

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

วิทยานิพนธ์ฉบับนี้ นำเสนอการวัดความสัมพันธ์ระหว่างคำโดยใช้ความหมายของประโยค ซึ่งการเปรียบเทียบความหมายระหว่างประโยคจำเป็นต้องใช้ทฤษฎีที่เกี่ยวข้องดังนี้

2.1.1 Semantic Similarity และ Semantic Relatedness

Semantic Similarity คือการเปรียบเทียบคำโดยพิจารณาว่าเหมือนกันในเชิงความหมายหรือไม่ เช่น “car” กับ “bicycle” จะมองว่ามีความหมายเหมือนกัน คือเป็นพาหนะที่มีล้อเหมือนกัน ในการเปรียบเทียบจะได้ผลเป็นมีความหมายเหมือนกันหากผลการเปรียบเทียบที่ได้อยู่ภายในค่า Threshold ที่กำหนด แต่หากเกินค่า Threshold จะถือว่าไม่มีความหมายเหมือนกัน

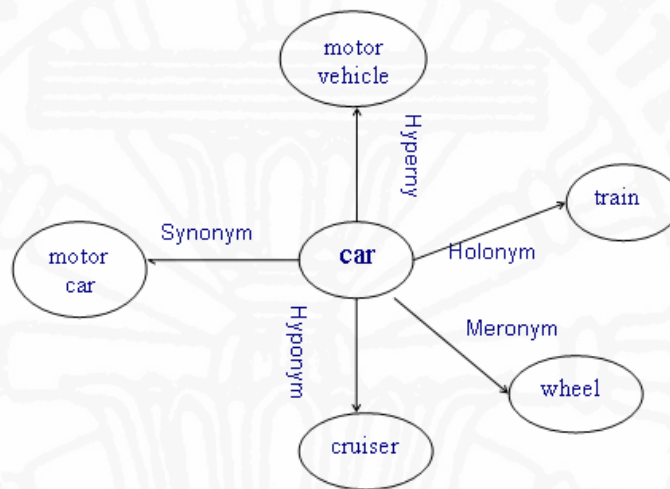
Semantic Relatedness คือการเปรียบเทียบคำในเชิงความหมาย โดยจะพิจารณาว่ามีความหมายที่มีความสัมพันธ์กันหรือไม่ ซึ่งการพิจารณาความสัมพันธ์เชิงความหมายนั้นเป็นแนวคิดที่เป็นทั่วไปมากกว่าการพิจารณาความหมายเหมือนกัน คือในขณะที่คำที่ไม่มีความหมายเหมือนกันในเชิงความหมาย แต่อาจจะเป็นคำที่มีความสัมพันธ์กันในเชิงความหมาย เช่น “car” กับ “wheel” ไม่มีความหมายที่เหมือนกัน แต่มีความสัมพันธ์กันในเชิงความหมายในลักษณะที่ “car” มี “wheel” เป็นส่วนประกอบ (Meronym) ซึ่งความสัมพันธ์กันในเชิงความหมายนั้นสามารถแบ่งลักษณะความสัมพันธ์ได้ดังนี้

- Synonym มีความหมายเหมือนกัน
- Antonym มีความหมายที่ตรงกันข้ามกัน
- Hypernym มีความหมายที่เป็นแบบทั่วไปกว่า
- Hyponym มีความหมายที่เป็นแบบเฉพาะเจาะจงกว่า
- Holonym มีความสัมพันธ์ทางความหมายในลักษณะที่เป็นส่วนประกอบหรือสมาชิกของ

- Meronym มีความสัมพันธ์ทางความหมายในลักษณะที่มีเป็นส่วนประกอบหรือมีเป็นสมาชิก

ภาพที่ 2.1

แสดงความสัมพันธ์ลักษณะต่าง ๆ ของ “car” กับคำอื่น ๆ



ซึ่งการเปรียบเทียบเพื่อหาความสัมพันธ์ในเชิงความหมายนี้ใช้ในการสรุปใจความเอกสาร (Text Summarization) และระบบค้นคืนสารสนเทศ (Information Retrieval) ซึ่งจะทำให้ระบบงานค้นคืนสารสนเทศสามารถค้นคืนข้อมูลหรือเอกสารที่มีความสัมพันธ์กันได้ครอบคลุมความต้องการมากขึ้น

ในการเปรียบเทียบเพื่อหาความสัมพันธ์ในเชิงความหมายจะพิจารณาจาก

- โครงข่ายของเวิร์ดเน็ต (WordNet) ซึ่งคำสองคำที่มีระยะห่างน้อยจะถือว่ามีค่าความสัมพันธ์กันมาก โดยหาความสัมพันธ์เชิงความหมายของคำจากการคำนวณจากเส้นทางระหว่างคำ (Synset) ที่สั้นที่สุดในเวิร์ดเน็ต งานวิจัยในกลุ่มนี้มีข้อดีคือง่าย แต่มีข้อจำกัดตรงที่คำที่นำมาเปรียบเทียบกันนั้นต้องมีอยู่ในฐานข้อมูล WordNet นอกจากนี้ คำที่เขียนเหมือนกันอาจอยู่ได้หลายที่ใน WordNet ขึ้นอยู่กับความหมาย เช่น คำว่า “mouse” อาจหมายถึง “หนูที่เป็นสัตว์” หรือ “อุปกรณ์คอมพิวเตอร์” ทำให้คำที่ได้อาจจะผิดพลาดได้

- พิจารณาความสัมพันธ์จากคำข้อมูล (Information Content หรือ Corpus Based) หรือข้อความที่เนื้อหาของคำทั้งสองสถิตอยู่ โดยวิธีนี้จะใช้การคำนวณทางสถิติ เช่น การกระจายของคำ, จำนวนครั้งที่เกิดขึ้น, ตำแหน่งของคำ ด้วยวิธีนี้สามารถแก้ปัญหาของคำที่ไม่มีใน WordNet ของกลุ่มแรกได้ แต่อย่างไรก็ตามวิธีการนี้จำเป็นต้องมีการสร้างคลังข้อมูลตัวอักษร (Corpus) และผลการคำนวณที่ได้ก็จะขึ้นอยู่กับขนาดและข้อมูลที่มีในคลังข้อมูลนั้นๆ
- พิจารณาจากการเปรียบเทียบคำจำกัดความหรือคำนิยามหรือคำแปลของคำทั้งคู่ (Gloss based) โดยดูคำนวณจากคำที่ใช้ร่วมกัน (Shared word) ในการอธิบายความหมายของคำทั้งคู่ ซึ่งในการหาค่าสัมพันธ์ในเชิงความหมายโดยพิจารณาจากข้อความที่เป็นเนื้อหาหรือการคำแปลที่มีการใช้คำร่วมกัน ต่อมาได้มีการปรับปรุงโดยในการเปรียบเทียบกันระหว่างเนื้อหาหรือคำแปลนั้นนอกจากจะพิจารณาความเหมือนของคำที่ใช้ร่วมกันแล้วยังมีการพิจารณาความสัมพันธ์ในเชิงความหมายที่เป็น Synonym, Hypernym และ Hyponym ด้วย โดยใช้เวิร์ดเน็ตเป็นฐานความรู้ (Knowledge-based) ในการพิจารณาเปรียบเทียบหาความสัมพันธ์ในระดับต่าง ๆ

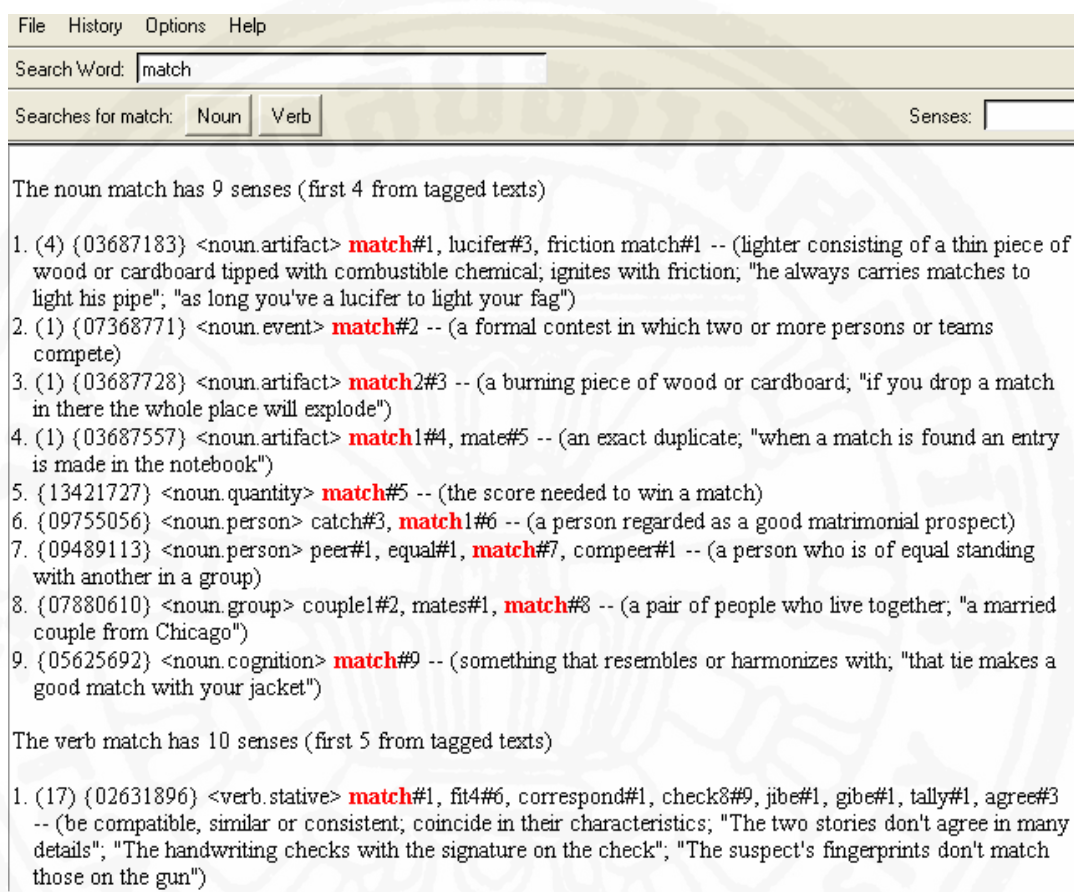
2.1.3 เวิร์ดเน็ต (WordNet)

เวิร์ดเน็ต (Cognitive Science Laboratory at Princeton University) เป็นฐานข้อมูลพจนานุกรมภาษาอังกฤษผู้ใช้สามารถdownload มาใช้หรือใช้งาน ในลักษณะออนไลน์ได้ โดยเวิร์ดเน็ตได้รวมคำในภาษาอังกฤษที่มีความหมาย(Sense) เหมือนกันเข้าเป็นกลุ่มเรียกว่า Synset เช่น “good” “right” และ “ripe” เป็น Synset เดียวกัน มีความหมายเหมือนกัน คือ “most suitable or right for a particular purpose”

ในระหว่าง Synset จะสัมพันธ์กันโดยมีความสัมพันธ์เชิงความหมายหลายรูปแบบคำ เช่น มีความหมายเหมือนกัน (Synonym) หรือมีความหมายตรงกันข้าม (Antonym) ซึ่งความสัมพันธ์จะขึ้นอยู่กับชนิดของคำโดยจะแยกความแตกต่างกันระหว่างคำที่เป็น Noun, Verb, Adjective และ Adverb ตามหลักไวยากรณ์ Synset ที่มีความสัมพันธ์กันเชิงความหมายจะมีหมายเลขที่เชื่อมต่อกัน ทุก Synset ที่เชื่อมต่อกันโดยตัวเลขของความเกี่ยวข้องเชิงความหมาย

ภาพที่ 2.2

ตัวอย่างการแสดงผลของเวิร์ดเน็ตจากการค้นหาคำว่า match



คำ 1 คำในเวิร์ดเน็ตอาจมีได้หลายความหมาย นั่นคือคำ 1 คำ อาจปรากฏอยู่ในหลาย Synset โดยที่แต่ละ Synset มีหมายเลขกำกับที่แตกต่างกัน ในขณะเดียวกันอาจจะมีคำหลายคำที่มีความหมายเดียวกัน จากภาพที่ 2.1 จะเห็นได้ว่า “match” ที่เป็นคำนามมี 9 ความหมาย และ “match” ที่เป็นคำกริยามี 10 ความหมาย ในขณะที่บางความหมายมีคำที่มีความหมายเหมือนกัน (Synonym) ด้วย เช่น ว่า “match” ที่เป็นคำนามความหมายที่ 1 มี lucifer, friction, match เป็นคำที่มีความหมายเดียวกัน

เนื่องจากเวิร์ดเน็ตเป็นฐานข้อมูลพจนานุกรมซึ่งได้รวบรวมความหมายต่าง ๆ ของทั้งที่เป็น Noun, Verb, Adjective และ Adverb และสามารถเชื่อมโยงความสัมพันธ์เชิงความหมายระหว่างคำในแบบต่าง ๆ ได้ครอบคลุม รวมทั้งเวิร์ดเน็ตถูกพัฒนาและมีการเพิ่มคำศัพท์มาอย่าง

ต่อเนื่องจนถึงปัจจุบัน ซึ่งประกอบไปด้วยข้อมูลของคำทั้งหมดประมาณ 90,000 คำ งานวิจัยนี้จึงเลือกใช้ เวิร์ดเน็ตเป็นฐานความรู้ในการเปรียบเทียบเพื่อคำนวณหาค่าความสัมพันธ์

2.1.4 การแทนความรู้ (Knowledge Representation)

วิทยานิพนธ์ฉบับนี้ใช้กราฟเชิงความคิด (Conceptual Graph) ในการแทนความรู้ กราฟเชิงความคิด (Chein et Marie-Laure Mugnier, M.,1992) คิดค้นโดย John F. Sowa เพื่อนำมาใช้เป็นรูปแบบการแทนความรู้ และนำมาประยุกต์ใช้กับการแสดงความหมายทางภาษาเพื่อตีความ (Knowledge Acquisition) ในการวิจัยทางด้านปัญญาประดิษฐ์ (Artificial Intelligence) จุดเด่นที่สำคัญของกราฟเชิงความคิดคือสามารถอธิบายภาษาธรรมชาติ (Natural Language) ออกมาในรูปแบบของสัญลักษณ์เชิงตรรกศาสตร์และเป็นตัวกลางในการสื่อความหมายระหว่างภาษาของคอมพิวเตอร์และภาษาของมนุษย์ ทำให้การสื่อแทนความหมายโดยใช้กราฟเชิงความคิดนั้นเข้าใจได้ง่ายและพัฒนาต่อไปในการหาเหตุผลเชิงตรรกะได้ ในการแทนความรู้โดยใช้กราฟเชิงความคิดประกอบด้วย

1. Concept Nodes แทนด้วย สัญลักษณ์รูปสี่เหลี่ยม ซึ่งจะแสดง Entities, Actions และ Attributes ซึ่งมี 2 Attributes

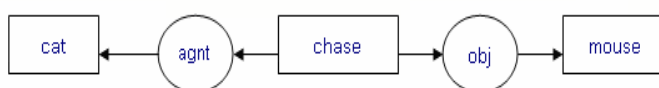
- Type จะเป็นตัวบอก class ของ element ที่แทนด้วย concept
- Referent จะเป็นตัวบอกความเฉพาะเจาะจงของ class ที่ถูกอ้างโดย node

2. Conceptual Relation Nodes แทนด้วย สัญลักษณ์รูปวงกลม จะแสดงความสัมพันธ์ระหว่าง Concept node โดย Attributes ของ Relation Nodes จะมี 2 Attributes คือ

- Valence จะเป็นตัวบอกจำนวนของ concept ที่อยู่ข้าง ๆ relation
- Type แสดงความเกี่ยวข้องของตัวมันเอง เช่น subject, attribute, object

ภาพที่ 2.3

แสดงตัวอย่าง Conceptual Graphs



ในภาพที่ 2.2 แสดง conceptual graph ของประโยค “Tom is chasing a brown mouse” ประกอบด้วย 3 concepts และ 2 relations ซึ่งสามารถเขียนในรูปแบบการแสดงผลที่ไม่เป็นกราฟิกได้ดังนี้

ภาพที่ 2.4

แสดงตัวอย่าง Conceptual Graphs

[cat: Tom]←-(Agnt)←-[chase]→(Ptnt)→[mouse]→(Attr)→[brown]

ทิศทางของลูกศรในรูปด้านบนแสดงถึงประธาน (Subject) และกรรม (Object) ของความเกี่ยวข้อง (Relation) concept [cat: Tom] เป็น concept ที่เฉพาะเจาะจงของ Type cat ในขณะที่ [chase] และ [mouse] เป็น generic concept ความเกี่ยวข้อง (Attr) บอก attribute ของ mouse ว่ามีสี brown

หลักการอีกส่วนหนึ่งที่สำคัญในการใช้กราฟเชิงความคิดในการแทนความรู้ คือการแสดงความเกี่ยวข้องระหว่างกลุ่มขององค์ประกอบแบบลำดับชั้น (Type hierarchy) หลักการดังกล่าวทำให้สามารถนำเอาคุณสมบัติการสืบทอด (Inheritance) มาประยุกต์ใช้ กล่าวคือ องค์ประกอบที่มีลักษณะเฉพาะเรียกว่าเป็น Subtype ขององค์ประกอบที่มีลักษณะทั่วไปซึ่งเรียกว่า Supertype สิ่งที่เป็น Subtype ขององค์ประกอบหนึ่ง ๆ จะสืบทอดคุณสมบัติขององค์ประกอบนั้น ๆ นอกจากนี้หลักการของกราฟเชิงความคิดได้ถูกนำไปใช้ในระบบงานค้นคืนสารสนเทศ (Montes-y-Gomez, Lopezs-Lopez, and Gelbukh, A., 2000) (Montes-y-Gomez, Lopezs-Lopez, Gelbukh, and Baeza-Yates, 2001) โดยอาศัยหลักการเปรียบเทียบกันระหว่างกราฟ (Graph Matching)

เนื่องจากกราฟเชิงความคิดสามารถแสดงส่วนประกอบ (concept) ในประโยค รวมทั้งแสดงความสัมพันธ์ของส่วนประกอบในประโยคนั้นด้วย ในงานวิจัยนี้จึงใช้กราฟเชิงความคิดในการแทนความหมายของประโยค

2.2 งานวิจัยอื่น ๆ ที่เกี่ยวข้อง

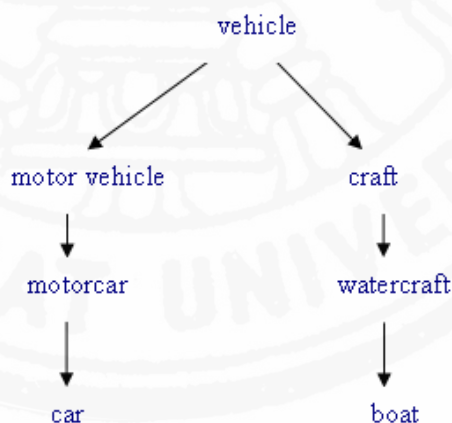
2.2.1 การเปรียบเทียบหาค่าความเหมือนเชิงความหมาย

Resnik (1995) และ Jiang and Conrath (1997) เสนอแนวคิดที่ว่าคำจะมีความหมายเหมือนกันนั้นสามารถพิจารณาได้จากความเหมือนกันของข้อความที่เนื้อหาของคำทั้งสอง ซึ่งวิธีการนี้จำเป็นต้องมีสร้างฐานความรู้คลังข้อมูลตัวอักษร (Corpus) และผลการคำนวณที่ได้ก็จะขึ้นอยู่กับขนาดและข้อมูลที่มีใน คลังข้อมูลตัวอักษร (Corpus) ด้วย

Leacock and Chodorow (1998) เสนอการคำนวณหาค่าความเหมือนกันทางความหมายของคำโดยคำนวณจากระยะห่างระหว่างคำ (Synset) ทั้งสองที่สั้นที่สุดโดยอาศัยโครงข่ายในเวิร์ดเน็ตโดยพิจารณาความสัมพันธ์แบบ “is-a”

ภาพที่ 2.5

แสดงระยะห่างระหว่าง “car” และ “boat” ในโครงข่ายเวิร์ดเน็ตเท่ากับ 6



2.2.2 การเปรียบเทียบค่าความสัมพันธ์เชิงความหมาย

Hist and St-On (1998) คำนวณค่าความสัมพันธ์เชิงความหมายโดยอาศัยโครงข่ายในเวิร์ดเน็ต จะพิจารณาจำนวนครั้งของการเปลี่ยนทิศทางของเส้นทาง มีสูตรในการคำนวณ คือ

$$Rel(c_1, c_2) = C - path\ length - k * d \quad (2.1)$$

โดยที่ d จำนวนครั้งของการเปลี่ยนทิศทางใน path, C และ k เป็นค่าคงที่ ซึ่งเส้นทางความสัมพันธ์ที่พิจารณานั้นประกอบด้วย Synonym, Hyponym, Hypernym และ Holonym โดยที่จะให้ความสำคัญในแต่ละรูปแบบความสัมพันธ์ที่เท่ากัน

Banerjee and Pederson (2003) เสนอวิธีการคำนวณหาความสัมพันธ์ในเชิงความหมายของคำซึ่งได้ประยุกต์ใช้หลักการของ Lesk (1986) ซึ่งจะนับจำนวนซึ่งสถิตในคำแปล (gloss) ของคำทั้งสองที่เหมือนกัน (overlap) โดยใช้ความคำแปลและโครงข่ายในเวิร์ดเน็ตเป็นฐานความรู้ แต่จะขยาย set ของคำแปลที่จะเปรียบเทียบโดยการนำเอาค่าผลการเปรียบเทียบคำที่สถิตอยู่เหมือนกันในคำแปลของคำที่เป็น Hyponym และ Hypernym ของคำที่ต้องการเปรียบเทียบทั้งสองมารวมคำนวณค่าสัมพันธ์ด้วย ดังนี้

$$\begin{aligned} Relatedness(A,B) = & score(gloss(A), gloss(B)) + score(hype(A), hype(B)) + \\ & score(hypo(A), hypo(B)) + score(hype(A), gloss(B)) + \\ & score(gloss(A), hype(B)) \end{aligned} \quad (2.2)$$

นอกจากนี้ยังมีการให้น้ำหนักตามความยาวของวลีที่ปรากฏด้วย เช่น ปรากฏคำว่า “bank” เหมือนกันจะได้คะแนนเป็น 1 ในขณะที่ปรากฏคำว่า “bank account” เหมือนกันคะแนนที่ได้จะเป็น 2^2 เท่ากับ 4 เป็นต้น

Strube and Ponzetto (2006) เสนอวิธีการ WikiRelate โดยใช้หลักการเดียวกันกับ Lesk (1986) โดยใช้คำแปลหรือคำนิยามของคำใน Wikipedia ซึ่งเป็นพจนานุกรมภาษาอังกฤษแบบออนไลน์ในการคำนวณหาความสัมพันธ์ในเชิงความหมาย โดยคำแปลที่นำมาคำนวณค่าความสัมพันธ์นั้นจะใช้ข้อความจากย่อหน้าแรกใน Wikipedia ในการเปรียบเทียบคำที่สถิตอยู่

เหมือนกันนั้นมีการให้น้ำหนัก ตามความยาวของวลีที่ปรากฏเช่นเดียวกับ Banerjee and Pederson (2003) โดยมีสูตรในการคำนวณหาความสัมพันธ์ดังนี้

$$Related\ gloss(t_1, t_2) = \tanh(\text{overlap}(t_1, t_2) / \text{length}(t_1) + \text{length}(t_2)) \quad (2.3)$$

ซึ่งเมื่อเปรียบเทียบผลที่ได้กับการคำนวณหาความสัมพันธ์กันโดยใช้นิยามของคำจากเวิร์ดเน็ต ปรากฏว่าให้ผลที่ดีกว่าการใช้คำนิยามจากเวิร์ดเน็ต

Zhang, Sun, Wang, and Bai, (2005) นำเสนอการวัดความสัมพันธ์ของประโยคกับหัวข้อเอกสารโดยการเปรียบเทียบหาความสัมพันธ์ในเชิงความหมายของประโยค ค่าความสัมพันธ์จะคำนวณจากผลรวมของค่าความสัมพันธ์ของคำในประโยค A กับประโยค B รวมกับผลรวมของค่าความสัมพันธ์ของคำในประโยค B กับประโยค Aหารด้วยจำนวนคำทั้งหมดในประโยคทั้งสอง โดยค่าความสัมพันธ์ของคำในประโยค A กับประโยค B คือระยะห่างระหว่างคำในประโยค A กับคำในประโยค B ที่สั้นที่สุด ซึ่งในการคำนวณหาระยะห่างระหว่างคำที่สั้นที่สุดนั้นจะพิจารณาความสัมพันธ์จากฐานข้อมูลในเวิร์ดเน็ตระดับ Synonym Hyponym Hypernym รวมทั้ง Antonym และ derivation ด้วย โดยจะให้น้ำในความสัมพันธ์ระดับ Antonym และ derivation ในระดับต่ำสุด ถึงแม้ว่าจะพิจารณาความสัมพันธ์เชิงความหมายในระดับประโยคแต่ก็ไม่ได้พิจารณาความสัมพันธ์กันของคำในประโยคนั้น ๆ แต่อย่างใด

Yang and Powers (2005) ได้เสนอวิธีการในการวัดค่าความสัมพันธ์ทางความหมายระหว่างคำจากระยะห่างโหนดโดยอาศัยโครงข่ายในเวิร์ดเน็ต โดยความสัมพันธ์ที่พิจารณาได้แก่ Synonym, Antonym, Hypernym, Hyponym, Antonym, Holonym และ Meronym คือนอกเหนือจากพิจารณาความสัมพันธ์ “is-a” แล้ว ยังพิจารณาความสัมพันธ์ “part-of”, อีกด้วย มีการให้น้ำหนักความสัมพันธ์ในแบบ Synonym และ Antonym มากกว่าความสัมพันธ์ในแบบอื่นๆ ซึ่งผลการทดลองกับชุดคู่ลำดับของคำมาตรฐาน 28 คู่ โดยการพิจารณาของมนุษย์ปรากฏว่าให้ผลที่ดีกว่าเมื่อเปรียบเทียบกับงานวิจัยที่ผ่านมา

ตารางที่ 2.1
เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ผู้เขียน	เสนอ	ข้อดี	ข้อเสีย
Hist and St-Ong (1998)	คำนวณค่า ความสัมพันธ์เชิง ความหมายโดยอาศัย โครงข่ายในเวิร์ดเน็ต จะ พิจารณาจำนวนครั้ง ของการเปลี่ยนทิศทาง ของเส้นทาง	ทำได้ง่าย	พิจารณาแค่ ความสัมพันธ์ในแบบ Synonym และ Hyponym/Hypernym เท่านั้น
Lesk M. (1986)	นับจำนวนซึ่งปรากฏใน คำแปล (gloss) ของคำ ทั้งสองที่เหมือนกัน (overlap)	ทำได้ง่าย	พิจารณาเฉพาะคำที่ เหมือนกันใน gloss เท่านั้น รวมทั้งมีข้อจำกัด ในเรื่องคำแปล (gloss) ที่ มีอยู่ใน dictionary อาจ ให้ข้อมูลที่ไม่เพียงพอ
Banerjee and Pederson (2003)	นับจำนวนซึ่งปรากฏใน คำแปล (gloss) ของคำ ทั้งสองที่เหมือนกัน (overlap) โดยมีการให้ น้ำหนักตามความยาว ของวลีที่ปรากฏด้วย เช่น ปรากฏคำว่า “bank account” เหมือนกันคะแนนที่ได้ จะเป็น 2 2	ไม่พิจารณาเฉพาะ gloss ของ คำนั้น ๆ เท่านั้น แต่ยัง พิจารณา gloss ของ คำที่เป็น Hyponym, Hypernym, Meronym Holonym and Troponym Synsets ของคำ กำหนดอีกด้วย	ไม่ได้พิจารณาคำอื่น ๆ ที่ อยู่ในประโยคหรือ ข้อความที่อยู่รอบ ๆ

ตารางที่ 2.1 (ต่อ)

ผู้เขียน	เสนอ	ข้อดี	ข้อเสีย
Strube and Ponzetto (2006)	ใช้ความหมายหรือคำนิยามของคำ ใน Wikipedia ซึ่งเป็นพจนานุกรมภาษาอังกฤษแบบออนไลน์ในการคำนวณหาความสัมพันธ์ในเชิงความหมาย ซึ่งอาศัยหลักการของ Lesk และ Banerjee, Pederson เช่นกัน โดยมีการ normalize คำที่คำนวณได้อีกที	เนื่องจาก Wikipedia เป็น on-line encyclopedia database ที่มีขนาดใหญ่ ทำให้สามารถคำนวณค่าความสัมพันธ์จาก คำแปล (gloss) หรือ full text page ใน Wikipedia ได้	เปรียบเทียบคำที่เหมือนกันเท่านั้น
Yang and Powers (2005)	เสนอการพิจารณาความสัมพันธ์แบบ Synonym, Antonym, Hypernym, Hyponym, Antonym, Holonym และ Meronym คือ นอกเหนือจากพิจารณาความสัมพันธ์ “is-a” แล้ว ยังพิจารณาความสัมพันธ์ “part-of”, อีกด้วย มีการให้น้ำหนักความสัมพันธ์	ครอบคลุมความสัมพันธ์ในเชิงความหมายมากขึ้น	เป็นการพิจารณาความสัมพันธ์ในระดับคำเท่านั้น

ตารางที่ 2.1 (ต่อ)

ผู้เขียน	เสนอ	ข้อดี	ข้อเสีย
Zhang, Sun, Wang and Bai (2005)	เสนอการหาประโยคจะ สัมพันธ์กับหัวข้อ เอกสารโดยการ เปรียบเทียบเชิง ความหมายของประโยค คำนวณค่า ความสัมพันธ์จาก ผลรวมของค่า ความสัมพันธ์ของคำใน ประโยค A กับประโยค B รวมกัน แล้วหารด้วย จำนวนคำทั้งหมดใน ประโยคทั้งสอง	เป็นการพิจารณา ความสัมพันธ์ในระดับ ประโยค	ในการเปรียบเทียบ ระหว่างคำภายใน ประโยคพิจารณาแค่ ความสัมพันธ์ในแบบ Synonym และ Hyponym/Hypernym เท่านั้น ไม่ได้ พิจารณาความสัมพันธ์ ระหว่างคำในประโยค

จากงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบหาความสัมพันธ์ในเชิงความหมายนั้นยังมีข้อจำกัด ซึ่งสรุปได้ดังนี้

1. ระดับของการเปรียบเทียบ งานวิจัยส่วนใหญ่ใช้การเปรียบเทียบในระดับคำ เพื่อเปรียบเทียบความเหมือนกันระหว่างคำ แต่เนื่องจากว่าคำหนึ่งคำอาจแทนความหมายได้หลายความหมาย อาจได้ผลลัพธ์ที่ไม่ถูกต้อง
2. ความสัมพันธ์ของคำ งานวิจัยโดยทั่วไปจะเปรียบเทียบเฉพาะคำที่มีความหมายเหมือนกัน (Synonym) ในความสัมพันธ์ “is-a” ซึ่งในการใช้งานจริงๆ แล้ว ยังมีความสัมพันธ์อื่นๆ มาใช้ในการเปรียบเทียบได้ด้วย เช่น ความสัมพันธ์ “part-of”, ความสัมพันธ์ “member-is-a-kind-of”, ความสัมพันธ์ “member-is-a-part-of” เป็นต้น

วิทยานิพนธ์ฉบับนี้จึงเสนอวิธีการวัดความสัมพันธ์เชิงความหมายของคำในภาษาอังกฤษ โดยพิจารณาจากความสัมพันธ์ 6 แบบ ได้แก่ Synonym, Hypernym, Hyponym,

Antonym, Holonym และ Meronym และการให้น้ำหนักความสัมพันธ์ในแต่ละแบบที่แตกต่างกัน ซึ่งในการพิจารณานั้นจะพิจารณาในระดับของประโยค โดยประโยคที่นำมาพิจารณาความเหมือนระหว่างคำสองคำนั้น จะถูกแทนด้วยกราฟเชิงความคิดก่อน แล้วนำมาคำนวณหาความเหมือนของกราฟ, ความเหมือนของประโยค และความเหมือนของคำ ตามลำดับ

