาพที่ 3.3 คำสำคัญ "parameter-free data mining" เมื่อดำเนินการตัดคำที่อยู่ด้านหน้า คือ parameter-free และนำนามวลีที่เหลือ คือ data mining ไปหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ต ปรากฏว่าค้นพบความสัมพันธ์ดังกล่าวในเวิร์ดเน็ต จึงกำหนดให้ "data mining" เป็น hypernym ของ "parameter-free data mining" ในกรณีที่ตัดคำด้านหน้าออกทีละคำแล้วไม่สามารถหาความสัมพันธ์แบบ hypernym ในเวิร์ดเน็ต ให้กำหนดหมวดหมู่สาขาวิชาของนักวิจัยเป็น hypernym

3.3 คำสำคัญในฐานข้อมูลไอเอสไออาจจะประกอบด้วยนามวลีและคำย่อ เช่น finite element method (FEM) หรือ PTPC (percutaneous transhepatic portal catheterization) เป็นต้น ในกรณีนี้จะตัดคำย่อออกและดำเนินการตามขั้นตอนที่ 3.2 กับนามวลีที่เป็นคำเต็ม ดังนั้น hypernym ของ "finite element method (FEM)" คือ method และ hypernym ของ "PTPC (percutaneous transhepatic portal catheterization)" คือ catheterization

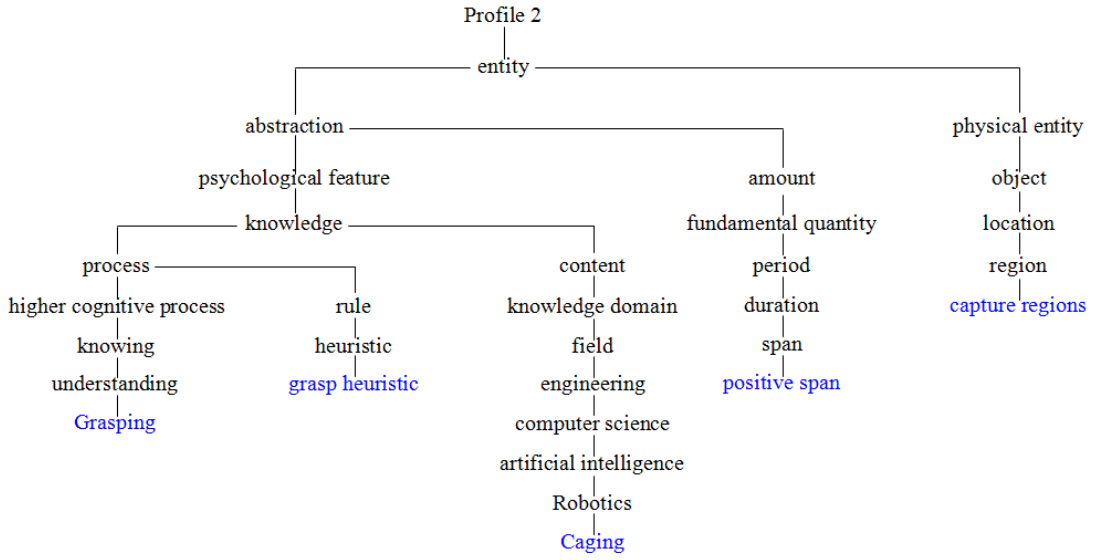
4. สร้างออนโทโลจี : ดำเนินการหาความสัมพันธ์แบบ hypernym จากเวิร์ดเน็ตตามขั้นตอนที่ 2 และ 3 ของแต่ละคำสำคัญ จนกระทั่งไม่สามารถหาความสัมพันธ์ดังกล่าวได้ ในที่นี้จะทำให้ได้สายของออนโทโลจี (Ontology Fragment) ของแต่ละคำสำคัญ

5. รวมออนโทโลจี : ขั้นตอนสุดท้ายคือการเชื่อมต่อสายของออนโทโลจีของแต่ละคำสำคัญเข้าด้วยกัน โดยการพิจารณาเทอมที่ปรากฏใน 2 สายของออนโทโลจี หากพบคู่เทอมที่เหมือนกันให้นำมาเชื่อมต่อกัน และจุดที่โหนดเชื่อมต่อกัน หากมีเส้นทาง (Path) หลายเส้นทางจากโหนดที่เชื่อมต่อไปยังราก (แต่ละเส้นทางประกอบด้วยโหนดที่แตกต่างกัน) ให้เลือกเส้นทางที่สั้นที่สุดสำหรับสร้างออนโทโลจี ซึ่งจะทำให้ออนโทโลจีมีขนาดไม่ใหญ่จนเกินไป จากภาพที่ 3.2 จะมีสายของออนโทโลจี จำนวน 5 สาย ซึ่งสามารถเชื่อมต่อกันที่โหนด learning,

knowledge และ psychological feature ตามลำดับ และเนื่องจากสามารถเชื่อมต่อแต่ละสายของออนโทโลจีทั้ง 5 สาย ที่โหนด psychological feature ดังนั้น hypernym ของ psychological feature (เช่น abstraction -> entity) จะถูกตัดออกไป

จบการทำงาน

เพื่อนำตัวอย่างไปอภิปรายผลการจับคู่ความคล้ายคลึงระหว่างออนโทโลจีนักวิจัยในข้อที่ 3.3 ในที่นี้ขอยกตัวอย่างออนโทโลจีที่ได้จากคำสำคัญของนักวิจัย Sudsang, A เพิ่มเติม ซึ่งอยู่ในหมวดหมู่สาขาวิชา Robotics ดังภาพที่ 3.4 ซึ่งมีคำสำคัญที่ได้จากงานวิจัย คือ Grasping, positive span, grasp heuristic, Caging และ capture regions

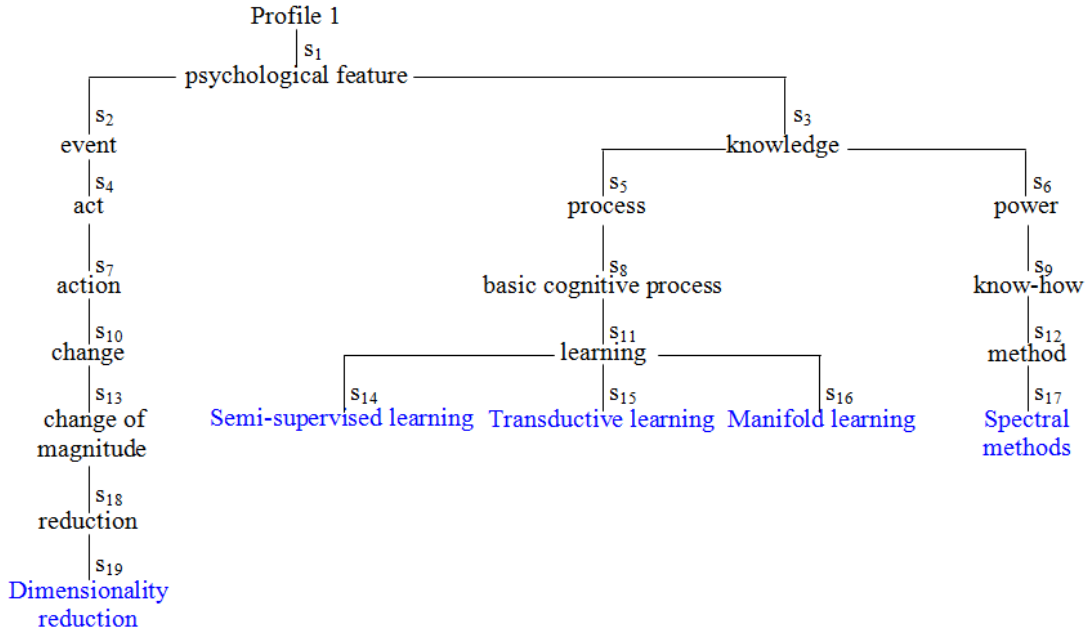


ภาพที่ 3.4 ออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Sudsang, A

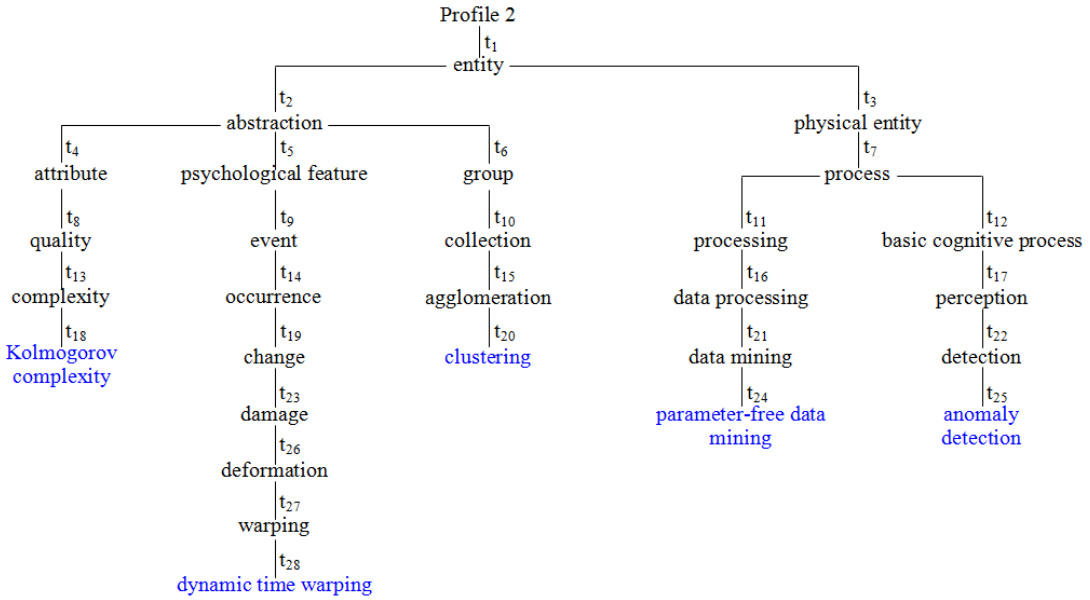
3.3 กระบวนการจับคู่ความคล้ายคลึงระหว่างออนโทโลจีนักวิจัย

แนวคิดในการดำเนินงานการจับคู่ความคล้ายคลึงกันระหว่างออนโทโลจีนักวิจัยมาจากงานวิจัยของ Alasoud และคณะ [5-6] โดยจะนำออนโทโลจีที่อยู่ในรูปโครงสร้างต้นไม้ที่ได้จากขั้นตอนที่ 3.2 มาดำเนินการตามอัลกอริทึม MLMA+ ดังนี้

1. นำออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Kijsirikul, B และ Ratanamahatana, CA ที่ได้จากขั้นตอนที่ 3.2 มากำหนดให้เป็นออนโทโลจี S และออนโทโลจี T ดังภาพที่ 3.5 และ 3.6



ภาพที่ 3.5 ออนโทโลจี S



ภาพที่ 3.6 ออนโทโลจี T

2. จากภาพที่ 3.5 และ 3.6 เปลี่ยนออนโทโลจี S และ T ที่อยู่ในรูปโครงสร้างกราฟให้อยู่ในรูปของเมตริกซ์ L ซึ่งขนาดของเมตริกซ์ L จะเท่ากับ จำนวนเทอมในออนโทโลจี S คูณกับจำนวนเทอมในออนโทโลจี T เช่น จำนวนเทอมในออนโทโลจี S เท่ากับ 19 จำนวนเทอมในออนโทโลจี T เท่ากับ 28 จะทำให้เมตริกซ์ L มีขนาดเท่ากับ 19 × 28 ดังภาพที่ 3.7 ค่าที่อยู่ภายในเมตริกซ์ L คือ ค่าสัมประสิทธิ์ความคล้ายคลึง (Similarity Coefficient) ซึ่งเป็นค่าเฉลี่ยระหว่างค่า Name Similarity ซึ่งได้มาจาก การคำนวณค่า Levenshtein Distance และค่า Linguistic Similarity (Wordnet) [17] เช่น s_{10} = change และ t_{23} = damage มีค่า Levenshtein Distance เท่ากับ 0.2 และ ค่า Linguistic Similarity เท่ากับ 0.933 ซึ่งจะทำให้ค่าสัมประสิทธิ์ความคล้ายคลึง (s_{10}, t_{23}) เท่ากับ 0.567
3. เปลี่ยนเมตริกซ์ L ให้อยู่ในรูป เมตริกซ์ $Map_{0\text{-}1}$ ค่าที่อยู่ในเมตริกซ์ $Map_{0\text{-}1}$ คือ ค่าที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์ความคล้ายคลึงในเมตริกซ์ L กับค่าขีดแบ่งเริ่มต้นความคล้าย ซึ่งกำหนดให้เป็น 0.5[1] หากค่าสัมประสิทธิ์ความคล้ายคลึงมีค่ามากกว่า

[1] ผู้วิจัยเลือกใช้ค่าขีดแบ่งเริ่มต้นความคล้ายที่มีค่ามากที่สุด ที่ไม่ทำให้การจับคู่ออนโทโลจีในที่สุดได้ผลลัพธ์เป็นคู่คำที่แต่ละคู่เป็นคำเดียวกัน เนื่องจากหากผลลัพธ์ที่ได้เป็นคู่คำที่เป็นคำเดียวกัน จะแสดงว่าการนำเวิร์ดเน็ตมาช่วยวิเคราะห์ความเหมือนในเชิงภาษาระหว่างคู่คำไม่ได้ทำให้เกิดประโยชน์แต่อย่างใดต่อการจับคู่ออนโทโลจี

หรือเท่ากับ 0.5 จะถูกเปลี่ยนเป็นเลข 1 ในเมตริกซ์ $Map_{0\text{-}1}$ ในกรณีที่มีค่าน้อยกว่า 0.5 จะถูกเปลี่ยนเป็นเลข 0 ในเมตริกซ์ $Map_{0\text{-}1}$ เช่นดังภาพที่ 3.8

L =

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}
s_1	0.359	0.454	0.324	0.402	1	0.4	0.428	0.36	0.47	0.365	0.25	0.39	0.328	0.428	0.328	0.227	0.428	0.027	0.392	0.299	0.207	0.359	0.36	0.021	0.027	0.36	0.31	0.025
s_2	0.386	0.421	0.372	0.389	0.47	0.417	0.472	0.384	1	0.413	0.419	0.356	0.341	0.51	0.375	0.369	0.426	0.027	0.545	0.306	0.353	0.456	0.456	0.021	0.034	0.435	0.405	0.027
s_3	0.341	0.417	0.286	0.413	0.473	0.389	0.517	0.389	0.456	0.345	0.25	0.441	0.328	0.419	0.311	0.216	0.5	0.027	0.396	0.3	0.208	0.407	0.37	0.02	0.039	0.353	0.322	0.027
s_4	0.334	0.389	0.258	0.363	0.425	0.384	0.5	0.356	0.555	0.456	0.232	0.332	0.3	0.467	0.417	0.202	0.456	0.025	0.512	0.281	0.2	0.491	0.459	0.02	0.034	0.421	0.384	0.028
s_5	0.396	0.384	0.467	0.389	0.428	0.391	1	0.375	0.472	0.462	0.57	0.491	0.3	0.407	0.431	0.456	0.491	0.028	0.438	0.281	0.405	0.456	0.452	0.021	0.032	0.379	0.341	0.027
s_6	0.357	0.379	0.411	0.45	0.428	0.384	0.517	0.517	0.521	0.328	0.37	0.409	0.472	0.407	0.296	0.324	0.472	0.025	0.456	0.291	0.308	0.389	0.357	0.021	0.032	0.331	0.313	0.027
s_7	0.386	0.452	0.414	0.472	0.389	0.372	0.563	0.435	0.488	0.473	0.456	0.312	0.384	0.44	0.424	0.402	0.473	0.025	0.538	0.287	0.389	0.5	0.484	0.022	0.042	0.475	0.378	0.032
s_8	0.246	0.343	0.227	0.337	0.39	0.297	0.491	0.317	0.356	0.275	0.192	1	0.255	0.334	0.261	0.184	0.492	0.023	0.31	0.24	0.168	0.463	0.291	0.02	0.025	0.292	0.258	0.028
s_9	0.285	0.232	0.358	0.363	0.388	0.335	0.447	0.35	0.389	0.3	0.217	0.38	0.277	0.353	0.277	0.188	0.435	0.025	0.341	0.26	0.185	0.368	0.322	0.019	0.034	0.317	0.291	0.027
s_{10}	0.405	0.446	0.339	0.419	0.392	0.405	0.438	0.488	0.545	0.403	0.389	0.31	0.44	0.524	0.372	0.343	0.403	0.025	1	0.306	0.336	0.425	0.567	0.002	0.032	0.512	0.484	0.028
s_{11}	0.272	0.403	0.221	0.409	0.359	0.313	0.484	0.405	0.379	0.291	0.217	0.489	0.265	0.341	0.268	0.189	0.493	0.027	0.338	0.272	0.205	0.475	0.313	0.024	0.034	0.306	0.347	0.034
s_{12}	0.284	0.344	0.218	0.35	0.359	0.322	0.475	0.341	0.441	0.452	0.204	0.356	0.27	0.379	0.419	0.179	0.452	0.025	0.425	0.25	0.179	0.425	0.44	0.022	0.036	0.372	0.326	0.027
s_{13}	0.212	0.411	0.198	0.263	0.337	0.261	0.38	0.244	0.387	0.363	0.171	0.274	0.232	0.363	0.35	0.163	0.362	0.028	0.506	0.217	0.159	0.381	0.455	0.024	0.032	0.455	0.291	0.028
s_{14}	0.022	0.024	0.025	0.023	0.024	0.022	0.023	0.023	0.024	0.023	0.027	0.023	0.025	0.024	0.024	0.027	0.025	0.024	0.022	0.028	0.025	0.024	0.022	0.027	0.023	0.024	0.025	0.03
s_{15}	0.027	0.03	0.03	0.027	0.027	0.025	0.027	0.027	0.027	0.028	0.032	0.027	0.028	0.027	0.028	0.03	0.027	0.024	0.027	0.034	0.034	0.028	0.025	0.028	0.03	0.027	0.03	0.042
s_{16}	0.032	0.034	0.036	0.032	0.028	0.03	0.032	0.032	0.032	0.039	0.039	0.027	0.036	0.032	0.036	0.039	0.032	0.027	0.032	0.042	0.042	0.034	0.032	0.027	0.039	0.036	0.039	0.036
s_{17}	0.036	0.042	0.039	0.036	0.03	0.034	0.039	0.036	0.034	0.039	0.036	0.025	0.036	0.036	0.036	0.032	0.042	0.027	0.036	0.034	0.039	0.039	0.034	0.022	0.034	0.042	0.032	0.028
s_{18}	0.272	0.453	0.311	0.281	0.333	0.287	0.409	0.277	0.42	0.417	0.363	0.275	0.25	0.384	0.366	0.321	0.417	0.027	0.521	0.243	0.317	0.478	0.477	0.023	0.042	0.493	0.319	0.03
s_{19}	0.025	0.03	0.03	0.025	0.024	0.022	0.022	0.025	0.023	0.03	0.023	0.024	0.024	0.024	0.027	0.024	0.027	0.023	0.022	0.024	0.024	0.03	0.023	0.021	0.039	0.028	0.023	0.024

ภาพที่ 3.7 เมตริกซ์ L ที่ได้จากการเปลี่ยนออนโทโลจี S และ T บนพื้นฐาน MLMA+

$Map_{0\text{-}1}$ =

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}
s_1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
s_4	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_5	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_6	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
s_8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{10}	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
s_{11}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{13}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_{14}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{15}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{17}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{18}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
s_{19}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ภาพที่ 3.8 เมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+

4. จากเมตริกซ์ Map_{0-1} คู่คำที่มีค่าเป็น 1 จะทำให้ได้สเตท (State) St_0 = {(s_1 , t_5), (s_2 , t_9), (s_2 , t_{14}), (s_2 , t_{19}), (s_3 , t_7), (s_3 , t_{17}), (s_4 , t_7), (s_4 , t_9), (s_4 , t_{19}), (s_5 , t_7), (s_5 , t_{11}), (s_6 , t_7), (s_6 , t_8), (s_6 , t_9) , (s_7 , t_7), (s_7 , t_{19}) , (s_7 , t_{22}), (s_8 , t_{12}) , (s_{10} , t_9), (s_{10} , t_{14}), (s_{10} , t_{19}) , (s_{10} , t_{23}), (s_{10} , t_{26}), (s_{13} , t_{19}) , (s_{18} , t_{19})} ซึ่งสเตท St_0 คือ สเตทเริ่มต้นที่โครงสร้างของ ออนโทโลจี S และ T มีความคล้ายคลึงกัน และผู้วิจัยจะใช้เป็นสเตทเริ่มต้นใน Neighbor Search Algorithm ดังที่กล่าวในหัวข้อที่ 2.2.2 เพื่อหาโครงสร้างเพื่อนบ้านที่มีความคล้ายคลึงกัน ซึ่งจำนวนโครงสร้างเพื่อนบ้านจะเท่ากับจำนวนของเทอมในออนโทโลจี S คือ 19 คูณกับจำนวนของเทอมในออนโทโลจี T คือ 28 ดังนั้น จำนวน โครงสร้างเพื่อนบ้านจะเท่ากับ 532 ตัวอย่างสเตทเพื่อนบ้านของเมตริกซ์ Map_{0-1} ที่มีการเปลี่ยนค่า (s_1, t_5) จาก 1 เป็น 0 แสดงดังภาพที่ 3.9 และสามารถแสดงค่าความคล้ายคลึงกัน (Score Value) ที่คำนวณได้จากโครงสร้างเพื่อนบ้านต่าง ๆ ของออนโทโลจี S และ T ดังตารางที่ 3.1

$Map_{0\text{-}1}$ =

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}
s_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
s_4	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_5	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_6	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
s_8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{10}	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
s_{11}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{13}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_{14}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{15}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{17}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{18}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
s_{19}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ภาพที่ 3.9 สเตทเพื่อนบ้านของเมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+

ตารางที่ 3.1 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+

Neighbor Number	Matched Pairs	Score Value
St_{n1}	{(event,event), (event,occurrence), (event,change), (knowledge,process), (knowledge,perception), (power,process), (power,event), (power,quality), (process,process), (process,processing), (act,process), (act,event), (act,change), (action,process), (action,change), (action,detection), (basic cognitive process,basic cognitive process), (change,event), (change,occurrence), (change,change), (change,damage), (change,deformation), (change of magnitude,change), (reduction,change)}	0.606

ตารางที่ 3.1 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ (ต่อ)

Neighbor Number	Matched Pairs	Score Value
⋮		
St_{n6}	{(psychological feature,psychological feature), (event,event),(event,occurrence), (event,change), (knowledge,process), (power,process), (power,event),(power,quality), (process,process), (process,processing), (act,process), (act,event), (act,change), (action,process), (action,change), (action,detection), (basic cognitive process,basic cognitive process), (change,event), (change,occurrence), (change,change), (change,damage), (change,deformation), (change of magnitude,change), (reduction,change)}	0.627
⋮		

ตารางที่ 3.1 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ (ต่อ)

Neighbor Number	Matched Pairs	Score Value
St_{n12}	{(psychological feature,psychological feature), (event,event), (event,occurrence), (event,change), (knowledge,process), (knowledge,perception), (power,process), (power,event), (power,quality), (process,process), (process,processing), (act,event),(act,change), (action,process), (action,change), (action,detection), (basic cognitive process,basic cognitive process), (change,event), (change,occurrence), (change,change), (change,damage), (change,deformation) (change of magnitude,change), (reduction,change)}	0.627
⋮		

ตารางที่ 3.1 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ (ต่อ)

Neighbor Number	Matched Pairs	Score Value
St_{n17}	{(psychological feature,psychological feature), (event,event), (event,occurrence), (event,change), (knowledge,process), (knowledge,perception), (power,process), (power,event), (power,quality), (process,process), (process,processing), (act,process), (act,event), (act,change), (action,process), (action,change), (basic cognitive process,basic cognitive process), (change,event),(change,occurrence), (change,change), (change,damage), (change,deformation), (change of magnitude,change), (reduction,change)}	0.627
⋮		

ตารางที่ 3.1 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ (ต่อ)

Neighbor Number	Matched Pairs	Score Value
St_{n532}	{(psychological feature,psychological feature), (event,event), (event,occurrence), (event,change), (knowledge,process), (knowledge,perception), (power,process), (power,event), (power,quality), (process,process), (process,processing), (act,process), (act,event), (act,change), (action,process), (action,change), (action,detection), (basic cognitive process,basic cognitive process), (change,event),(change,occurrence), (change,change), (change,damage), (change,deformation), (change of magnitude,change), (reduction,change), (Dimensionality reduction,dynamic time warping)}	0.599

จากตารางที่ 3.1 การคำนวณ Score Value สามารถคำนวณได้จากสมการที่ (2.2) ดังที่กล่าวในหัวข้อที่ 2.2.2 และสามารถสรุปได้ว่านักวิจัยมีระดับความสนใจร่วมกันเป็นเท่าไรโดยดูผลลัพธ์จาก Score Value ที่มีค่ามากที่สุด และสรุปผลด้านงานวิจัยที่มีความสนใจร่วมกันจาก Matched Pairs ของ Score Value ที่มีค่ามากที่สุดนั้น ตัวอย่างเช่น St_{n6}, St_{n12} และ St_{n17} มีค่า Score Value มากที่สุด คือ 0.627 ในกรณีที่ Score Value ที่มีค่ามากที่สุด มีมากกว่า 1 ค่า จะนำ Matched Pairs ทั้งหมดของ Score Value นั้น ๆ มารวมกัน แล้วทำการตัด Matched Pairs ที่ซ้ำกันออกไป ทำให้สามารถสรุปได้ว่านักวิจัยมีความสนใจร่วมกันในด้าน (psychological feature,

psychological feature), (event, event), (event, occurrence), (event, change), (knowledge, process), (power, process), (power, event), (power, quality), (process, process), (process, processing), (act, process), (act, event), (act, change), (action, process), (action, change), (action, detection), (basic cognitive process, basic cognitive process), (change, event), (change, occurrence), (change, change), (change, damage), (change, deformation), (change of magnitude, change), (reduction, change), (knowledge, perception)

จากผลลัพธ์ข้างต้นจะเห็นได้ว่า Matched Pairs ที่ได้จากแนวคิดการจับคู่ออนโทโลจีโดยใช้วิธี MLMA+ มีจำนวนมาก และส่วนใหญ่มาจากคู่คำที่อยู่ในระดับบนของกราฟ และเป็นคู่คำที่มาจากเวิร์ดเน็ต ผู้วิจัยจึงมีแนวคิดที่จะเพิ่มการพิจารณาน้ำหนักความลึก (Depth Weights) สำหรับการจับคู่ออนโทโลจี โดยที่น้ำหนักความลึกของคู่คำที่ถูกจับคู่จะถูกกำหนดโดยระยะทางจากรากของออนโทโลจีจนถึงตำแหน่งของคู่คำที่ปรากฏในออนโทโลจี [7] เนื่องจากมีแนวคิดว่าคู่คำที่เหมือนกันและอยู่ลึกลงไปในระดับล่างของออนโทโลจีควรจะได้รับความสนใจมากกว่าคู่คำที่อยู่ด้านบนและใกล้กับรากของออนโทโลจี ซึ่งถึงแม้จะเหมือนกันแต่น่าจะให้ความสำคัญน้อยกว่าคำที่อยู่ระดับล่าง เพราะคำที่อยู่ระดับล่างถือเป็นความสนใจที่เฉพาะเจาะจงมากกว่า (Specific) ในการนี้สามารถอธิบายขั้นตอนการทำงานตามแนวคิดที่นำเสนอ คือ MLMA+ with Depth Weights ซึ่งปรับจาก MLMA+ ดังนี้

1. นำออนโทโลจีที่ได้จากคำสำคัญของงานวิจัย Kijsirikul, B และ Ratanamahatana, CA ที่ได้จากขั้นตอนที่ 3.2 มากำหนดให้เป็นออนโทโลจี S และออนโทโลจี T ดังภาพที่ 3.5 และ 3.6
2. จากภาพที่ 3.5 และ 3.6 เปลี่ยนออนโทโลจี S และ T ที่อยู่ในรูปโครงสร้างกราฟให้อยู่ในรูปของเมตริกซ์ L ซึ่งค่าสัมประสิทธิ์ความคล้ายคลึงภายในเมตริกซ์ L ได้จากการนำค่าสัมประสิทธิ์ความคล้ายคลึงของคู่คำ s_i และ t_j ตามวิธี MLMA+ เดิม (ค่าเฉลี่ยระหว่างค่า Levenshtein Distance ร่วมกับค่า Linguistic Similarity (Wordnet) [17]) มาคูณกับค่าน้ำหนักความลึกของคู่คำ s_i และ t_j สูตรการคำนวณน้ำหนักความลึก w_{ij} ของคู่คำ s_i และ t_j จากออนโทโลจี S และ T เป็นดังสมการที่ (3.1)

$$w_{ij} = \frac{(\text{rdepth}(s_i) + \text{rdepth}(t_j)}{2} ; w_{ij} \text{ is in } (0, 1] \qquad (3.1)$$

โดยที่ rdepth(t) คือ ระดับความลึกของเทอม t ในออนโทโลจีหารด้วยความสูงของออนโทโลจี

การคูณค่าน้ำหนักความลึกเป็นการลดทอนค่าสัมประสิทธิ์ความคล้ายคลึงของคู่คำ โดยคู่คำที่อยู่ระดับบนของออนโทโลจีจะถูกลดทอนความคล้ายคลึงมากกว่าคู่คำที่อยู่ระดับล่าง และหากคู่คำอยู่ที่ระดับล่างสุดทั้งคู่คือค่าน้ำหนักความลึกเป็น 1 คู่คำนั้นจะไม่ถูกลดทอนความคล้ายคลึง

ในที่นี้ขอยกตัวอย่างการคำนวณค่าสัมประสิทธิ์ความคล้ายคลึงระหว่างเทอมในออนโทโลจี 2 ออนโทโลจี ที่อยู่ภายในเมตริกซ์ L ตามวิธี MLMA+ with Depth Weights สำหรับ s_2 = event และ t_{14} = occurrence ดังภาพที่ 3.5 และ 3.6 จะได้ว่า rdepth(s_2) เท่ากับ 2/8 และ rdepth(t_{14}) เท่ากับ 5/10 จากนั้นนำค่าที่ได้มาหาค่าน้ำหนักความลึกระหว่างคู่คำโดยใช้สมการที่ (3.1) จะได้เป็น (2/8 + 5/10)/2 เท่ากับ 0.375 ค่าน้ำหนักความลึกนี้จะถูกนำมาคูณกับค่าสัมประสิทธิ์ความคล้ายคลึงตามวิธี MLMA+ เดิม ซึ่งมีค่าเท่ากับ 0.51 ดังภาพที่ 3.7 (ค่า Levenshtein Distance ของ s_2 และ t_{14} คือ 0.111 และค่า Linguistic Similarity (Wordnet) ของ s_2 และ t_{14} คือ 0.909 ทำให้ได้ค่าสัมประสิทธิ์ความคล้ายคลึง คือ (0.111 + 0.909)/2 เท่ากับ 0.51) จะทำให้ค่าสัมประสิทธิ์ความคล้ายคลึง (s_2, t_{14}) ในเมตริกซ์ L ตามวิธี MLMA+ with Depth Weights เท่ากับ 0.191 ดังภาพที่ 3.10 แต่สำหรับ s_{10} = change และ t_{23} = damage ดังภาพที่ 3.5 และ 3.6 จะได้ rdepth(s_{10}) เท่ากับ 5/8 และ rdepth(t_{23}) เท่ากับ 7/10 ซึ่งทำให้ค่าน้ำหนักความลึกจะเท่ากับ 0.663 เมื่อนำค่าน้ำหนักความลึกนี้มาคูณกับค่าสัมประสิทธิ์ความคล้ายคลึงตามวิธี MLMA+ เดิม คือ 0.567 ดังภาพที่ 3.7 จะทำให้ค่าสัมประสิทธิ์ความคล้ายคลึง (s_{10}, t_{23}) ในเมตริกซ์ L ตามวิธี MLMA+ with Depth Weights เท่ากับ 0.376 ดังภาพที่ 3.10 จะเห็นว่าการจับคู่ระหว่าง s_2 กับ t_{14} จะถูกลดทอนความคล้ายคลึงมากกว่าการจับคู่ระหว่าง s_{10} กับ t_{23} ซึ่งอยู่ระดับล่างของออนโทโลจีมากกว่า

3. เปลี่ยนเมตริกซ์ L ให้อยู่ในรูป เมตริกซ์ $Map_{0\text{-}1}$ สำหรับค่าที่อยู่ในเมตริกซ์ $Map_{0\text{-}1}$ คือ ค่าที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์ความคล้ายคลึงในเมตริกซ์ L กับค่าขีดแบ่งเริ่มต้นความคล้าย ซึ่งกำหนดให้เป็น 0.35 หากค่าที่อยู่ในเมตริกซ์ L มีค่ามากกว่าหรือเท่ากับ 0.35 จะถูกเปลี่ยนเป็นเลข 1 ในเมตริกซ์ $Map_{0\text{-}1}$ ในกรณีที่มีค่าน้อยกว่า 0.35

จะถูกเปลี่ยนเป็นเลข 0 ในเมตริกซ์ $Map_{0\text{-}1}$ เช่นดังภาพที่ 3.11 จะเห็นว่าสำหรับคู่คำ s_2 = event และ t_{14} = occurrence และคู่คำ s_{10} = change และ t_{14} = occurrence จะถูกพิจารณาว่าคล้ายคลึงกันในภาพที่ 3.8 แต่เนื่องจากเป็นคู่คำทั่วไป และสร้างจากเวิร์ดเน็ตจึงถูกลดทอนความคล้ายคลึงด้วยค่าน้ำหนักความลึกและถูกพิจารณาว่าไม่คล้ายคลึงกันในภาพที่ 3.11 แต่สำหรับคู่คำ s_{10} = change และ t_{23} = damage แม้จะถูกลดทอนความคล้ายคลึงลงบ้างแต่ยังถูกพิจารณาว่าคล้ายคลึงกันทั้งภาพที่ 3.8 และ 3.11

L =

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}
s_1	0.04	0.074	0.053	0.085	0.213	0.085	0.091	0.095	0.123	0.096	0.066	0.102	0.103	0.134	0.103	0.071	0.134	0.01	0.142	0.108	0.009	0.13	0.149	0.009	0.011	0.167	0.159	0.014
s_2	0.068	0.095	0.084	0.107	0.129	0.115	0.13	0.125	0.325	0.134	0.136	0.116	0.128	0.191	0.141	0.138	0.16	0.011	0.232	0.13	0.15	0.194	0.217	0.01	0.016	0.228	0.233	0.017
s_3	0.06	0.094	0.064	0.114	0.13	0.107	0.142	0.126	0.148	0.116	0.081	0.143	0.123	0.157	0.117	0.081	0.188	0.011	0.168	0.128	0.088	0.173	0.176	0.01	0.019	0.185	0.185	0.017
s_4	0.079	0.112	0.074	0.123	0.143	0.13	0.169	0.138	0.215	0.177	0.09	0.129	0.131	0.204	0.182	0.088	0.2	0.012	0.25	0.137	0.098	0.239	0.247	0.011	0.018	0.247	0.245	0.019
s_5	0.094	0.11	0.134	0.131	0.144	0.132	0.338	0.145	0.183	0.179	0.221	0.19	0.131	0.178	0.189	0.2	0.215	0.014	0.214	0.137	0.197	0.222	0.243	0.011	0.017	0.223	0.217	0.019
s_6	0.085	0.109	0.118	0.152	0.144	0.13	0.174	0.2	0.202	0.127	0.143	0.158	0.207	0.178	0.13	0.142	0.207	0.012	0.222	0.142	0.15	0.19	0.192	0.011	0.017	0.194	0.2	0.019
s_7	0.116	0.158	0.145	0.189	0.156	0.149	0.225	0.196	0.22	0.213	0.205	0.14	0.192	0.22	0.212	0.201	0.237	0.014	0.296	0.158	0.214	0.275	0.29	0.013	0.025	0.309	0.265	0.024
s_8	0.074	0.12	0.079	0.135	0.156	0.119	0.196	0.143	0.16	0.124	0.086	0.45	0.128	0.167	0.131	0.092	0.246	0.013	0.171	0.132	0.092	0.255	0.175	0.012	0.015	0.19	0.181	0.021
s_9	0.086	0.125	0.081	0.145	0.155	0.134	0.179	0.158	0.175	0.135	0.098	0.171	0.139	0.177	0.139	0.094	0.218	0.014	0.188	0.143	0.102	0.202	0.193	0.011	0.02	0.206	0.204	0.02
s_{10}	0.147	0.184	0.14	0.194	0.181	0.187	0.203	0.25	0.279	0.207	0.199	0.159	0.248	0.295	0.209	0.193	0.227	0.015	0.613	0.187	0.206	0.26	0.376	0.013	0.021	0.365	0.369	0.023
s_{11}	0.099	0.166	0.091	0.189	0.166	0.145	0.224	0.208	0.194	0.149	0.111	0.251	0.149	0.192	0.151	0.106	0.277	0.017	0.207	0.167	0.126	0.291	0.207	0.016	0.023	0.218	0.265	0.028
s_{12}	0.103	0.142	0.09	0.162	0.166	0.149	0.22	0.175	0.226	0.232	0.105	0.182	0.152	0.213	0.236	0.101	0.254	0.015	0.26	0.153	0.11	0.26	0.292	0.015	0.024	0.265	0.249	0.022
s_{13}	0.09	0.195	0.094	0.138	0.177	0.137	0.2	0.14	0.223	0.209	0.098	0.158	0.145	0.227	0.219	0.102	0.226	0.019	0.342	0.146	0.107	0.257	0.33	0.017	0.023	0.353	0.24	0.025
s_{14}	0.009	0.011	0.012	0.012	0.013	0.012	0.012	0.013	0.014	0.013	0.016	0.013	0.016	0.015	0.015	0.017	0.016	0.016	0.015	0.019	0.017	0.016	0.016	0.02	0.017	0.019	0.021	0.026
s_{15}	0.011	0.014	0.014	0.014	0.014	0.013	0.014	0.016	0.016	0.016	0.018	0.016	0.018	0.017	0.018	0.019	0.017	0.016	0.018	0.023	0.023	0.019	0.018	0.02	0.022	0.021	0.025	0.037
s_{16}	0.014	0.016	0.017	0.017	0.015	0.016	0.017	0.018	0.018	0.022	0.022	0.016	0.023	0.02	0.023	0.024	0.02	0.018	0.022	0.028	0.028	0.023	0.023	0.02	0.028	0.028	0.032	0.032
s_{17}	0.015	0.02	0.019	0.019	0.016	0.018	0.02	0.021	0.02	0.022	0.021	0.014	0.023	0.023	0.023	0.02	0.026	0.018	0.024	0.023	0.026	0.026	0.025	0.016	0.025	0.033	0.026	0.025
s_{18}	0.133	0.243	0.167	0.165	0.196	0.169	0.24	0.177	0.268	0.266	0.231	0.175	0.172	0.264	0.252	0.221	0.287	0.02	0.384	0.179	0.234	0.353	0.376	0.018	0.033	0.413	0.283	0.028
s_{19}	0.014	0.018	0.018	0.016	0.016	0.014	0.014	0.018	0.016	0.021	0.016	0.017	0.018	0.018	0.02	0.018	0.02	0.018	0.018	0.019	0.019	0.024	0.02	0.018	0.033	0.025	0.022	0.024

ภาพที่ 3.10 เมตริกซ์ L ที่ได้จากการเปลี่ยนออนโทโลจี S และ T บนพื้นฐาน MLMA+ with Depth Weights

$Map_{0\text{-}1}$ =

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}
s_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{10}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0
s_{11}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{13}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
s_{14}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{15}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{17}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_{18}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0
s_{19}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ภาพที่ 3.11 เมตริกซ์ $Map_{0\text{-}1}$ บนพื้นฐาน MLMA+ with Depth Weights

4. จากเมตริกซ์ Map_{0-1} คู่คำที่มีค่าเป็น 1 จะทำให้ได้สเตท (State) St_0 = {(s_8 , t_{12}), (s_{10} , t_{19}), (s_{10} , t_{23}), (s_{10} , t_{26}), (s_{10} , t_{27}), (s_{13} , t_{26}), (s_{18} , t_{19}), (s_{18} , t_{22}), (s_{18} , t_{23}), (s_{18} , t_{26})} ซึ่งสเตท St_0 คือ สเตทเริ่มต้นที่โครงสร้างของออนโทโลจี S และ T มีความคล้ายคลึงกัน และผู้วิจัยจะใช้เป็นสเตทเริ่มต้นใน Neighbor Search Algorithm ดังที่กล่าวในหัวข้อที่ 2.2.2 เพื่อ หาโครงสร้างเพื่อนบ้านที่มีความคล้ายคลึงกัน ซึ่งจำนวนโครงสร้างเพื่อนบ้านจะเท่ากับจำนวนของเทอมในออนโทโลจี S คือ 19 คูณกับจำนวนของเทอมในออนโทโลจี T คือ 28 ดังนั้นจำนวนโครงสร้างเพื่อนบ้านจะเท่ากับ 532 ซึ่งค่าความคล้ายคลึงกัน (Score Value) ที่คำนวณได้จากโครงสร้างเพื่อนบ้านต่าง ๆ แสดงดังตารางที่ 3.2

ตารางที่ 3.2 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ with Depth Weights

Neighbor Number	Matched Pairs	Score Value
St_{n1}	{(change,change), (change,damage), (change,deformation), (change,warping), (change of magnitude,deformation), (reduction,change), (reduction,detection), (reduction,damage), (reduction,deformation)}	0.4
⋮		
St_{n6}	{(basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (reduction,change), (reduction,detection), (reduction,damage), (reduction,deformation)}	**0.411**

ตารางที่ 3.2 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านของออนโทโลจี S และ T บนพื้นฐาน MLMA+ with Depth Weights (ต่อ)

Neighbor Number	Matched Pairs	Score Value
St_{n7}	{(basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping) (change of magnitude,deformation), (reduction,detection), (reduction,damage), (reduction,deformation)}	0.408
St_{n8}	{(basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (change of magnitude,deformation), (reduction,change), (reduction,damage), (reduction,deformation)}	**0.411**
⋮		
St_{n532}	{(basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (change of magnitude,deformation), (reduction,change), (reduction,detection), (reduction,damage), (reduction,deformation), (Dimensionality reduction,dynamic time warping)}	0.371

จากตารางที่ 3.2 การคำนวณ Score Value สามารถคำนวณได้จากสมการที่ (2.2) ดังที่กล่าวในหัวข้อที่ 2.2.2 และสามารถสรุปได้ว่านักวิจัยมีระดับความสนใจร่วมกันเป็นเท่าไรโดยดูผลลัพธ์จาก Score Value ที่มีค่ามากที่สุด และสรุปผลด้านงานวิจัยที่มีความสนใจร่วมกันจาก

Matched Pairs ของ Score Value ที่มีค่ามากที่สุดนั้น ตัวอย่างเช่น St_{n6} และ St_{n8} มีค่า Score Value มากที่สุด คือ 0.411 ในกรณีที่ Score Value ที่มีค่ามากที่สุด มีมากกว่า 1 ค่า จะนำ Matched Pairs ทั้งหมดของ Score Value นั้น ๆ มารวมกัน แล้วทำการตัด Matched Pairs ที่ซ้ำกันออกไป ทำให้สามารถสรุปได้ว่านักวิจัยมีความสนใจร่วมกันในด้าน (basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (reduction,change), (reduction,detection) (reduction,damage), (reduction,deformation), (change of magnitude,deformation)

จากข้อมูลดังกล่าวข้างต้น Matched Pairs ที่ได้จากแนวคิดการจับคู่ออนโทโลจีโดยวิธี MLMA+ จะมีจำนวนมาก และเป็นคำทั่วไปที่ได้จากเวิร์ดเน็ต ในขณะที่แนวคิด MLMA+ with Depth Weights ที่นำเสนอจะได้ Matched Pairs ที่น้อยกว่า เนื่องจากถูกลดทอนความคล้ายคลึงกันของคู่คำด้วยค่าน้ำหนักความลึก ดังตารางที่ 3.3

ตารางที่ 3.3 การเปรียบเทียบ Matched Pairs ระหว่างแนวคิด MLMA+ และ MLMA+ with Depth Weights

Matched Pairs	
MLMA+	**MLMA+ with Depth Weights**
(psychological feature, psychological feature), (event, event), (event, occurrence), (event, change), (knowledge, process), (power, process), (power, event), (power, quality), (process, process), (process, processing), (act, process), (act, event), (act, change), (action, process), (action, change), (action, detection), (basic cognitive process, basic cognitive process), (change, event), (change, occurrence), (change, change), (change, damage), (change, deformation), (change of magnitude, change), (reduction, change), (knowledge, perception)	(basic cognitive process, basic cognitive process), (change, change), (change, damage), (change, deformation), (change, warping), (reduction, change), (reduction, detection), (reduction, damage), (reduction, deformation), (change of magnitude, deformation)

นอกจากนี้ผู้วิจัยได้ทดลองสร้างออนโทโลจีนักวิจัยของ Sudsang, A ซึ่งอยู่ในหมวดหมู่สาขาวิชา Robotics ดังภาพที่ 3.4 และทำการจับคู่ความสนใจด้านงานวิจัยกับ Kijsirikul, B ซึ่งอยู่ในหมวดหมู่สาขาวิชา Computer Science ตามแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ และแนวคิด MLMA+ with Depth Weights ดังที่กล่าวในข้อที่ 3.3 และคำนวณค่าความคล้ายคลึงกันจากสมการที่ (2.2) ดังที่กล่าวในหัวข้อที่ 2.2.2 ผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกันเป็นดังตารางที่ 3.4 ซึ่งทั้งสองวิธีต่างพบว่า Kijsirikul, B มีความสนใจร่วมกันกับ Ratanamahatana, CA มากกว่า Sudsang, A

ตารางที่ 3.4 ผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกัน

Algorithm	Author 1	Author 2	Matching Score
MLMA+	Kijsirikul, B	Ratanamahatana, CA	0.627
		Sudsang, A	0.581
MLMA+ with Depth Weights	Kijsirikul, B	Ratanamahatana, CA	0.411
		Sudsang, A	0.372

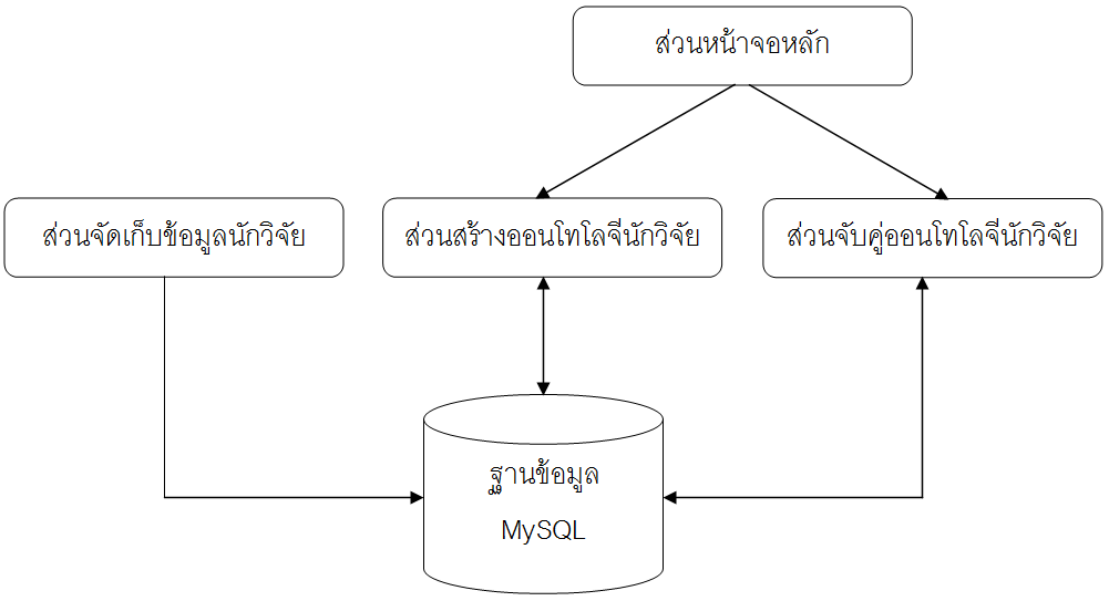
จากตารางที่ 3.4 ผลคะแนนการจับคู่ความสนใจด้านงานวิจัยด้วยวิธี MLMA+ และวิธี MLMA+ with Depth Weights นั้น ไม่สามารถเปรียบเทียบกันได้ เนื่องจากใช้ค่าขีดแบ่งเริ่มต้นความคล้ายคนละค่า ในตัวอย่างนี้แนวคิด MLMA+ ใช้ค่าขีดแบ่งเริ่มต้นความคล้าย เท่ากับ 0.5 แต่แนวคิด MLMA+ with Depth Weights ใช้ค่าขีดแบ่งเริ่มต้นความคล้าย เท่ากับ 0.35 อีกทั้งทั้งสองวิธียังมีการพิจารณาความคล้ายคลึงของคู่คำที่แตกต่างกัน ดังนั้นจึงไม่สามารถเปรียบเทียบค่าคะแนนกันได้ ตัวอย่างเช่น ไม่สามารถเปรียบเทียบค่าความคล้ายคลึงระหว่าง Kijsirikul, B กับ Sudsang, A ด้วยวิธี MLMA+ ซึ่งได้คะแนน 0.581 กับค่าความคล้ายคลึงระหว่าง Kijsirikul, B กับ Ratanamahatana, CA ด้วยวิธี MLMA+ with Depth Weights ซึ่งได้คะแนน 0.411 และสรุปว่า Kijsirikul, B มีความสนใจที่คล้ายคลึงกับ Sudsang, A มากกว่า Ratanamahatana, CA ได้ การเปรียบเทียบคะแนนความคล้ายคลึงจะต้องเปรียบเทียบคะแนนจากวิธีเดียวกัน

บทที่ 4
การพัฒนาเครื่องมือ

ในบทนี้จะกล่าวถึงรายละเอียดการพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย โดยจะกล่าวถึงโครงสร้างของเครื่องมือ และสภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ ซึ่งมีรายละเอียดดังนี้

4.1 โครงสร้างของเครื่องมือ

โครงสร้างของเครื่องมือที่พัฒนาประกอบด้วยการทำงาน 3 ส่วนหลัก คือ ส่วนจัดเก็บข้อมูลนักวิจัย ส่วนสร้างออนโทโลจีนักวิจัย และส่วนจับคู่ออนโทโลจีนักวิจัย ซึ่งสามารถอธิบายด้วยแผนภาพส่วนประกอบของเครื่องมือ ดังภาพที่ 4.1 และมีรายละเอียดดังนี้



ภาพที่ 4.1 แผนภาพแสดงส่วนประกอบโครงสร้างของเครื่องมือ

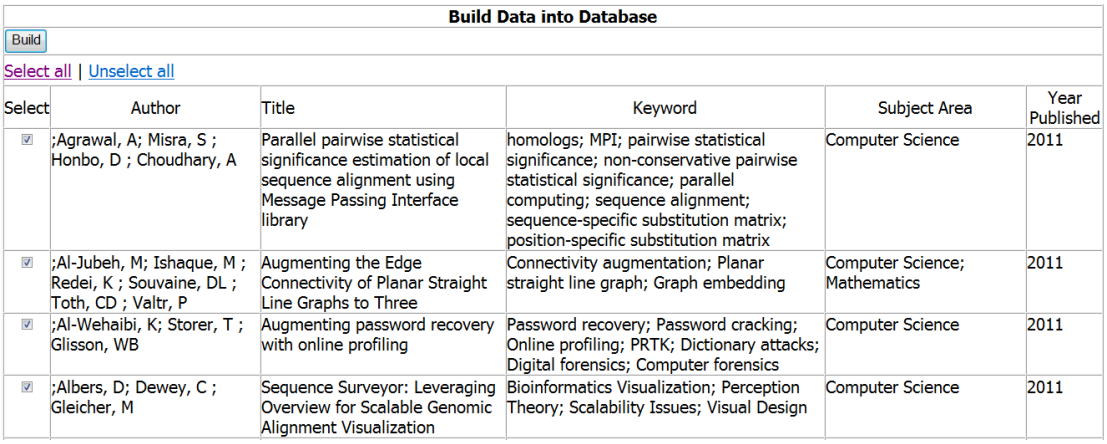
4.1.1 ส่วนจัดเก็บข้อมูลนักวิจัย

ส่วนจัดเก็บข้อมูลนักวิจัยจะบริหารจัดการโดยผู้ดูแลระบบ ผู้วิจัยได้ใช้รายละเอียดข้อมูลจากฐานข้อมูลไอเอสไอตั้งแต่ปีคริสต์ศักราช 2002-2011 ซึ่งจัดเก็บเฉพาะชื่อนักวิจัยที่อยู่ลำดับสูงสุด 100 คน ในแต่ละหมวดหมู่สาขาวิชา คำสำคัญของงานวิจัย ชื่อหมวดหมู่สาขาวิชาที่อยู่ในลำดับสูงสุด 100 สาขาวิชา และปีที่ผลงานวิจัยได้รับการตีพิมพ์ ซึ่งหน้าจอในส่วนจัดเก็บข้อมูลนักวิจัยเป็นดังภาพที่ 4.2 และ 4.3

ภาพที่ 4.2 หน้าจอสำหรับการอัพโหลดเท็กซ์ไฟล์

จากภาพที่ 4.2 เท็กซ์ไฟล์ที่อัพโหลดเข้าสู่ส่วนจัดเก็บข้อมูลได้มาจากการส่งออกข้อมูลนักวิจัยจากฐานข้อมูลไอเอสไอ ซึ่งรายละเอียดภายในเท็กซ์ไฟล์จะปรากฏดังภาพที่ 3.1 และเมื่อดำเนินการอัพโหลดเท็กซ์ไฟล์เรียบร้อยแล้ว จะเข้าสู่ขั้นตอนจัดเก็บข้อมูลนักวิจัย เช่น ชื่อผู้วิจัย คำสำคัญของ งานวิจัย หมวดหมู่สาขาวิชา และปีที่ผลงานวิจัยได้รับการตีพิมพ์ ลงในฐานข้อมูลชื่อ

ผู้วิจัย ฐานข้อมูลคำสำคัญของงานวิจัย ฐานข้อมูลหมวดหมู่สาขาวิชา และฐานข้อมูลปีที่ผลงานวิจัยได้รับการตีพิมพ์ โดยการคลิกที่ปุ่ม Build ดังภาพที่ 4.3



Build Data into Database

Build

Select all | Unselect all

Select	Author	Title	Keyword	Subject Area	Year Published
☑	;Agrawal, A; Misra, S ; Honbo, D ; Choudhary, A	Parallel pairwise statistical significance estimation of local sequence alignment using Message Passing Interface library	homologs; MPI; pairwise statistical significance; non-conservative pairwise statistical significance; parallel computing; sequence alignment; sequence-specific substitution matrix; position-specific substitution matrix	Computer Science	2011
☑	;Al-Jubeh, M; Ishaque, M ; Redei, K ; Souvaine, DL ; Toth, CD ; Valtr, P	Augmenting the Edge Connectivity of Planar Straight Line Graphs to Three	Connectivity augmentation; Planar straight line graph; Graph embedding	Computer Science; Mathematics	2011
☑	;Al-Wehaibi, K; Storer, T ; Glisson, WB	Augmenting password recovery with online profiling	Password recovery; Password cracking; Online profiling; PRTK; Dictionary attacks; Digital forensics; Computer forensics	Computer Science	2011
☑	;Albers, D; Dewey, C ; Gleicher, M	Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization	Bioinformatics Visualization; Perception Theory; Scalability Issues; Visual Design	Computer Science	2011

ภาพที่ 4.3 หน้าจอสำหรับจัดเก็บข้อมูลนักวิจัย

4.1.2 ส่วนสร้างออนโทโลจีนักวิจัย

ส่วนสร้างออนโทโลจีนักวิจัยมีการออกแบบส่วนต่อประสานกับผู้ใช้ (User Interface) ให้สามารถเลือกช่วงปีที่ผลงานวิจัยได้รับการตีพิมพ์ หมวดหมู่สาขาวิชา และชื่อนักวิจัยที่ต้องการสร้างออนโทโลจีโปรไฟล์งานวิจัย ซึ่งจะปรากฏคำสำคัญของงานวิจัยที่ได้จัดเรียงลำดับตามความถี่ของคำจากมากไปหาน้อย ดังภาพที่ 4.4 เมื่อผู้ใช้คลิกที่ปุ่ม Build Research Profiles เครื่องมือจะสร้างออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 4.5 และ 4.6 สำหรับหลักเกณฑ์ในการเลือกคำสำคัญของงานวิจัย ผู้ใช้ควรเลือกคำสำคัญตั้งแต่ 2 คำ ขึ้นไป เพื่อที่เครื่องมือจะได้สามารถสร้างออนโทโลจีโปรไฟล์งานวิจัยในรูปแบบกราฟได้

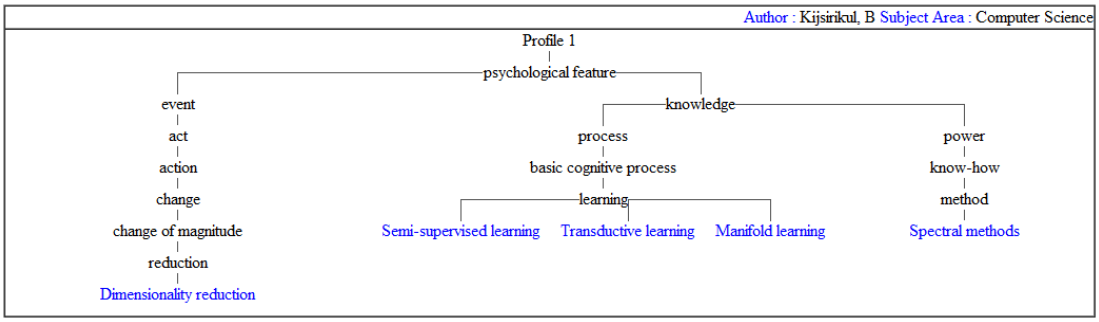
Ontological Research Profile Building and Matching

Research Profile 1	
Year Published :	2002 - 2011
Subject Area :	Computer Science
Author :	Kijsirikul, B
Select all keywords | Unselect all keywords	
Select	Keyword
☑	Dimensionality reduction
☑	Semi-supervised learning
☑	Transductive learning
☑	Spectral methods
☑	Manifold learning
☐	first-order Bayesian networks
☐	inductive logic programming
☐	overfitting problem
☐	feature extraction
☐	propositionalisation
☐	Distance metric learning
☐	Representer theorem
☐	Kernel machines
☐	Kernel alignment
☐	Zone analysis
☐	Bias-variance analysis
☐	Ensemble methods
☐	Bagging
☐	Boosting
☐	Analysis of algorithms
☐	On-line algorithms
☐	On-line learning
☐	Prediction on graphs
☐	Support vector machines
☐	Evolutionary strategies
☐	Kernel methods
☐	Thai tone
☐	tone recognition
☐	combination of neural networks
☐	combination rules
☐	voting techniques

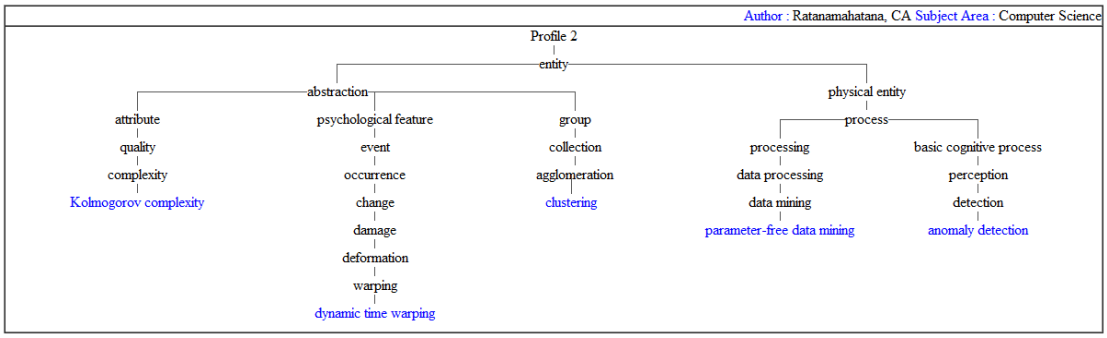
Research Profile 2	
Year Published :	2002 - 2011
Subject Area :	Computer Science
Author :	Ratanamahatana, CA
Select all keywords | Unselect all keywords	
Select	Keyword
☑	Kolmogorov complexity
☑	dynamic time warping
☑	parameter-free data mining
☑	anomaly detection
☑	clustering
☐	visualization
☐	high-dimensional data
☐	dimensionality reduction
☐	clipping
☐	time series data mining
☐	nearest neighbor search
☐	scaled and warped matching
☐	subsequence matching
☐	uniform scaling
☐	Time series
☐	Indexing

Build Research Profiles Match Research Profiles

ภาพที่ 4.4 ส่วนต่อประสานกับผู้ใช้สำหรับสร้างและจับคู่ออนโทโลจี

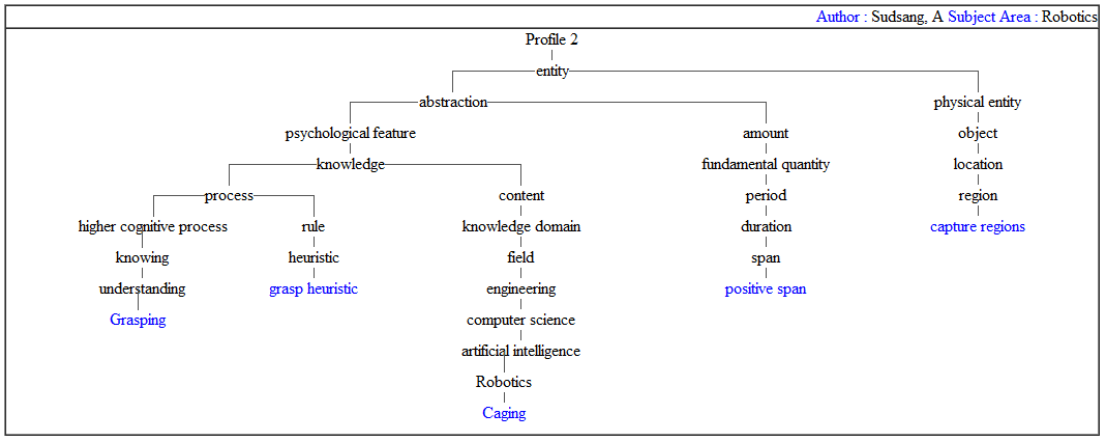


ภาพที่ 4.5 ออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

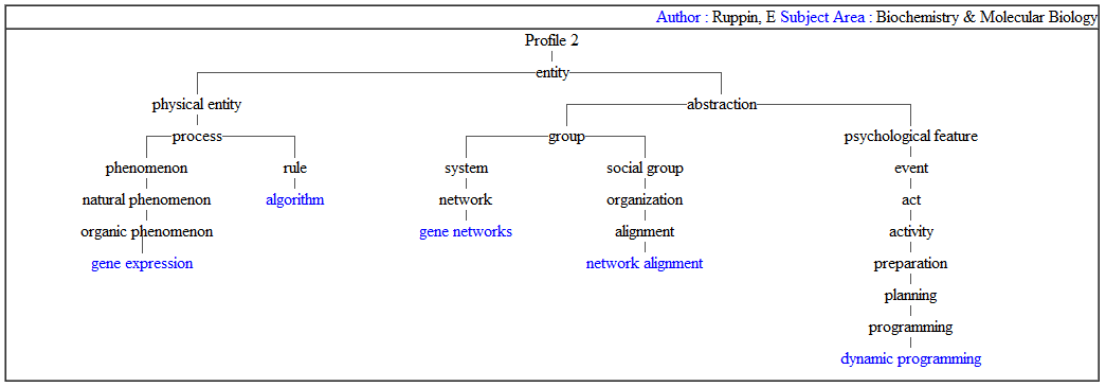


ภาพที่ 4.6 ออนโทโลจีโปรไฟล์งานวิจัยของ Ratanamahatana, CA เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

จากนั้นผู้วิจัยได้ทดลองสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Sudsang, A ในหมวดหมู่สาขาวิชา Robotics และ Ruppin, E ในหมวดหมู่สาขาวิชา Biochemistry & Molecular Biology ดังภาพที่ 4.7 และ 4.8



ภาพที่ 4.7 ออนโทโลจีโปรไฟล์งานวิจัยของ Sudsang, A เมื่อใช้คำสำคัญเริ่มต้น 5 คำ



ภาพที่ 4.8 ออนโทโลจีโปรไฟล์งานวิจัยของ Ruppin, E เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

4.1.3 ส่วนจับคู่ออนโทโลจีนักวิจัย

ส่วนจับคู่ออนโทโลจีนักวิจัย จากภาพที่ 4.4 เมื่อผู้ใช้คลิกที่ปุ่ม Match Research Profiles เครื่องมือจะเริ่มดำเนินการสร้างเมตริกซ์ L เพื่อหาค่าสัมประสิทธิ์ความคล้ายคลึงกันระหว่างเทอมที่ปรากฏอยู่ในออนโทโลจีโปรไฟล์งานวิจัย 2 โปรไฟล์ ดังภาพที่ 4.9

Author : Kijsirikul, B Subject area : Computer Science
Author : Ratanamahatana, CA Subject area : Computer Science

psychological feature,entity	psychological feature,physical entity	psychological feature,abstraction	psychological feature,attribute	psychological feature,group	psychological feature,process
levenshtein=0.05 wordnet=0.667 similarity=0.359 depth weight=0.1125 similarity * depth weight=0.04	levenshtein=0.077 wordnet=0.571 similarity=0.324 depth weight=0.1625 similarity * depth weight=0.053	levenshtein=0.05 wordnet=0.857 similarity=0.454 depth weight=0.1625 similarity * depth weight=0.074	levenshtein=0.053 wordnet=0.75 similarity=0.402 depth weight=0.2125 similarity * depth weight=0.085	levenshtein=0.05 wordnet=0.75 similarity=0.4 depth weight=0.2125 similarity * depth weight=0.085	levenshtein=0.056 wordnet=0.8 similarity=0.428 depth weight=0.2125 similarity * depth weight=0.091

ภาพที่ 4.9 เมตริกซ์ L และค่าสัมประสิทธิ์ความคล้ายคลึงกันระหว่างเทอมในออนโทโลจีโปรไฟล์งานวิจัย

จากภาพที่ 4.10 เมื่อระบบทำการคำนวณค่าความคล้ายคลึงกันระหว่างเทอมที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัย 2 โปรไฟล์ เสร็จเรียบร้อยแล้ว ระบบจะให้ผู้ใช้กรอกค่าขีดแบ่ง เริ่มต้นความคล้าย ดังภาพที่ 4.10 เพื่อทำการสร้างเมตริกซ์ $Map_{0\text{-}1}$ ที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์ความคล้ายคลึงในเมตริกซ์ L กับค่าขีดแบ่งเริ่มต้นความคล้าย ดังภาพที่ 4.11

Input Threshold Value to Calculate Research Profile Matching : 0.35 Calculate

ภาพที่ 4.10 ช่องสำหรับกรอกค่าขีดแบ่งเริ่มต้นความคล้าย

จากภาพที่ 4.11 เมื่อระบบทำการสร้างเมตริกซ์ $Map_{0\text{-}1}$ เสร็จเรียบร้อยแล้ว จะแสดงค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านระหว่างออนโทโลจีโปรไฟล์งานวิจัย 2 โปรไฟล์ ดังภาพที่ 4.12 พร้อมทั้งคำนวณค่าความสอดคล้องกันของนักวิจัย ดังภาพที่ 4.13

Author : Kijsirikul, B Subject area : Computer Science
Author : Ratanamahatana, CA Subject area : Computer Science

Input Threshold Value to Calculate Research Profile Matching : 0.35 Calculate

psychological feature,entity	psychological feature,physical entity	psychological feature,abstraction	psychological feature,attribute	psychological feature,group	psychological feature,process
levenshtein=0.05 wordnet=0.667 similarity=0.359 depth weight=0.1125 similarity * depth weight =0.04 0	levenshtein=0.077 wordnet=0.571 similarity=0.324 depth weight=0.1625 similarity * depth weight =0.053 0	levenshtein=0.05 wordnet=0.857 similarity=0.454 depth weight=0.1625 similarity * depth weight =0.074 0	levenshtein=0.053 wordnet=0.75 similarity=0.402 depth weight=0.2125 similarity * depth weight =0.085 0	levenshtein=0.05 wordnet=0.75 similarity=0.4 depth weight=0.2125 similarity * depth weight =0.085 0	levenshtein=0.056 wordnet=0.8 similarity=0.428 depth weight=0.2125 similarity * depth weight =0.091 0

ภาพที่ 4.11 เมตริกซ์ $Map_{0\text{-}1}$ ที่ได้จากการเปรียบเทียบค่าในเมตริกซ์ L กับค่าขีดแบ่งเริ่มต้นความคล้าย

Neighbor number	Matched pairs	Score value
STn1	{(change,change)(change,damage)(change,deformation)(change,warping)(change of magnitude,deformation)(reduction,change)(reduction,detection)(reduction,damage)(reduction,deformation)}	0.4
STn2	{(basic cognitive process,basic cognitive process)(change,damage)(change,deformation)(change,warping)(change of magnitude,deformation)(reduction,change)(reduction,detection)(reduction,damage)(reduction,deformation)}	0.382
STn3	{(basic cognitive process,basic cognitive process)(change,change)(change,deformation)(change,warping)(change of magnitude,deformation)(reduction,change)(reduction,detection)(reduction,damage)(reduction,deformation)}	0.408
STn4	{(basic cognitive process,basic cognitive process)(change,change)(change,damage)(change,warping)(change of magnitude,deformation)(reduction,change)(reduction,detection)(reduction,damage)(reduction,deformation)}	0.41
STn5	{(basic cognitive process,basic cognitive process)(change,change)(change,damage)(change,deformation)(change of magnitude,deformation)(reduction,change)(reduction,detection)(reduction,damage)(reduction,deformation)}	0.409

ภาพที่ 4.12 ค่าความคล้ายคลึงกันของโครงสร้างเพื่อนบ้านระหว่างออนโทโลจีโปรไฟล์งานวิจัย 2 โปรไฟล์

MLMA+ with Depth Weights

The researchers have common interest score : 0.411

Matched fields : (basic cognitive process,basic cognitive process) (change,change) (change,damage) (change,deformation) (change,warping) (reduction,change) (reduction,detection) (reduction,damage) (reduction,deformation) (change of magnitude,deformation)

Fields of common interest in summary : basic cognitive process , change , change of magnitude , damage , deformation , detection , reduction , warping

ภาพที่ 4.13 ค่าความสอดคล้องกันของนักวิจัย

ในกรณีที่ผลลัพธ์ความสอดคล้องกันของนักวิจัยยังไม่เป็นที่พอใจของผู้ใช้ ผู้ใช้สามารถคำนวณความสอดคล้องกันของนักวิจัยใหม่ได้ จากภาพที่ 4.11 ผู้ใช้สามารถกรอกค่าขีดแบ่งเริ่มต้นความคล้ายใหม่ได้ที่ช่อง Input Threshold Value to Calculate Research Profile Matching แล้วคลิกที่ปุ่ม Calculate เพื่อคำนวณค่าความสอดคล้องกันของนักวิจัยอีกครั้ง

4.2 สภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ

สภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย ผู้วิจัยได้ใช้ฮาร์ดแวร์ และซอฟต์แวร์ ซึ่งมีรายละเอียดดังนี้

4.2.1 ฮาร์ดแวร์

ฮาร์ดแวร์สำหรับพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย มีรายละเอียดดังนี้

3. เครื่องคอมพิวเตอร์โน้ตบุ๊ก หน่วยประมวลผล อินเทล คอร์ i3 2.13 กิกะเฮิร์ทซ์ (Intel(R) Core(TM) i3 CPU 2.13 GHz)
4. หน่วยความจำ (Memory) 4.0 กิกะไบต์
5. จานบันทึกแบบแข็ง (Hard disk) ความจุ 300 กิกะไบต์

4.2.2 ซอฟต์แวร์

ซอฟต์แวร์สำหรับพัฒนาเครื่องมือเพื่อการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยสำหรับนักวิจัย มีรายละเอียดดังนี้

1. ระบบปฏิบัติการไมโครซอฟท์วินโดวส์ 7 อัลทิเมต (Microsoft Windows 7 Ultimate)
2. ระบบปฏิบัติการลีนุกซ์อูบุนตู (Ubuntu) สำหรับคำนวณค่าความคล้ายคลึงกันโดยใช้เวิร์ดเน็ต ซึ่งฟังก์ชันการทำงานถูกพัฒนาโดยใช้โปรแกรมเชิงวัตถุของภาษาเพิร์ล (Perl)
3. อาปาเช่ เว็บเซิร์ฟเวอร์ (Apache Web Server) รุ่น 2.2.8 สำหรับให้เครื่องคอมพิวเตอร์มีคุณสมบัติเป็นเว็บเซิร์ฟเวอร์ (Web Server)
4. พีเอชพีมายแอดมิน (phpMyAdmin) รุ่น 2.10.3 สำหรับบริหารจัดการฐานข้อมูลมายซีควล (MySQL) ผ่านเว็บเบราว์เซอร์ (Web Browser)
5. มายซีควล รุ่น 5.0.51b สำหรับจัดเก็บข้อมูลที่เกี่ยวข้อง เช่น ชื่อนักวิจัย คำสำคัญ หมวดหมู่สาขาวิชา และปีที่ผลงานวิจัยได้รับการตีพิมพ์ เป็นต้น
6. ภาษาพีเอชพี (PHP Language) รุ่น 5.2.6 สำหรับพัฒนาเว็บเพจแบบไดนามิก
7. โปรแกรมอะโดบี ดรีมวีฟเวอร์ รุ่น CS5 (Adobe Dreamweaver CS5) สำหรับเขียนรหัสต้นฉบับ (Source Code) ภาษาพีเอชพี (PHP)

บทที่ 5

ผลการทดสอบ

ผู้วิจัยทดสอบระเบียบวิธีวิเคราะห์ความสนใจร่วมกันของนักวิจัยและการทำงานของเครื่องมือโดยการสร้างและแสดงออนโทโลจีโปรไฟล์งานวิจัย พร้อมทั้งหาค่าความสอดคล้องของนักวิจัยที่สนใจในหมวดหมู่สาขาวิขาที่เหมือนกัน ใกล้เคียงกัน และแตกต่างกัน แล้วเปรียบเทียบค่า Score Value รวมทั้งวัดความสามารถของเครื่องมือในด้านสมรรถนะในการประมวลผลและในด้านความแม่นยำ เมื่อคำสำคัญของงานวิจัยมีจำนวนต่าง ๆ ซึ่งมีรายละเอียดดังนี้

5.1 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่เหมือนกัน

การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่เหมือนกัน ผู้วิจัยได้ทดลองจับคู่นักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA ซึ่งอยู่ในหมวดหมู่สาขาวิชา Computer Science ซึ่งจะได้ออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 4.5 และภาพที่ 4.6 เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

ผู้วิจัยได้ทดลองสร้างออนโทโลจีโปรไฟล์งานวิจัยเมื่อคำสำคัญมีจำนวนต่าง ๆ ซึ่งทำให้ได้คำศัพท์ที่ปรากฏอยู่ในออนโทโลจีโปรไฟล์งานวิจัย มีจำนวนที่แตกต่างกัน ดังตารางที่ 5.1

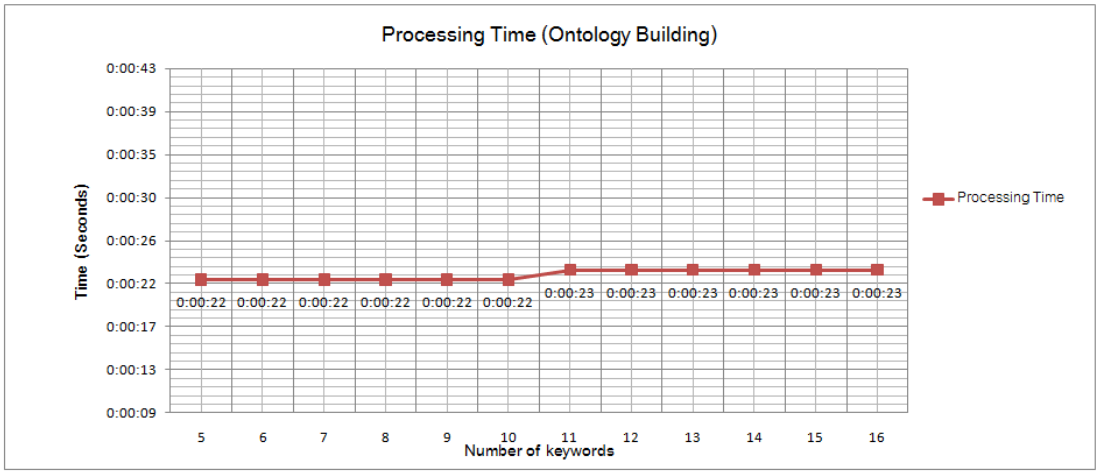
ตารางที่ 5.1 จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA

Number of keywords	Number of terms in Research Profile (Kijsirikul, B)	Number of terms in Research Profile (Ratanamahatana, CA)
5	20	28
6	24	33
7	30	35
8	36	39
9	38	46
10	44	47

ตารางที่ 5.1 จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Number of terms in Research Profile (Kijsirikul, B)	Number of terms in Research Profile (Ratanamahatana, CA)
11	45	49
12	50	55
13	57	56
14	61	60
15	68	64
16	69	67

จากจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA ดังตารางที่ 5.1 จะใช้เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 5.1



ภาพที่ 5.1 เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ratanamahatana, CA

จากนั้นผู้วิจัยได้ทดลองหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.35 ทำให้ได้ค่าความสอดคล้องของนักวิจัย ดังตารางที่ 5.2

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA

Number of keywords	Matching Score	Matched Pairs
5	0.411	(basic cognitive process, basic cognitive process), (change, change), (change, damage), (change, deformation), (change, warping), (reduction, change), (reduction, detection), (reduction, damage), (reduction, deformation), (change of magnitude, deformation)
6	0.409	(event,event), (knowledge,knowledge), (process,process), (basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (change of magnitude,deformation), (reduction,change), (reduction,damage), (reduction,deformation)
7	0.409	(event,event), (knowledge,knowledge), (process,process), (basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,deformation), (change,warping), (change of magnitude,deformation), (reduction,change), (reduction,damage), (reduction,deformation),

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
8	0.455	(act,act), (action,action), (action,reduction), (basic cognitive process,basic cognitive process), (change,change), (change,damage), (change,change of magnitude), (change,deformation), (change,reduction), (change,warping), (change of magnitude,change), (change of magnitude,change of magnitude), (change of magnitude,deformation), (change of magnitude,reduction), (reduction,action), (reduction,change), (reduction,damage), (reduction,change of magnitude), (reduction,deformation), (reduction,reduction), (process,process)

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
9	0.44	(basic cognitive process,basic cognitive process), (action,action), (action,reduction), (knowledge,knowledge), (change,change), (change,damage), (change,reduction), (change,deformation), (change,warping), (change of magnitude,change of magnitude), (change of magnitude,reduction), (change of magnitude,deformation), (reduction,section), (reduction,action), (reduction,change), (reduction,passage), (reduction,damage), (reduction,change of magnitude), (reduction,reduction), (reduction,deformation), (reduction,clipping), (method,damage)
10	0.481	(knowledge,knowledge), (action,action), (basic cognitive process,basic cognitive process), (change,change), (content,content), (change of magnitude,change of magnitude), (reduction,change of magnitude), (reduction,deformation), (reduction,clipping), (reduction,reduction), (change of magnitude,reduction)

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
11	0.481	(action,action), (knowledge,knowledge), (basic cognitive process,basic cognitive process), (change,change), (content,content), (change of magnitude,change of magnitude), (reduction,change of magnitude), (reduction,clipping), (reduction,deformation), (reduction,reduction), (change of magnitude,reduction)
12	0.499	(knowledge,knowledge), (action,action), (basic cognitive process,basic cognitive process), (change,change), (content,content), (knowledge domain,knowledge domain), (change of magnitude,change of magnitude), (field,field), (field,engineering), (field,computer science), (reduction,change of magnitude), (reduction,clipping), (reduction,reduction), (reduction,deformation), (engineering,field), (engineering,engineering), (engineering,computer science), (computer science,field), (computer science,engineering), (computer science,computer science), (change of magnitude,reduction)

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
13	0.486	(basic cognitive process,basic cognitive process), (action,action), (knowledge,knowledge), (content,content), (change,change), (knowledge domain,knowledge domain), (change of magnitude,change of magnitude), (field,field), (field,engineering), (field,computer science), (reduction,change of magnitude), (reduction,clipping), (reduction,reduction), (reduction,deformation), (engineering,field), (engineering,engineering), (engineering,computer science), (machine,action), (machine,image), (computer science,field), (computer science,engineering), (computer science,computer science), (change of magnitude,reduction)

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
14	0.485	(basic cognitive process,basic cognitive process), (action,action), (knowledge,knowledge), (content,content), (organization,organization), (change,change), (knowledge domain,knowledge domain), (change of magnitude,change of magnitude), (reduction,change of magnitude), (reduction,clipping), (reduction,reduction), (reduction,deformation), (field,field), (field,engineering), (field,computer science), (machine,action), (machine,image), (machine,organization), (engineering,field), (engineering,engineering), (engineering,computer science), (computer science,field), (computer science,engineering), (computer science,computer science), (change of magnitude,reduction)

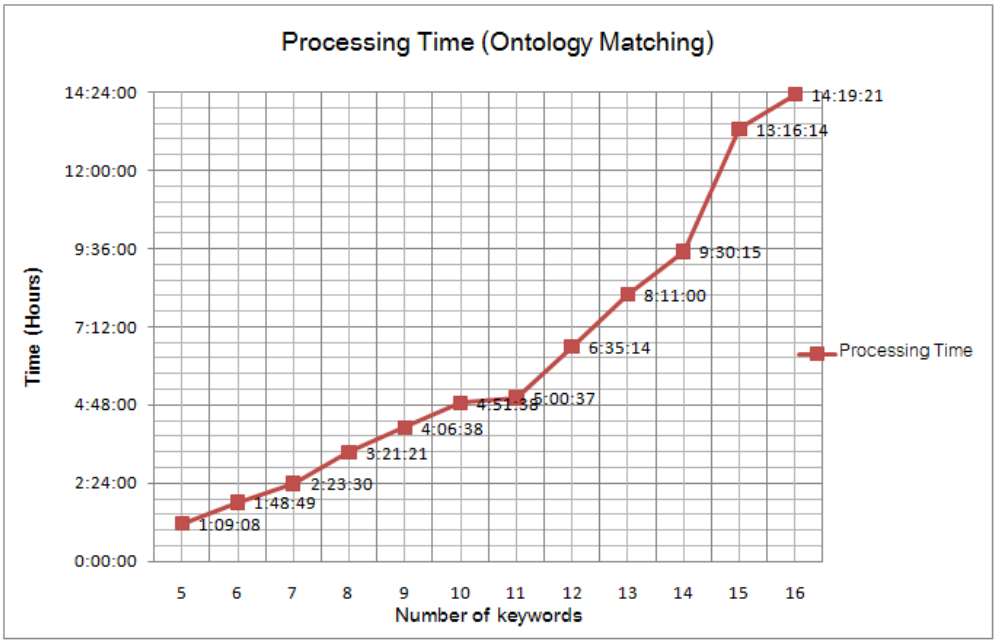
ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
15	0.454	(knowledge,knowledge), (event,event), (action,action), (basic cognitive process,basic cognitive process), (change,change), (content,content), (act,act), (organization,organization), (activity,activity), (knowledge domain,knowledge domain), (change of magnitude,change of magnitude), (change of magnitude,reduction), (reduction,change of magnitude), (reduction,reduction), (reduction,deformation), (reduction,clipping), (field,field), (field,engineering), (field,computer science), (work,search), (work,scaling), (investigation,search), (investigation,scaling), (machine,action), (machine,organization), (machine,image), (engineering,field), (engineering,engineering), (engineering,computer science), (analysis,action), (analysis,detection), (analysis,search), (analysis,scaling), (computer science,field), (computer science,engineering), (computer science,computer science), (activity,scaling)

ตารางที่ 5.2 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA (ต่อ)

Number of keywords	Matching Score	Matched Pairs
16	0.451	(event,event), (action,action), (basic cognitive process,basic cognitive process), (knowledge,knowledge), (act,act), (organization,organization), (content,content), (change,change), (knowledge domain,knowledge domain), (activity,activity), (change of magnitude,change of magnitude), (change of magnitude,reduction), (work,search), (work,scaling), (field,field), (field,engineering), (field,computer science), (reduction,change of magnitude), (reduction,deformation), (reduction,reduction), (reduction,clipping), (investigation,search), (investigation,scaling), (machine,action), (machine,organization), (machine,image), (engineering,field), (engineering,engineering), (engineering,computer science), (computer science,field), (computer science,engineering), (computer science,computer science), (analysis,action), (analysis,detection), (analysis,search), (analysis,scaling), (analysis,Indexing), (activity,scaling)

จากตารางที่ 5.2 จะพบว่าจำนวนของ Matched Pairs ที่ปรากฏในตารางจะขึ้นอยู่กับคำสำคัญที่ใช้ในการสร้างออนโทโลจีโปรไฟล์นักวิจัย และจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย หากจำนวนคำศัพท์มีจำนวนมากมักจะทำให้จำนวนของ Matched Pairs มากขึ้นด้วย รวมทั้งอาจขึ้นอยู่กับค่าความเหมือนกันระหว่างเทอมที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย 2 โปรไฟล์ และในการหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA จะใช้เวลาในการประมวลผลเพิ่มขึ้นตามจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย ดังภาพที่ 5.2



ภาพที่ 5.2 เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Ratanamahatana, CA

5.2 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่ใกล้เคียงกัน

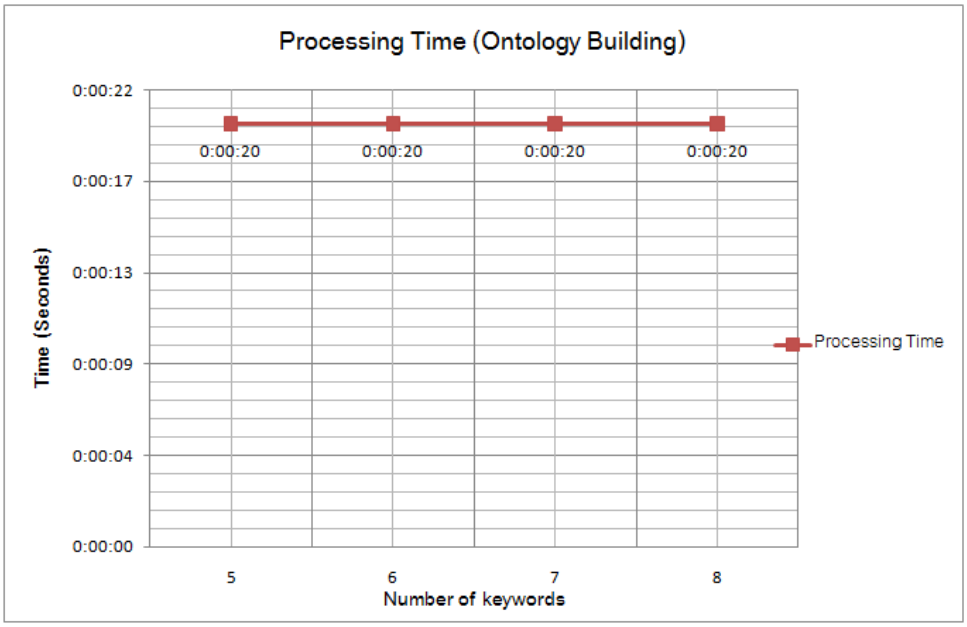
การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่ใกล้เคียงกัน ผู้วิจัยได้ทดลองจับคู่นักวิจัยระหว่าง Kijsirikul, B ซึ่งอยู่ในหมวดหมู่สาขาวิชา Computer Science และ Sudsang, A ซึ่งอยู่ในหมวดหมู่สาขาวิชา Robotics จะได้ออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 4.5 และภาพที่ 4.7 เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

ผู้วิจัยได้ทดลองสร้างออนโทโลจีโปรไฟล์งานวิจัยเมื่อคำสำคัญมีจำนวนต่าง ๆ ซึ่งทำให้ได้คำศัพท์ที่ปรากฏอยู่ในออนโทโลจีโปรไฟล์งานวิจัย มีจำนวนที่แตกต่างกัน ดังตารางที่ 5.3

ตารางที่ 5.3 จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Sudsang, A

Number of keywords	Number of terms in Research Profile (Kijsirikul, B)	Number of terms in Research Profile (Sudsang, A)
5	20	31
6	24	37
7	30	39
8	36	43

จากจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Sudsang, A ดังตารางที่ 5.3 จะใช้เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 5.3



ภาพที่ 5.3 เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Sudsang, A

จากนั้นผู้วิจัยได้ทดลองหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.30 ทำให้ได้ค่าความสอดคล้องของนักวิจัย ดังตารางที่ 5.4

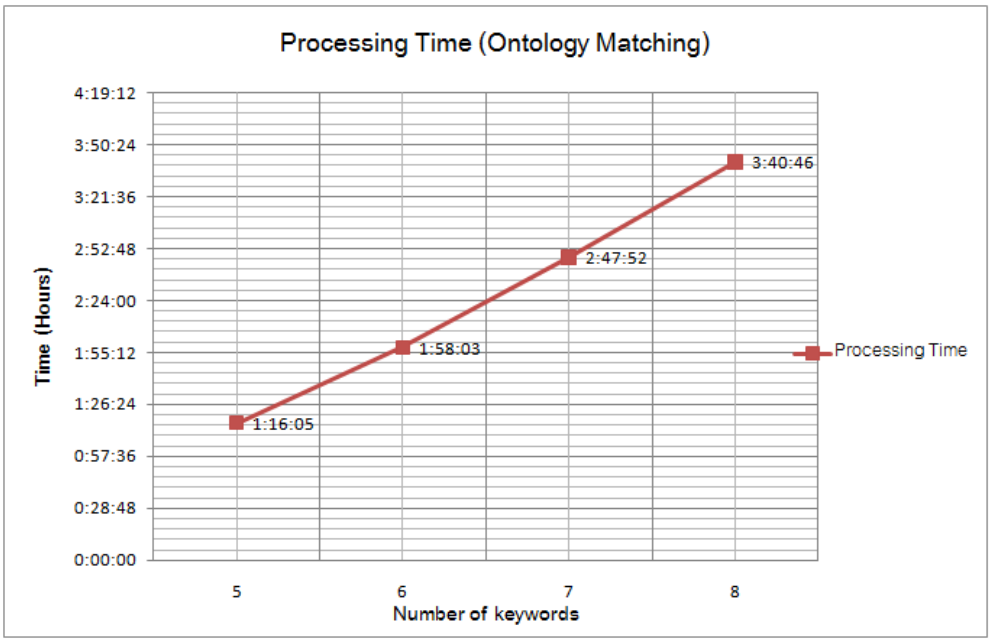
ตารางที่ 5.4 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A

Number of keywords	Matching Score	Matched Pairs
5	0.339	(process,process), (method,Robotics), (change,span), (reduction,span), (reduction,Grasping), (method,Grasping)
6	0.331	(event,event), (knowledge,knowledge), (act,act), (process,process), (change,influence), (change,span), (change,Grasping), (learning,Grasping), (method,field), (method,Grasping), (method,Robotics), (change of magnitude,span), (reduction,act), (reduction,causing), (reduction,span), (reduction,manipulation), (reduction,Grasping), (action,Grasping)
7	0.329	(event,event), (knowledge,knowledge), (act,act), (process,process), (method,field), (method,Grasping), (method,Robotics), (learning,Grasping), (change,influence), (change,span), (change,Grasping), (change of magnitude,span), (reduction,act), (reduction,causing), (reduction,propulsion), (reduction,pushing), (reduction,span), (reduction,manipulation), (reduction,Grasping), (action,Grasping)

ตารางที่ 5.4 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A (ต่อ)

Number of keywords	Matching Score	Matched Pairs
8	0.326	(event,event), (knowledge,knowledge), (act,act), (process,process), (action,activity), (learning,planning), (learning,Grasping), (method,field), (method,Grasping), (method,Robotics), (change,preparation), (change,influence), (change,span), (change,Grasping), (change of magnitude,preparation), (change of magnitude,span), (reduction,act), (reduction,causing), (reduction,activity), (reduction,propulsion), (reduction,preparation), (reduction,pushing), (reduction,span), (reduction,planning), (reduction,manipulation), (reduction,Grasping), (action,Grasping)

จากตารางที่ 5.4 จะพบว่าจำนวนของ Matched Pairs ที่ปรากฏในตารางจะขึ้นอยู่กับคำสำคัญที่ใช้ในการสร้างออนโทโลจีโปรไฟล์นักวิจัย และจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย หากจำนวนคำศัพท์มีจำนวนมากมักจะทำให้จำนวนของ Matched Pairs มากขึ้นด้วย รวมทั้งอาจขึ้นอยู่กับค่าความเหมือนกันระหว่างเทอมที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย 2 โปรไฟล์ และในการหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A จะใช้เวลาในการประมวลผลเพิ่มขึ้นตามจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย ดังภาพที่ 5.4



ภาพที่ 5.4 เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A

5.3 การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่แตกต่างกัน

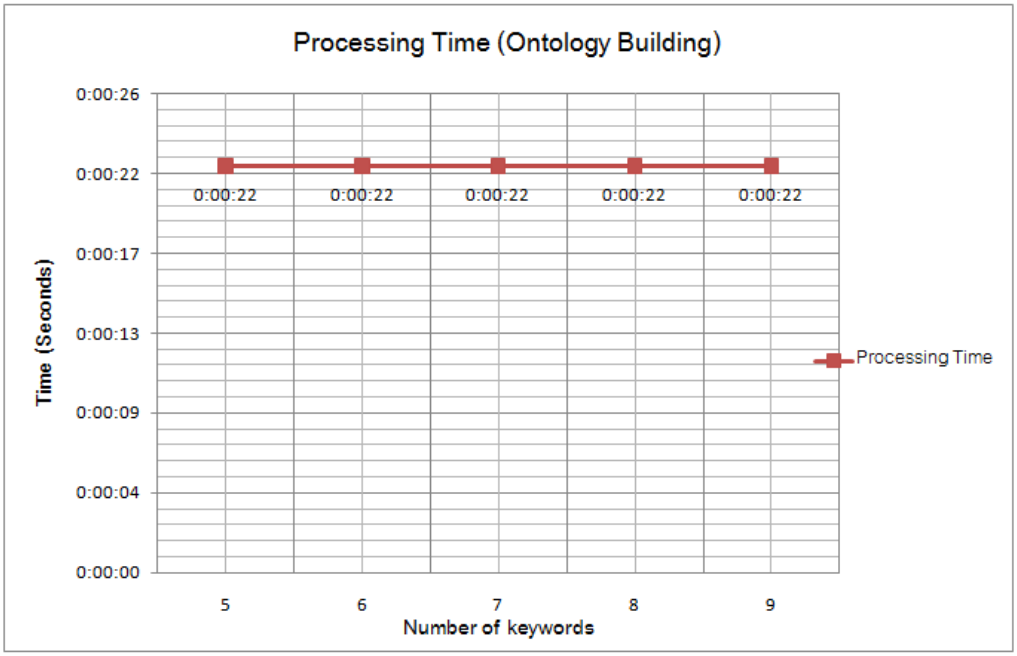
การสร้างออนโทโลจีโปรไฟล์งานวิจัย และหาค่าความสอดคล้องของนักวิจัยในหมวดหมู่สาขาวิชาที่แตกต่างกัน ผู้วิจัยได้ทดลองจับคู่นักวิจัยระหว่าง Kijsirikul, B ซึ่งอยู่ในหมวดหมู่สาขาวิชา Computer Science และ Ruppin, E ซึ่งอยู่ในหมวดหมู่สาขาวิชา Biochemistry & Molecular Biology จะได้ออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 4.5 และภาพที่ 4.8 เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

ผู้วิจัยได้ทดลองสร้างออนโทโลจีโปรไฟล์งานวิจัยเมื่อคำสำคัญมีจำนวนต่าง ๆ ซึ่งทำให้ได้คำศัพท์ที่ปรากฏอยู่ในออนโทโลจีโปรไฟล์งานวิจัย มีจำนวนที่แตกต่างกัน ดังตารางที่ 5.5

ตารางที่ 5.5 จำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ruppin, E

Number of keywords	Number of Terms in Research Profile (Kijsirikul, B)	Number of terms in Research Profile (Ruppin, E)
5	20	26
6	24	36
7	30	40
8	36	43
9	38	47

จากจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ruppin, E ดังตารางที่ 5.5 จะใช้เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัย ดังภาพที่ 5.5



ภาพที่ 5.5 เวลาในการสร้างออนโทโลจีโปรไฟล์งานวิจัยของ Kijsirikul, B และ Ruppin, E

จากนั้นผู้วิจัยได้ทดลองหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.30 ทำให้ได้ค่าความสอดคล้องของนักวิจัย ดังตารางที่ 5.6

ตารางที่ 5.6 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E

Number of keywords	Matching Score	Matched Pairs
5	0.332	(event,event), (process,process), (act,act), (action,activity), (method,activity), (method,preparation), (method,planning), (method,programming), (change,preparation) (change,planning), (learning,planning), (change of magnitude,preparation), (reduction,act), (reduction,organization), (reduction,activity), (reduction,preparation), (reduction,planning),(reduction,programming), (action,planning)
6	0.345	(event,event), (system,system), (knowledge,knowledge), (act,act), (network,network), (process,process), (learning,planning), (method,biology), (reduction,preparation), (action,activity)

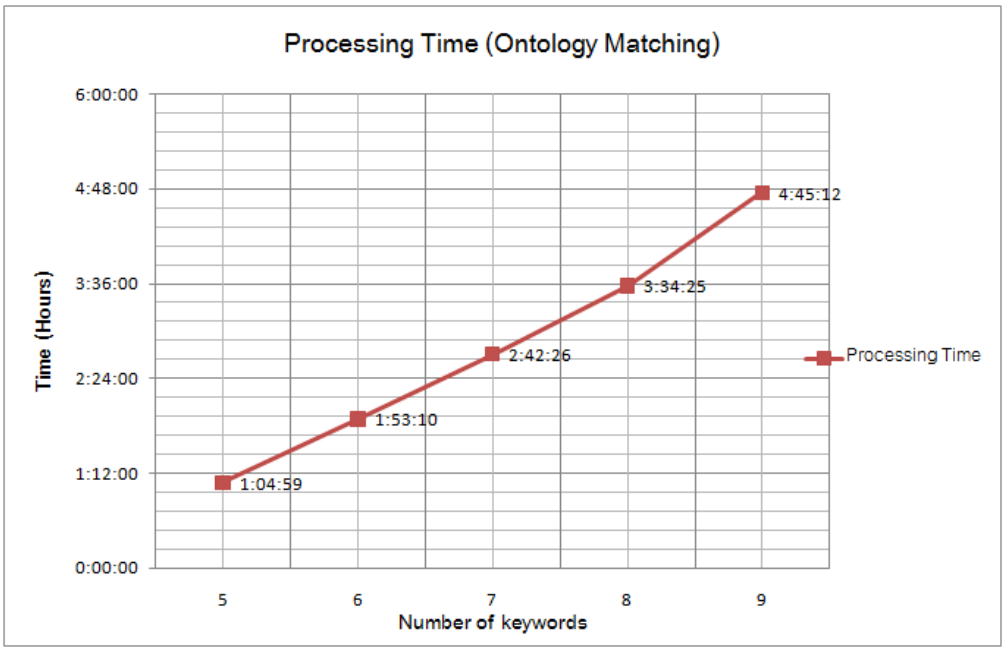
ตารางที่ 5.6 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E (ต่อ)

Number of keywords	Matching Score	Matched Pairs
7	0.357	(event,event), (system,system), (knowledge,knowledge), (network,network), (act,act), (process,process), (method,biology), (logic programming,programming), (learning,planning), (change,change), (change,variation), (change of magnitude,change), (change of magnitude,variation), (reduction,change), (reduction,variation), (reduction,preparation), (action,activity)
8	0.371	(event,event), (knowledge,knowledge), (system,system), (act,act), (network,network), (process,process), (basic cognitive process,basic cognitive process), (learning,learning), (learning,planning), (method,biology), (logic programming,programming), (change,change), (change,variation), (change of magnitude,change), (change of magnitude,variation), (reduction,change), (reduction,variation), (reduction,preparation), (action,activity)

ตารางที่ 5.6 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E (ต่อ)

Number of keywords	Matching Score	Matched Pairs
9	0.347	(psychological feature,psychological feature), (process,process), (basic cognitive process,basic cognitive process), (system,system), (knowledge,knowledge), (power,field), (power,science), (power,biology), (change,change), (change,variation), (learning,learning), (learning,work), (network,network), (difficulty,work), (difficulty,investigation), (difficulty,analysis), (change of magnitude,change), (change of magnitude,variation), (know-how,science), (reduction,change), (reduction,variation), (reduction,preparation), (method,system), (method,act), (method,activity), (method,variation), (method,work), (method,preparation), (method,field), (method,planning) (method,investigation) (method,science) (method,analysis) (method,programming) (method,natural science) (method,life science) (method,biology) (method,molecular biology) (logic programming,planning) (logic programming,programming) (group,group)

จากตารางที่ 5.6 จะพบว่าจำนวนของ Matched Pairs ที่ปรากฏในตารางจะขึ้นอยู่กับคำสำคัญที่ใช้ในการสร้างออนโทโลจีโปรไฟล์นักวิจัย และจำนวนคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย หากจำนวนคำศัพท์มีจำนวนมากมักจะทำให้จำนวนของ Matched Pairs มากขึ้นด้วย รวมทั้งอาจขึ้นอยู่กับค่าความเหมือนกันระหว่างเทอมที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย 2 โปรไฟล์ และในการหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Sudsang, A จะใช้เวลาในการประมวลผลเพิ่มขึ้นตามจำนวนคำศัพท์ที่ปรากฏในออนโทโลจี โปรไฟล์นักวิจัย ดังภาพที่ 5.6



ภาพที่ 5.6 เวลาในการหาค่าความสอดคล้องนักวิจัยระหว่าง Kijsirikul, B และ Ruppin, E

5.4 การวัดประสิทธิภาพของการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย

ผู้วิจัยดำเนินการวัดประสิทธิภาพของการวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัยที่พัฒนาขึ้นมา คือ MLMA+ with Depth Weights โดยเปรียบเทียบกับแนวคิดการจับคู่ออนโทโลจีที่มีอยู่แล้ว คือ MLMA+ โดยใช้ชุดข้อมูลทดสอบของโอเออีไอ 2011 (OAEI Benchmark 2011) ซึ่งเป็นเกณฑ์สำหรับการวัดเปรียบเทียบและการประเมินเทคโนโลยีการจับคู่ออนโทโลจี โดยจะจัดเตรียมชุดข้อมูลทดสอบในโดเมนที่เกี่ยวกับบรรณานุกรม แต่ละชุดข้อมูลทดสอบจะประกอบด้วยออนโทโลจีทดสอบ (Test Ontology) ในรูปแบบภาษาอาวล์ (OWL) และข้อมูลความ

สอดคล้องกับออนโทโลจีอ้างอิง (Reference Alignment) แต่ละออนโทโลจีทดสอบจะเป็นออนโทโลจีที่ถูกเปลี่ยนแปลงจากออนโทโลจีอ้างอิง (Reference Ontology) ชุดที่ 101 และมีข้อมูลความสอดคล้องกับออนโทโลจีอ้างอิง ซึ่งจะระบุว่าคำใดในออนโทโลจีทดสอบเหมือนกับคำใดในออนโทโลจีอ้างอิง (Expected Alignments) สำหรับใช้ในการทดสอบประสิทธิภาพของอัลกอริทึมในการจับคู่ออนโทโลจีว่าสามารถจับคู่คำที่อยู่ในออนโทโลจีทดสอบกับคำที่อยู่ในออนโทโลจีอ้างอิง ซึ่งถูกระบุว่ามีความสอดคล้องกันได้หรือไม่ ดังนั้นในชุดทดสอบที่ 101 ออนโทโลจีอ้างอิงจะถูกจับคู่กับตัวมันเอง และในชุดข้อมูลทดสอบอื่น ๆ คำในออนโทโลจีทดสอบของชุดข้อมูลทดสอบนั้น ๆ จะถูกจับคู่กับคำที่ปรากฏในออนโทโลจีอ้างอิง ในการนี้ผู้วิจัยใช้ชุดข้อมูลทดสอบ 101- 104, 201-210, 221-247 และ 301-304 เฉพาะข้อมูลที่เป็นซูเปอร์คลาสและซับคลาสของชุดทดสอบ

ชุดข้อมูลทดสอบที่ 101-104 เป็นชุดข้อมูลทดสอบที่ประกอบด้วยออนโทโลจีที่มีการเปลี่ยนตำแหน่งหรือแทนที่ตำแหน่งโครงสร้างของอาวด์ ชุดข้อมูลทดสอบที่ 201-210 เป็นชุดข้อมูลทดสอบที่มีการเปลี่ยนชื่อคลาสหรือซับคลาส โดยใช้วิธีเช่น การสุ่มสายอักขระ การสะกดคำผิด การนำคำพ้องความหมายเข้ามาแทนที่ และการเปลี่ยนจากภาษา อังกฤษไปเป็นภาษาอื่น เป็นต้น ชุดข้อมูลทดสอบที่ 221-247 เป็นชุดข้อมูลทดสอบที่มีการเปลี่ยนแปลงโครงสร้างของคลาสหรือซับคลาส เช่น การยุบซับคลาสมารวมกับคลาสเพื่อไม่ให้มีลำดับชั้น ขยายลำดับชั้นของออนโทโลจีไม่มีอินสแตนซ์ และไม่มีคุณสมบัติของคลาส เป็นต้น และชุดข้อมูลทดสอบที่ 301-304 เป็นชุดข้อมูลทดสอบที่ประกอบด้วยออนโทโลจีข้อมูลทางบรรณานุกรมที่มีใช้อยู่จริง

ผู้วิจัยพิจารณาประสิทธิภาพของอัลกอริทึมจากค่าความแม่นยำ (Precision) ค่าเรียกคืน (Recall) และเอฟเมเชอร์ (F-measure) ซึ่งมีสูตรการคำนวณดัง สมการที่ (5.1) - (5.3)

$$\text{Precision} = \frac{\text{no. of expected alignments found as matched by algo}}{\text{no. of matched pairs found by algo}} \tag{5.1}$$

$$\text{Recall} = \frac{\text{no. of expected alignments found as matched by algo}}{\text{no. of expected alignments}} \tag{5.2}$$

$$\text{F}-\text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.3}$$

ผู้วิจัยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.5 ซึ่งมีความเหมาะสมกับค่าสัมประสิทธิ์ความคล้ายคลึงระหว่างเทอมที่ปรากฏในออนโทโลจีโปรไฟล์นักวิจัย ทำให้ได้ผลการวัดประสิทธิภาพของอัลกอริทึม ดังเช่นตารางที่ 5.7

ตารางที่ 5.7 ผลการวัดประสิทธิภาพของการจับคู่ออนโทโลจี

Test Set	MLMA+			MLMA+ with Depth Weights		
	Prec.	Rec.	F-measure	Prec.	Rec.	F-measure
#101-104	0.74	1.0	0.85	0.93	0.84	0.88
#201-210	0.35	0.24	0.26	0.68	0.18	0.27
#221-247	0.71	0.99	0.82	0.94	0.66	0.75
#301-304	0.56	0.75	0.64	0.90	0.57	0.68
Average	**0.59**	**0.74**	**0.64**	**0.86**	**0.56**	**0.64**

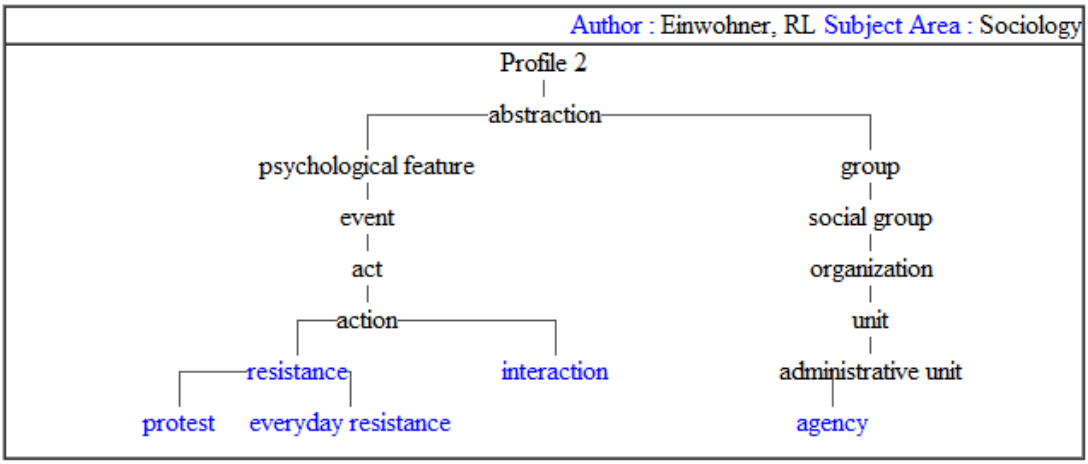
จากตารางที่ 5.7 จะเห็นได้ว่าแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ with Depth Weights จะให้ค่าเอฟเมเชอร์ที่มีคุณภาพเช่นเดียวกับแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ และให้ค่าความแม่นยำมากกว่า แต่ให้ค่าเรียกคืนน้อยกว่า เนื่องจาก MLMA+ with Depth Weights จะค้นพบคู่คำศัพท์ที่สอดคล้องกันมีจำนวนน้อยกว่า MLMA+ ด้วยเหตุนี้จึงทำให้ได้ค่าความแม่นยำที่สูงกว่า แต่ในขณะเดียวกันค่าเรียกคืนจะได้รับผลกระทบด้วย ทั้งนี้ เป็นเพราะข้อมูลความสอดคล้องกับออนโทโลจีอ้างอิงของชุดข้อมูลทดสอบในโอเออีไอ 2011 เพียงแต่แสดงรายการเฉพาะคู่คำที่เหมือนกัน ตัวอย่างเช่น ถ้าออนโทโลจีที่ใช้ในการทดสอบและออนโทโลจีที่ใช้อ้างอิงมีคู่คำที่เหมือนกัน จะหมายความว่าอัลกอริทึมควรจะสามารถค้นพบว่าคู่คำนี้เหมือนกัน แต่แนวคิดการจับคู่โดย MLMA+ with Depth Weights จะพิจารณาตำแหน่งของคำที่ปรากฏในออนโทโลจีด้วย ดังนั้นคู่คำที่เหมือนกันทุกประการอาจจะถูกพิจารณาว่าไม่เหมือนกัน หากมีตำแหน่งอยู่ใกล้กับรากของออนโทโลจี จึงถูกลดทอนความคล้ายคลึงลงไป จากการที่คำที่อยู่ใกล้กับรากของออนโทโลจีจะเป็นคำประเภทนามธรรม การค้นพบคู่คำนามธรรมที่เหมือนกันจำนวนมากโดยวิธี MLMA+ จึงไม่เป็นประโยชน์ต่อการนำผลลัพธ์ไปใช้งานต่อไปนัก ในขณะที่การลดทอนความคล้ายคลึงของคู่คำนามธรรมที่อยู่ใกล้กับรากของออนโทโลจีโดยวิธี MLMA+ with Depth Weights ทำให้ผลลัพธ์ของการจับคู่มีปริมาณลดลง แต่เป็นผลลัพธ์ที่มีคุณค่าต่อการนำไปใช้งานมากกว่า

เพราะจะได้คู่คำที่เหมือนกันและมีความเฉพาะเจาะจงมากกว่า จึงสะท้อนการจับคู่ความสนใจในด้านงานวิจัยที่เจาะจงและชัดเจนกว่า

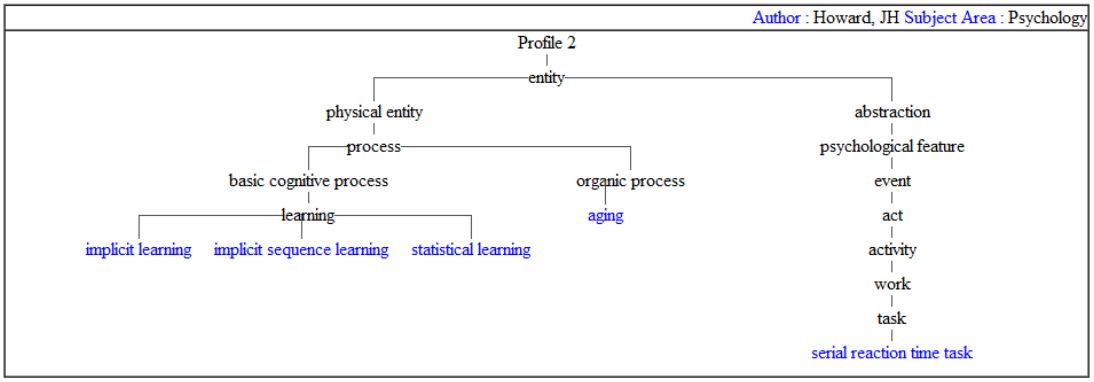
นอกจากนี้ค่าขีดแบ่งเริ่มต้นความคล้ายในขั้นตอนของการหาค่าความสอดคล้องของนักวิจัย ถือเป็นพารามิเตอร์ที่มีผลกระทบต่อค่าความแม่นยำและค่าเรียกคืน หากกำหนดค่าขีดแบ่งเริ่มต้นความคล้ายสูงเกินไปมักจะทำให้ได้คู่คำที่เป็นคำเดียวกัน เช่น (psychological feature, psychological feature) เป็นต้น ซึ่งแสดงให้เห็นว่าคู่คำดังกล่าวปรากฏอยู่ใกล้รากของออนโทโลจีและเป็นคำเชิงนามธรรม ในทางตรงกันข้ามหากกำหนดค่าขีดแบ่งเริ่มต้นความคล้ายต่ำเกินไปจะทำให้ค้นพบคู่คำที่สอดคล้องกันเป็นจำนวนมาก ถึงแม้คู่คำเหล่านั้นจะถูกลดทอนความคล้ายคลึงลงด้วยน้ำหนักความลึกแล้ว แต่ยังมีค่าความคล้ายคลึงที่ผ่านเกณฑ์ค่าขีดแบ่งเริ่มต้นความคล้ายอยู่ ดังนั้นค่าขีดแบ่งเริ่มต้นความคล้ายที่ใช้สำหรับข้อมูลทดสอบในงานวิจัยนี้จะเลือกใช้ค่าที่สูงที่สุดแต่ไม่ทำให้ค้นพบคู่คำที่เป็นคำเดียวกันทั้งหมด และคู่คำที่ค้นพบไม่มีจำนวนมากจนเกินไปซึ่งจะทำให้การใช้ประโยชน์หรือตีความจากผลลัพธ์ที่ได้ทำได้สะดวกขึ้น

5.5 การทดสอบประโยชน์ของแนวคิด MLMA+ with Depth Weights

เพื่อเป็นการทดสอบประโยชน์ของแนวคิด MLMA+ with Depth Weights ในที่นี้ขอยกตัวอย่างออนโทโลจีโปรไฟล์งานวิจัย พร้อมทั้งหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B ในหมวดหมู่สาขาวิชา Computer Science และ Einwohner, RL ในหมวดหมู่สาขาวิชา Sociology และหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B ในหมวดหมู่สาขาวิชา Computer Science และ Howard, JH ในหมวดหมู่สาขาวิชา Psychology ดังภาพที่ 4.5, 5.7 และ 5.8 เมื่อใช้คำสำคัญเริ่มต้น 5 คำ ตามแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ และแนวคิด MLMA+ with Depth Weights เพื่อพิจารณาความสนใจร่วมกันระหว่างนักวิจัยดังกล่าวข้างต้นว่ามีความสนใจร่วมกันในด้านงานวิจัยมากน้อยเพียงใด



ภาพที่ 5.7 ออนโทโลจีโปรไฟล์งานวิจัยของ Einwohner, RL เมื่อใช้คำสำคัญเริ่มต้น 5 คำ



ภาพที่ 5.8 ออนโทโลจีโปรไฟล์งานวิจัยของ Howard, JH เมื่อใช้คำสำคัญเริ่มต้น 5 คำ

จากภาพที่ 4.5 และ 5.7 ผู้วิจัยได้ทดลองหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Einwohner, RL ตามแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.5 และการจับคู่ออนโทโลจีตามแนวคิด MLMA+ with Depth Weights โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.35 ทำให้ได้ค่าความสอดคล้องของนักวิจัยดังตารางที่ 5.8 และ 5.9

ตารางที่ 5.8 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Einwohner, RL บนพื้นฐาน MLMA+

Number of keywords	Matching Score	Matched Pairs
5	0.633	(psychological feature,psychological feature), (event,event), (event,act), (power,event), (process,action), (process,protest), (act,event), (act,act), (act,action), (act,agency), (action,act), (action,action), (action,resistance), (action,interaction), (action,agency), (change,event), (change,act), (change,action), (change,resistance), (reduction,action), (process,act)

ตารางที่ 5.9 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Einwohner, RL บนพื้นฐาน MLMA+ with Depth Weights

Number of keywords	Matching Score	Matched Pairs
5	0.396	(act,act), (act,agency), (action,action), (action,resistance), (action,interaction), (action,protest), (action,agency), (change,action), (change,resistance), (change,interaction), (change,agency), (change of magnitude,resistance), (change of magnitude,interaction), (reduction,action), (reduction,resistance), (reduction,interaction), (reduction,protest), (reduction,agency), (process,protest)

จากภาพที่ 4.5 และ 5.8 ผู้วิจัยได้ทดลองหาค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Howard, JH ตามแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.5 และการจับคู่ออนโทโลจีตามแนวคิด MLMA+ with Depth Weights โดยใช้ค่าขีดแบ่งเริ่มต้นความคล้ายเท่ากับ 0.35 ทำให้ได้ค่าสอดคล้องของนักวิจัย ดังตารางที่ 5.10 และ 5.11

ตารางที่ 5.10 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Howard, JH บนพื้นฐาน MLMA+

Number of keywords	Matching Score	Matched Pairs
5	0.638	(psychological feature,psychological feature), (event,event), (event,act), (knowledge,process), (power,process), (power,event), (power,work), (process,process), (process,act), (process,activity), (process,work), (act,process), (act,event), (act,act), (act,activity), (act,work), (act,task), (action,process), (action,act), (action,activity), (action,work), (action,task), (basic cognitive process,basic cognitive process), (method,activity), (change,event), (change,act), (learning,learning), (learning,work), (process,organic process)

ตารางที่ 5.11 ค่าความสอดคล้องของนักวิจัยระหว่าง Kijsirikul, B และ Howard, JH บนพื้นฐาน MLMA+ with Depth Weights

Number of keywords	Matching Score	Matched Pairs
5	0.426	(process,process), (act,act), (action,task), (basic cognitive process,basic cognitive process), (method,task), (learning,learning), (learning,work)

จากตารางที่ 5.8-5.11 สามารถเปรียบเทียบผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกันเป็นดังตารางที่ 5.12 ซึ่งทั้งสองวิธีต่างพบว่า Kijsirikul, B มีความสนใจร่วมกันกับ Howard, JH มากกว่า Einwohner, RL

ตารางที่ 5.12 เปรียบเทียบผลการวิเคราะห์หรือจับคู่นักวิจัยที่สนใจงานวิจัยที่เป็นเรื่องเดียวกัน บนพื้นฐาน MLMA+ และ MLMA+ with Depth Weights

Algorithm	Author 1	Author 2	Matching Score
MLMA+	Kijsirikul, B	Einwohner, RL	0.633
		Howard, JH	0.638
MLMA+ with Depth Weights	Kijsirikul, B	Einwohner, RL	0.396
		Howard, JH	0.426

จากตารางที่ 5.12 สามารถอภิปรายได้ว่าแนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ เมื่อพิจารณา Matching Score จะเห็นว่า Kijsirikul, B มีความสนใจร่วมกันกับ Howard, JH มากกว่า Einwohner, RL แต่มากกว่าเพียงเล็กน้อย คือ ค่าความแตกต่างของ Matching Score มีค่าเพียง 0.005 จึงสามารถพิจารณาได้ว่า Kijsirikul, B มีความสนใจที่สอดคล้องกับนักวิจัยทั้งสองพอ ๆ กัน แต่แนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ with Depth Weights จะสามารถแยกได้ชัดเจนขึ้นว่า Kijsirikul, B มีความสนใจร่วมกันกับ Howard, JH มากกว่า Einwohner, RL เนื่องจาก Matching Score ระหว่างนักวิจัยดังกล่าวมีความแตกต่างมากขึ้น คือ ค่าความแตกต่างของ

Matching Score เท่ากับ 0.03 ตัวอย่างนี้แสดงให้เห็นว่าการใช้ Depth Weights สามารถลดทอนความคล้ายคลึงของคู่คำที่เหมือนกันแต่เป็นคำนามธรรมที่อยู่ใกล้กับรากของออนโทโลจี โดยที่คู่คำเหล่านั้นจะถูกกรองออกจากผลลัพธ์ของการจับคู่ออนโทโลจี ผลลัพธ์ที่ได้จึงมีเฉพาะคู่คำที่เหมือนกันมากและเฉพาะเจาะจงเหลืออยู่ ทำให้ได้ค่าคะแนน Matching Score ที่ชัดเจนขึ้นในการแยกแยะความคล้ายคลึงของความสนใจของนักวิจัย ในขณะที่เมื่อใช้วิธี MLMA+ คู่คำนามธรรมเหล่านั้นไม่ได้ถูกกรองออกไป ทำให้สามารถส่งผลให้คะแนน Matching Score ค่อนข้างใกล้เคียงกันได้ จึงแยกแยะความคล้ายคลึงของความสนใจของนักวิจัยได้ไม่ชัดเจนนัก

อีกทั้งเมื่อพิจารณาผลลัพธ์ของการจับคู่ความสนใจระหว่าง Kijsirikul, B และ Howard, JH ในตารางที่ 5.10 และ 5.11 แนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ จะได้คู่คำจำนวนมากถึง 29 คู่ แต่แนวคิดการจับคู่ออนโทโลจีด้วยวิธี MLMA+ with Depth Weights จะทำให้ได้คู่คำที่ลดลงเหลือ 7 คู่ อันเป็นผลจากการกรองคู่คำที่เป็นนามธรรมและอยู่ในระดับบนของออนโทโลจีออกไปเป็นจำนวนมาก ซึ่ง 6 ใน 7 คู่คำที่ปรากฏในตารางที่ 5.11 จะปรากฏในตารางที่ 5.10 ด้วย และได้คู่คำที่ไม่ปรากฏในตารางที่ 5.10 จำนวน 1 คู่ คือ (method, task) ซึ่งเป็นคู่คำที่เฉพาะเจาะจงและอยู่ในระดับล่างของออนโทโลจี ผู้วิจัยเห็นว่า แม้ว่าวิธี MLMA+ with Depth Weights จะให้ผลลัพธ์การจับคู่ออนโทโลจีเป็นจำนวนลดลง มีค่าเรียกคืนลดลง แต่ผลลัพธ์เป็นคู่คำที่มีคุณค่าต่อการนำไปใช้ประโยชน์มากกว่าและชัดเจนกว่าผลลัพธ์จากวิธี MLMA+

บทที่ 6

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

6.1 สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอระเบียบวิธีและพัฒนาเครื่องมือสำหรับสร้างออนโทโลจีโปรไฟล์งานวิจัย และวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย เพื่อวิเคราะห์หรือจับคู่นักวิจัยที่สนใจในหมวดหมู่สาขาวิชาที่เหมือนกัน ใกล้เคียงกัน และแตกต่างกัน ซึ่งจะช่วยให้นักวิจัยเกิดการพัฒนาความรู้ความเข้าใจเกี่ยวกับงานวิจัยที่ให้ความสนใจ และเป็นการเพิ่มขีดความสามารถในการนำความรู้ความเข้าใจไปประยุกต์ใช้กับงานวิจัยต่อไป รวมทั้งทำให้รู้จักนักวิจัยอื่น ๆ ที่มีความสนใจในด้านวิจัยที่คล้ายคลึงกัน เพื่อแลกเปลี่ยนประสบการณ์ซึ่งและกัน ซึ่งจะช่วยให้เกิดแนวคิดที่หลากหลายในการแก้ไขปัญหาที่เกิดขึ้นระหว่างการทำงานวิจัย

ผู้วิจัยดำเนินการพัฒนาเครื่องมือดังกล่าวโดยอาศัยแนวคิดเกี่ยวกับออนโทโลจีเข้ามาช่วยดำเนินการ เนื่องจากแนวคิดเกี่ยวกับออนโทโลจีถูกนำมาใช้งานอย่างแพร่หลายในงานวิจัยด้านเว็บเชิงความหมาย การจัดการองค์ความรู้ และการค้นคืนสารสนเทศ เป็นต้น

การพัฒนาเครื่องมือดังกล่าวได้แบ่งการทำงานออกเป็น 2 ส่วน คือ ส่วนสร้างออนโทโลจีโปรไฟล์งานวิจัย และส่วนจับคู่ออนโทโลจีโปรไฟล์งานวิจัย โดยประยุกต์แนวคิดการสร้างออนโทโลจีและการจับคู่ออนโทโลจีที่มีอยู่แล้ว ในส่วนของการสร้างออนโทโลจีโปรไฟล์งานวิจัยได้ใช้คำสำคัญจากผลงานวิจัยที่ได้รับการตีพิมพ์และจัดทำดัชนีโดยฐานข้อมูลไอเอสไอร่วมกับความสัมพันธ์เชิงความหมายของคำศัพท์แบบ Hypernym/Hyponym จากเวิร์ดเน็ต สำหรับการเปรียบเทียบความเหมือนกันระหว่างออนโทโลจีโปรไฟล์ งานวิจัยได้เสนอวิธี MLMA+ with Depth Weights ซึ่งจะพิจารณาค่าเฉลี่ยระหว่างค่าความเหมือนกันทางสายอักขระร่วมกันกับค่าความเหมือนกันเชิงความหมายจากเวิร์ดเน็ต รวมทั้งได้เพิ่มแนวคิดของน้ำหนักความลึกในการจับคู่ออนโทโลจี เพื่อมุ่งความสนใจไปยังคู่คำที่อยู่ในระดับล่างของออนโทโลจีโปรไฟล์งานวิจัย เนื่องจากเป็นคู่คำที่แสดงความสนใจร่วมกันที่เฉพาะเจาะจงมากกว่าคู่คำในระดับบนซึ่งอยู่ใกล้กับรากของออนโทโลจี

ในการวัดประสิทธิภาพของเครื่องมือพบว่าจำนวนคำสำคัญเริ่มต้นไม่ส่งผลมากนักต่อการสร้างออนโทโลจีโปรไฟล์งานวิจัย แต่จะส่งผลต่อการจับคู่ออนโทโลจีเพราะจำนวนคำสำคัญเริ่มต้นยิ่งมีจำนวนมากยิ่งทำให้ออนโทโลจีโปรไฟล์มีขนาดใหญ่ขึ้นและใช้เวลาจับคู่นานขึ้น ผู้วิจัยดำเนินการวัดประสิทธิภาพระหว่างแนวคิดที่นำเสนอกับแนวคิดการจับคู่ออนโทโลจี

ที่มีอยู่แล้ว โดยใช้มาตรฐานของโอเออีไอ 2011 ซึ่งแนวคิดของน้ำหนักความลึกจะให้ค่าความแม่นยำที่ดีขึ้น แต่ค่าเรียกคืนจะลดลง

6.2 อภิปรายผลการวิจัย

จากการนำเสนอระเบียบวิธีและพัฒนาเครื่องมือสำหรับสร้างออนโทโลจีโปรไฟล์งานวิจัยและวิเคราะห์ความสนใจร่วมกันในด้านงานวิจัย สามารถอภิปรายผลการวิจัยได้ดังนี้

6.2.1 การสร้างออนโทโลจีโปรไฟล์งานวิจัยใช้คำสำคัญจากผลงานวิจัยที่ได้รับการตีพิมพ์และจัดทำดัชนีโดยฐานข้อมูลไอเอสไอร่วมกับความสัมพันธ์เชิงความหมายของคำศัพท์แบบ Hypernym/Hyponym จากเวิร์ดเน็ต คำสำคัญจากผลงานวิจัยจะเป็นคำศัพท์ทางเทคนิค หรือเป็นคำศัพท์ที่เฉพาะเจาะจง ซึ่งคำสำคัญดังกล่าวมักไม่พบในเวิร์ดเน็ต เนื่องจากเวิร์ดเน็ตเป็นฐานข้อมูลชุดคำศัพท์ภาษาอังกฤษ และไม่มีคำศัพท์ที่เฉพาะเจาะจงหรือคำศัพท์ที่เกี่ยวข้องกับงานวิจัยมากนัก ในการนี้ผู้วิจัยได้นำหมวดหมู่สาขาวิชา (Subject Area) หรือคำนามหลักของคำสำคัญเข้ามาช่วยสร้างออนโทโลจีโปรไฟล์งานวิจัย

6.2.2 จากการที่นำคำศัพท์จากเวิร์ดเน็ตเข้ามาช่วยสร้างออนโทโลจีโปรไฟล์งานวิจัย ในกรณีที่คำสำคัญจากผลงานวิจัยของนักวิจัยเป็นคำศัพท์ทางเทคนิค หรือเป็นคำศัพท์ที่เฉพาะเจาะจง ดังนั้นจึงทำให้เทอมของคำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัยโดยส่วนมากจะเป็นคำศัพท์พื้นฐานและคำเชิงนามธรรม

6.2.3 แนวคิดการจับคู่ออนโทโลจีที่มีอยู่แล้ว และแนวคิดที่ผู้วิจัยนำเสนอจะให้ผลลัพธ์แบบ False Positive เมื่อปรากฏคำพ้องรูปในออนโทโลจีโปรไฟล์งานวิจัยทั้ง 2 ออนโทโลจี เช่น คำว่า “date” ในออนโทโลจีข้อมูลบรรณานุกรมจะหมายถึงวันที่ แต่หากอยู่ในออนโทโลจีไวน์จะหมายถึงผลอินทผลัม แต่จะพบสองคำนี้เหมือนกันหากนำ 2 ออนโทโลจีนี้มาจับคู่กัน

6.2.4 สืบเนื่องจากการที่คำศัพท์ที่ปรากฏในออนโทโลจีโปรไฟล์งานวิจัย มักจะเป็นคำศัพท์พื้นฐานและคำเชิงนามธรรม ผลลัพธ์การจับคู่คำที่เป็นคำพื้นฐานและคำเชิงนามธรรมอาจจะไม่เป็นประโยชน์นักในแง่การสรุปผลความสนใจร่วมกันของนักวิจัย ดังนั้นถึงแม้ว่าแนวคิดเกี่ยวกับน้ำหนักความลึกจะทำให้การจับคู่ออนโทโลจีได้ค่าเรียกคืนลดลง แต่ถือว่าช่วยกรองผลลัพธ์การจับคู่ให้มีจำนวนลดลง ซึ่งน่าจะเป็นประโยชน์ต่อการตีความและนำไปใช้ประโยชน์ต่อไป

6.3 ข้อเสนอแนะ

ประเด็นที่สามารถศึกษา และสามารถทำการวิจัยเพิ่มเติมได้ในอนาคต มีดังนี้

6.3.1 ศึกษาชุดข้อมูลคำศัพท์ภาษาอังกฤษอื่น ๆ ที่มีคำศัพท์ทางเทคนิค หรือคำศัพท์ที่เฉพาะเจาะจง และมีความความสัมพันธ์เชิงความหมายของคำศัพท์แบบ Hypernym/Hyponym ที่เหมือนกันกับเวิร์ดเน็ต เข้ามาช่วยสร้างออนโทโลจีโปรไฟล์งานวิจัย

6.3.2 เพิ่มแนวคิดการจับคู่ออนโทโลจีโดยพิจารณาโครงสร้าง และบริบท (Context) ของเทอมที่ปรากฏระหว่างออนโทโลจี 2 ออนโทโลจี

6.3.3 ทดลองเพิ่มขึ้นเกี่ยวกับการจับคู่การทำงานร่วมกันของนักวิจัยที่อยู่คนละหมวดหมู่สาขาวิชา เนื่องจากเวิร์ดเน็ตเป็นฐานข้อมูลที่มีคำศัพท์พื้นฐาน ดังนั้นน่าจะเป็นประโยชน์โดยเฉพาะต่อการเชื่อมโยงโปรไฟล์ของนักวิจัยที่อยู่คนละวงการและดูเหมือนไม่เกี่ยวข้องกันโดยเชื่อมโยงผ่านคำศัพท์พื้นฐานเพราะคำศัพท์เฉพาะทางของแต่ละสาขาวิชาจะแตกต่างกันอย่างมากอยู่แล้ว

รายการอ้างอิง

[1] Cox, A. What are communities of practice? A comparative review of four seminal works. Journal of Information Science 31 (December 2005): 527-540.

[2] Community of Practice. [Online]. Available from: http://en.wikipedia.org/wiki/Community_of_practice [2010, October]

[3] Okubo, Y. Bibliometric Indicators and Analysis of Research Systems: Methods and Examples. Paris, OECD Publishing, 1997.

[4] An, Y. J., Geller, J., Wu, Y., and Chun, S. A. Automatic Generation of Ontology from the Deep Web. Proceeding of the 18th International Workshop on Database and Expert Systems Applications (DEXA'07), 2007, pp.470-474.

[5] Alasoud, A., Haarslev, V., and Shiri, N. An empirical comparison of ontology matching techniques. Journal of Information Science 35 (March 2009): 379-397.

[6] Alasoud, A., Haarslev, V., and Shiri, N. An Effective Ontology Matching Technique. Proceeding of the 17th International Conference on Foundations of Intelligent Systems, 2008, pp.585-590.

[7] Yang, H., Liu, S., Fu, P., Qin, H., and Gu, L. A Semantic Distance Measure for Matching Web Services. Proceeding of Computational Intelligence and Software Engineering (CiSE), 2009, pp.1-3.

[8] McGuinness, D. L., and Harmelen, F. V. OWL Web Ontology Language Overview. [Online]. Available from: http://www.w3.org/TR/owl-features/ [2010, September]

[9] Daconta, M. C., Obrst, L. J., and Smith, K. T. Understanding Ontologies. In Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. pp.181-237. Indianapolis: Wiley Publishing, 2003.

[10] Guarino, N. Formal Ontology and Information Systems. In Proceedings of FOIS, 1998, pp.3-15. Amsterdam: IOS Press, 1998.

[11] Miller, G. A. Wordnet: A lexical Database for English. [Online]. Available from: http://wordnet.princeton.edu/ [2010, November]

[12] Assem, M. V., Gangemi, A., and Schreiber, G. RDF/OWL Representation of Wordnet. [Online]. Available from: http://www.w3.org/TR/wordnet-rdf/ [2010, November]

[13] Navarro, G. A Guided Tour to Approximate String Matching. ACM Computing Surveys. 33 (March 2001): 31-88.

[14] Levenshtein Edit Distance. [Online]. Available from: http://www.miislita.com/searchito/levenshtein-edit-distance.html [2011, May]

[15] Winkler, W. E. Overview of Record Linkage and Current Reseach Directions. [Online]. Available from: http://www.census.gov/srd/papers/pdf/rrs2006-02.pdf [2010, December]

[16] Yancey, W. E. An Adaptive String Comparator for Record Linkage. [Online]. Available from: http://www.census.gov/srd/papers/pdf/rrs2004-02.pdf [2010, December]

[17] Pedersen, T., Patwardhan, S., and Michelizzi, J. Wordnet:Similarity – Measuring the Relatedness of Concepts. Proceedings of 19th National Conference on Airtifical Intelegence (AAAAI 2004), 2004, pp.1024-1025.

[18] WordNet::Similarity. [Online]. Available from: http://sourceforge.net/projects/wn-similarity/ [2011, April]

[19] Pedersen, T., Patwardhan, S., Michelizzi, J., and Banerjee, S., WordNet::Similarity. [Online]. Available from: http://wn-similarity.sourceforge.net/ [2011, April]

[20] Hearst, M. A. Automated Discovery of WordNet Relation. Wordnet: An Electronic Lexical Database and Some of its Applications, pp.132-152. MIT Press,1998.

[21] Diederich, J., and Iofciu T. Finding Communities of Practice from User Profiles Based On Folksonomies. Proceedings of the 1st International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice (TEL-CoPs'06), 2006.

[22] Sriharee, N., and Punnarut, R. Constructing Semantic Campus for Academic Collaboration. Proceedings of 2nd International ISWC+ASWC ExpertFinder Workshop (FEWS2007), pp.23-32, 2007.

[23] Punnarut, R., and Sriharee, G. A researcher Expertise Search system using Ontology-Based Data Mining. Proceedings 7th Asia-Pacific Conference on Conceptual Modelling (APCCM 2010), pp.71-78, 2010.

[24] ISI Web of Knowledge. [Online]. Available from: http://www.isiknowledge.com/ [2010, September]

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวนวรัตน์ คำเสียง เกิดเมื่อวันที่ 22 มกราคม พ.ศ. 2522 ที่จังหวัดชลบุรี สำเร็จการศึกษาระดับปริญญาบัณฑิต หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ จากมหาวิทยาลัยบูรพา เมื่อ พ.ศ. 2543 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาศาสตร์คอมพิวเตอร์ ณ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ. 2552