

## แบบเสนอโครงการพิเศษ (ปริญญานิพนธ์)

สาขาวิชาวิศวกรรมสารสนเทศและเครือข่าย  
ภาควิชาเทคโนโลยีสารสนเทศ  
คณะเทคโนโลยีและการจัดการอุตสาหกรรม

### 1. ข้อมูลขั้นต้นของโครงการ

#### 1.1 ชื่อโครงการ

(ภาษาไทย) แพลตฟอร์มวิศวกรรมความรู้เชิงสัมพันธ์การปรากฏร่วม  
(ภาษาอังกฤษ) Co-Occurrence Knowledge Engineering Platform

#### 1.2 ชื่อนักศึกษาผู้ทำโครงการ

ชื่อ-นามสกุล นายยงยุทธ ขวนขุนทด  
สาขาวิชา วิศวกรรมสารสนเทศและเครือข่าย  
ภาควิชา เทคโนโลยีสารสนเทศ  
ภาคเรียนที่ 1  
ปีการศึกษา 2568

#### 1.3 ชื่ออาจารย์ที่ปรึกษา

ชื่อ-นามสกุล รองศาสตราจารย์ ดร. อนิราช มิ่งขวัญ

### 2. รายละเอียดของโครงการ

#### 2.1 ความเป็นมาและความสำคัญของปัญหา

ในยุคดิจิทัลปัจจุบัน ข้อมูลและความรู้ที่เกิดขึ้นในแต่ละวันมีปริมาณที่เพิ่มขึ้นอย่างรวดเร็ว โดยเฉพาะอย่างยิ่งเอกสารทางวิชาการ หนังสือ และงานวิจัยต่าง ๆ ที่มีการเผยแพร่ในรูปแบบดิจิทัล เช่น ไฟล์ PDF ซึ่งเป็นแหล่งความรู้ที่มีคุณค่าสูง อย่างไรก็ตาม การจัดการ การวิเคราะห์ และการค้นหาความเชื่อมโยงระหว่างความรู้จากเอกสารเหล่านี้ยังคงเป็นปัญหาที่ท้าทาย

ปัญหาหลักที่พบในปัจจุบันคือ การที่ผู้ใช้งานไม่สามารถมองเห็นภาพรวมของความสัมพันธ์และความเชื่อมโยงระหว่างแนวคิดต่าง ๆ ที่ปรากฏในเอกสารหลายฉบับได้อย่างชัดเจน การอ่านและทำความเข้าใจเอกสารแต่ละฉบับแยกกันทำให้เกิดการสูญเสียโอกาสในการค้นพบความรู้ใหม่ที่อาจเกิดขึ้นจากการเชื่อมโยงข้อมูลจากแหล่งที่ต่างต่างกัน

นอกจากนี้ การวิเคราะห์ความถี่ของการใช้คำและการปรากฏร่วมของความรู้ (Co-occurrence) ในเอกสารยังเป็นกระบวนการที่ซับซ้อนและใช้เวลานาน หากต้องทำด้วยมือหรือเครื่องมือพื้นฐาน ทำให้การสกัดความรู้และการสร้างความเข้าใจเชิงลึกจากเอกสารเป็นไปได้ยาก

ด้วยเหตุนี้ จึงจำเป็นต้องมีระบบที่สามารถแปลงเอกสารจากแหล่งต่าง ๆ ให้กลายเป็นกราฟเครือข่าย (Network Graph) ที่แสดงความสัมพันธ์และความเชื่อมโยงระหว่างแนวคิดได้อย่างชัดเจน รวมถึงสามารถสกัดส่วนของกราฟเพื่อนำไปผสมผสานกับข้อมูลจากแหล่งอื่น ๆ เพื่อสร้างความรู้ใหม่และค้นพบความเป็นไปได้ที่ไม่เคยมีมาก่อน ซึ่งจะช่วยเพิ่มประสิทธิภาพในการจัดการความรู้และส่งเสริมการเกิดนวัตกรรมใหม่ ๆ ในอนาคต

## 2.2 วัตถุประสงค์ของโครงการ

- 2.2.1 เพื่อพัฒนาแพลตฟอร์มวิศวกรรมความรู้เชิงสัมพันธ์การปรากฏร่วมที่สามารถแปลงเอกสารให้กลายเป็นกราฟเครือข่ายความรู้ได้อย่างมีประสิทธิภาพ
- 2.2.2 เพื่อพัฒนาระบบวิเคราะห์การปรากฏร่วมของคำและแนวคิด (Co-occurrence Analysis) ที่สามารถระบุความถี่และความสัมพันธ์ระหว่างคำศัพท์ในเอกสารได้อย่างแม่นยำ
- 2.2.3 เพื่อสร้างส่วนติดต่อผู้ใช้ (User Interface) ที่ช่วยให้ผู้ใช้สามารถแสดงผลและโต้ตอบกับกราฟเครือข่ายความรู้ได้อย่างง่ายดายและเข้าใจง่าย
- 2.2.4 เพื่อพัฒนาฟิเจอร์การจัดการส่วนของกราฟ (Graph Management) เพื่อนำไปใช้ในการสร้างกราฟเครือข่ายใหม่หรือผสมผสานกับข้อมูลจากแหล่งอื่น
- 2.2.5 เพื่อพัฒนาระบบการผสมผสานความรู้จากหลายแหล่งข้อมูลเพื่อค้นหาความเชื่อมโยงและสร้างความรู้ใหม่ที่มีความเชื่อมโยงกัน
- 2.2.6 เพื่อพัฒนาระบบบูรณาการกับ Large Language Models (LLM) ที่สามารถใช้ข้อมูลจากกราฟเครือข่ายในการปรับปรุงความแม่นยำและประสิทธิภาพของการค้นหาและสกัดความรู้

## 2.3 ขอบเขตของการทำโครงการพิเศษ (Scope of Special Project)

- 2.3.1 การพัฒนาระบบประมวลผลและวิเคราะห์เอกสาร เช่น PDF ที่สามารถดึงข้อความ วิเคราะห์โครงสร้าง และแยกแยะเนื้อหาสำคัญจากเอกสารได้อย่างมีประสิทธิภาพ รวมถึงการจัดการกับรูปแบบการจัดวางข้อความและภาษาที่หลากหลาย
- 2.3.2 การพัฒนาระบบวิเคราะห์การปรากฏร่วม (Co-occurrence Analysis) ที่สามารถ
  - 2.3.2.1 วิเคราะห์ความถี่ของคำและวลีในเอกสาร
  - 2.3.2.2 ระบุความสัมพันธ์และการปรากฏร่วมของแนวคิดต่าง ๆ
  - 2.3.2.3 คำนวณค่าความแข็งแกร่งของความเชื่อมโยงระหว่างคำหรือแนวคิด
  - 2.3.2.4 สร้างเมทริกซ์ความสัมพันธ์สำหรับการสร้างกราฟเครือข่าย
- 2.3.3 การพัฒนาระบบสร้างและจัดการกราฟเครือข่ายความรู้ (Knowledge Network Graph) ที่สามารถ
  - 2.3.3.1 แปลงข้อมูลการวิเคราะห์ให้เป็นโครงสร้างกราฟ
  - 2.3.3.2 จัดกลุ่มและจัดระเบียบโหนดและขอบเชื่อมตามความสัมพันธ์
  - 2.3.3.3 คำนวณคุณสมบัติของกราฟ เช่น ความหนาแน่น ความเป็นศูนย์กลาง
  - 2.3.3.4 สนับสนุนการแสดงผลแบบ Interactive และ Dynamic
- 2.3.4 การพัฒนาส่วนติดต่อผู้ใช้ (User Interface) ที่มีคุณสมบัติ
  - 2.3.4.1 อัปโหลดและจัดการไฟล์ PDF
  - 2.3.4.2 แสดงผลกราฟเครือข่ายแบบโต้ตอบได้
  - 2.3.4.3 เครื่องมือการค้นหาและกรองข้อมูล

- 2.3.4.4 ส่งออกผลลัพธ์ในรูปแบบต่าง ๆ (รูปภาพ, JSON, CSV)
- 2.3.5 การพัฒนาฟีเจอร์การสกัดและจัดการส่วนของกราฟ (Graph Management) ที่สามารถ
  - 2.3.5.1 เลือกและสกัดส่วนที่สนใจจากกราฟใหญ่
  - 2.3.5.2 บันทึกและจัดเก็บส่วนกราฟที่สกัดไว้
  - 2.3.5.3 ผสมผสานกราฟจากหลายแหล่งข้อมูล
  - 2.3.5.4 สร้างกราฟใหม่จากการรวมข้อมูลหลายเอกสาร
- 2.3.6 การพัฒนาระบบฐานข้อมูลสำหรับจัดเก็บข้อมูล
  - 2.3.6.1 ข้อมูลเอกสารและเมตาดาต้า
  - 2.3.6.2 ผลการวิเคราะห์และกราฟเครือข่าย
  - 2.3.6.3 ประวัติการทำงานและการแก้ไข
  - 2.3.6.4 การตั้งค่าและ Preferences ของผู้ใช้
- 2.3.7 การพัฒนาระบบบูรณาการกับ Large Language Models (LLM) ที่สามารถ
  - 2.3.7.1 แปลงข้อมูล Network Graph ให้เป็นรูปแบบที่ LLM สามารถเข้าใจและประมวลผลได้
  - 2.3.7.2 สร้างระบบ Query Interface ที่ช่วยให้ผู้ใช้สามารถสอบถามข้อมูลจากกราฟผ่าน LLM ได้
  - 2.3.7.3 พัฒนา Context-aware Search ที่ใช้ความรู้จากกราฟเครือข่ายในการปรับปรุงความแม่นยำของการค้นหา
  - 2.3.7.4 สร้างระบบ Knowledge Discovery ที่ใช้ LLM ในการวิเคราะห์และสกัดความรู้ใหม่จากความสัมพันธ์ในกราฟ
- 2.4 รายละเอียดของทฤษฎีที่ใช้ในการจัดทำปริณิญาพนธ์
  - 2.4.1 สมมติฐาน หรือ ข้อตกลงเบื้องต้นในการจัดทำโครงการพิเศษ (Assumption of the Study)
    - 2.4.1.1 เอกสารที่นำเข้าสู่ระบบจะอยู่ในรูปแบบ PDF เป็นต้น ฯลฯ ที่มีข้อความที่สามารถสกัดได้ (Text-extractable) และมีคุณภาพเพียงพอ สำหรับการ ประมวล ผล ด้วย เทคนิค Optical Character Recognition (OCR) ในกรณีที่เป็น
    - 2.4.1.2 เอกสารที่ใช้ในการวิเคราะห์จะเป็นเอกสารทางวิชาการ งานวิจัย หรือเอกสารที่มีโครงสร้างและเนื้อหาที่ชัดเจน โดยมีการใช้คำศัพท์และแนวคิดที่สามารถระบุและวิเคราะห์ความสัมพันธ์ได้
    - 2.4.1.3 ระบบจะทำงานภายใต้สมมติฐานที่ว่าปรากฏการณ์ของคำหรือแนวคิดในระยะทางที่ใกล้กันภายในเอกสารแสดงถึงความสัมพันธ์หรือความเชื่อมโยงระหว่างแนวคิดเหล่านั้น
    - 2.4.1.4 ผู้ใช้งานระบบจะมีความรู้พื้นฐานในการตีความและวิเคราะห์กราฟเครือข่าย รวมถึงสามารถระบุความสัมพันธ์และความหมายของความสัมพันธ์ที่แสดงในกราฟได้
    - 2.4.1.5 ระบบจะมีประสิทธิภาพในการประมวลผลเอกสารที่มีขนาดปานกลางถึงใหญ่ โดยสมมติว่าทรัพยากรคอมพิวเตอร์ที่ใช้งานมีความสามารถเพียงพอสำหรับการประมวลผลและการแสดงผลกราฟเครือข่ายแบบโต้ตอบได้
    - 2.4.1.6 การบูรณาการกับ Large Language Models (LLM) จะช่วยปรับปรุงความแม่นยำในการระบุและจัดกลุ่มแนวคิด โดยสมมติว่า LLM จะสามารถเข้าใจบริบทและความหมายของข้อความในเอกสารได้อย่างถูกต้อง
    - 2.4.1.7 ผลลัพธ์ที่ได้จากระบบจะมีความเชื่อถือได้และสามารถนำไปใช้ในการตัดสินใจหรือการวิจัยเพิ่มเติมได้ โดยระบบจะมีกลไกในการตรวจสอบและปรับปรุงความแม่นยำของผลการวิเคราะห์

- 2.4.1.8 การผสมผสานข้อมูลจากหลายแหล่งจะช่วยสร้างความรู้ใหม่ที่มีคุณค่า โดยสมมติว่าข้อมูลจากแหล่งต่าง ๆ จะมีความเข้ากันได้และสามารถรวมเข้าด้วยกันได้อย่างมีความหมาย
- 2.4.2 คำจำกัดความ (Key Word)
- 2.4.2.1 การปรากฏร่วม (Co-occurrence) หมายถึง การที่คำหรือแนวคิดสองคำหรือมากกว่าปรากฏในตำแหน่งที่ใกล้เคียงกันภายในเอกสารหรือข้อความ ซึ่งแสดงถึงความสัมพันธ์หรือความเชื่อมโยงระหว่างแนวคิดเหล่านั้น
- 2.4.2.2 วิศวกรรมความรู้ (Knowledge Engineering) หมายถึง กระบวนการในการสกัด จัดระเบียบ จัดเก็บ และจัดการความรู้จากแหล่งข้อมูลต่าง ๆ เพื่อให้สามารถนำไปใช้งานได้อย่างมีประสิทธิภาพ
- 2.4.2.3 แพลตฟอร์ม (Platform) หมายถึง ระบบซอฟต์แวร์ที่ให้บริการและเครื่องมือครบครันสำหรับการดำเนินงานเฉพาะด้าน ในที่นี้คือการวิเคราะห์และจัดการความรู้เชิงสัมพันธ์
- 2.4.2.4 กราฟเครือข่าย (Network Graph) หมายถึง โครงสร้างข้อมูลที่ประกอบด้วยโหนด (Nodes) และขอบเชื่อม (Edges) ที่แสดงความสัมพันธ์ระหว่างแนวคิดหรือเอนทิตีต่าง ๆ ในรูปแบบที่เข้าใจได้ง่าย
- 2.4.2.5 การวิเคราะห์การปรากฏร่วม (Co-occurrence Analysis) หมายถึง เทคนิคการวิเคราะห์ข้อมูลที่ใช้ในการระบุและวัดความถี่ของการปรากฏร่วมของคำหรือแนวคิดในข้อความ เพื่อค้นหาความสัมพันธ์และรูปแบบต่าง ๆ
- 2.4.2.6 โหนด (Node) หมายถึง จุดหรือองค์ประกอบพื้นฐานในกราฟเครือข่ายที่แทนคำ แนวคิด หรือเอนทิตีต่าง ๆ ที่ได้จากการวิเคราะห์เอกสาร
- 2.4.2.7 ขอบเชื่อม (Edge) หมายถึง เส้นเชื่อมระหว่างโหนดในกราฟเครือข่ายที่แสดงความสัมพันธ์หรือความเชื่อมโยงระหว่างแนวคิดหรือเอนทิตีต่าง ๆ รวมถึงค่าน้ำหนักที่บ่งบอกถึงความแข็งแกร่งของความสัมพันธ์
- 2.4.2.8 การสกัดข้อความ (Text Extraction) หมายถึง กระบวนการในการดึงข้อความจากเอกสารดิจิทัล เช่น ไฟล์ PDF โดยใช้เทคนิคการประมวลผลเอกสารหรือ OCR (Optical Character Recognition)
- 2.4.2.9 เมทริกซ์ความสัมพันธ์ (Relationship Matrix) หมายถึง ตารางสองมิติที่แสดงค่าความแข็งแกร่งของความสัมพันธ์ระหว่างคำหรือแนวคิดต่าง ๆ ที่ได้จากการวิเคราะห์การปรากฏร่วม
- 2.4.2.10 Large Language Models (LLM) หมายถึง โมเดลปัญญาประดิษฐ์ที่ได้รับการฝึกฝนด้วยข้อมูลข้อความขนาดใหญ่ เพื่อให้สามารถเข้าใจและประมวลผลภาษาธรรมชาติได้อย่างมีประสิทธิภาพ
- 2.4.2.11 การแสดงผลแบบโต้ตอบ (Interactive Visualization) หมายถึง การแสดงผลข้อมูลในรูปแบบที่ผู้ใช้สามารถโต้ตอบและปรับเปลี่ยนมุมมองหรือรายละเอียดได้ตามต้องการ
- 2.4.2.12 การจัดการส่วนของกราฟ (Graph Management) หมายถึง ระบบที่ช่วยในการเลือก สกัด บันทึก และจัดการส่วนต่าง ๆ ของกราฟเครือข่ายเพื่อนำไปใช้งานหรือวิเคราะห์เพิ่มเติม
- 2.4.2.13 การค้นพบความรู้ (Knowledge Discovery) หมายถึง กระบวนการในการค้นหาและระบุรูปแบบความสัมพันธ์ หรือความรู้ใหม่ที่ซ่อนอยู่ในข้อมูลขนาดใหญ่ผ่านเทคนิคการวิเคราะห์ข้อมูลต่าง ๆ
- 2.4.3 รายงานการค้นคว้า การศึกษา หรือการวิจัยที่เกี่ยวข้อง
- 2.4.3.1 Centroid Terms as Text Representatives
- งาน วิจัย ของ Mario M. Kubek, Herwig Unger (2016) เรื่อง "Centroid Terms as Text Representatives" ได้ศึกษาเทคนิคการสร้างกราฟความรู้จากข้อความ โดยเฉพาะการใช้เทคนิค Co-occurrence Analysis ในการระบุความสัมพันธ์ระหว่างเอนทิตีต่าง ๆ การศึกษานี้ได้ชี้ให้เห็นว่า อัลกอริทึมสำหรับการจัดกลุ่มและการจำแนกข้อความตามหัวข้อนั้นอาศัยข้อมูลเกี่ยวกับระยะทาง

และความคล้ายคลึงเชิงความหมาย วิธีการมาตรฐานที่อิงตามแบบจำลอง bag-of-words ในการกำหนดข้อมูลนี้จะให้เพียงการประมาณแบบคร่าว ๆ เกี่ยวกับความเกี่ยวข้องของข้อความ นอกจากนี้วิธีการเหล่านี้ยังไม่สามารถค้นหาคำศัพท์ที่เป็นนามธรรมหรือคำที่สามารถอธิบายเนื้อหาของข้อความได้อย่างครอบคลุม งานวิจัยนี้จึงได้นำเสนอวิธีการใหม่ในการกำหนดคำศูนย์กลาง (Centroid Terms) ในข้อความและการประเมินความคล้ายคลึงโดยใช้คำที่เป็นตัวแทนเหล่านั้น ซึ่งแสดงให้เห็นว่าการวิเคราะห์การปรากฏร่วมสามารถช่วยในการค้นพบรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ และสามารถนำไปประยุกต์ใช้ในการพัฒนาระบบวิศวกรรมความรู้เชิงสัมพันธ์ได้

#### 2.4.3.2 Spreading activation: a fast calculation method for text centroids

งานวิจัยของ Mario M. Kubeck, Thomas Böhme, Herwig Unger (2017) เรื่อง "Spreading activation: a fast calculation method for text centroids" ได้นำเสนอเทคนิคการคำนวณ Centroid Terms แบบใหม่ที่มีประสิทธิภาพสูง โดยใช้ Spreading Activation Algorithm ซึ่งเป็นเทคนิคที่ทำงานบนพื้นฐานของกราฟและหลักการงานแบบเฉพาะที่ (Local Working Principle) การศึกษานี้ชี้ให้เห็นว่า Centroids เป็นเครื่องมือที่สะดวกในการแสดงคำค้นหาและข้อความทั้งหมดด้วยคำศัพท์เชิงบรรยายเพียงคำเดียว ซึ่งสามารถนำไปใช้ในการกำหนดความคล้ายคลึงของเนื้อหาข้อความและการจัดกลุ่มเอกสารแบบลำดับชั้น อย่างไรก็ตาม การคำนวณตามคำจำกัดความแบบดั้งเดิมอาจใช้เวลามากและเป็นอุปสรรคต่อการนำไปใช้งานจริง ดังนั้น การพัฒนาอัลกอริทึมแบบใหม่ที่อิงตามกราฟ Co-occurrence จึงมีความสำคัญต่อการเพิ่มประสิทธิภาพการประมวลผล ซึ่งสอดคล้องกับแนวทางที่จะใช้ในโครงงานนี้สำหรับการวิเคราะห์การปรากฏร่วมและการสร้างกราฟเครือข่ายความรู้ที่มีประสิทธิภาพในการประมวลผลเอกสารขนาดใหญ่

#### 2.4.3.3 Enhancing Retrieval-Augmented Generation Systems by Text-Representing Centroid

การศึกษานี้ของ Yanakorn Ruamsuk, Anirach Mingkhawn, Herwig Unger (2025) เรื่อง "Enhancing Retrieval-Augmented Generation Systems by Text-Representing Centroid" ได้นำเสนอแนวทางใหม่ในการปรับปรุงระบบ Retrieval-Augmented Generation (RAG) โดยการบูรณาการเทคนิค Text-Representing Centroid (TRC) เพื่อแก้ไขข้อจำกัดของฐานข้อมูลเวกเตอร์แบบดั้งเดิม วิธีการนี้สามารถรักษาความสัมพันธ์เชิงโครงสร้างและปรับตัวตามความซับซ้อนของเนื้อหา ส่งผลให้การค้นคืนข้อมูลมีประสิทธิภาพและความแม่นยำที่สูงขึ้น การมีส่วนร่วมสนับสนุนที่สำคัญได้แก่การสร้างกราฟขั้นสูง อัลกอริทึมการให้คะแนนความเกี่ยวข้อง และการตรวจสอบความถูกต้องอย่างครอบคลุม พร้อมการอภิปรายเกี่ยวกับการประยุกต์ใช้ที่เป็นไปได้และการวิจัยในอนาคต หลักฐานเชิงประจักษ์แสดงให้เห็นว่าเทคนิค TRC สามารถบรรลุอัตราความสำเร็จ 75 เปอร์เซ็นต์จากคำถาม 100 ข้อ ซึ่งมีประสิทธิภาพเหนือกว่าวิธีการเวกเตอร์แบบดั้งเดิม การศึกษานี้มีความเกี่ยวข้องโดยตรงกับโครงงานที่เสนอ เนื่องจากแสดงให้เห็นถึงความเป็นไปได้ในการใช้เทคนิค Co-occurrence Analysis และ Centroid-based Methods ในการพัฒนาระบบวิศวกรรมความรู้ที่มีประสิทธิภาพสูง โดยเฉพาะในด้านการประมวลผลเอกสารและการสร้างกราฟเครือข่ายความรู้ที่สามารถตอบคำถามและค้นหาข้อมูลได้อย่างแม่นยำ ซึ่งสอดคล้องกับวัตถุประสงค์ในการพัฒนาแพลตฟอร์มที่บูรณาการกับ Large Language Models เพื่อเพิ่มประสิทธิภาพในการสกัดและจัดการความรู้จากเอกสารหลายแหล่ง

#### 2.4.4 เนื้อหา เหตุผล และทฤษฎีที่สำคัญ

โครงงาน Co-Occurrence Knowledge Engineering Platform นี้มีพื้นฐานทางทฤษฎีที่แข็งแกร่งและเหตุผลที่ชัดเจนในการพัฒนา โดยอาศัยหลักการทางวิศวกรรมความรู้และเทคนิคการวิเคราะห์ข้อมูลขั้นสูงหลายแนวทาง

## เหตุผลในการพัฒนาโครงการ

ในยุคสารสนเทศปัจจุบัน ปริมาณข้อมูลและเอกสารดิจิทัลเพิ่มขึ้นอย่างรวดเร็ว แต่การจัดการและการค้นหาความเชื่อมโยงระหว่างความรู้จากแหล่งต่าง ๆ ยังคงเป็นความท้าทายสำคัญ ผู้ใช้งานส่วนใหญ่ไม่สามารถมองเห็นภาพรวมของความสัมพันธ์ระหว่างแนวคิดที่ปรากฏในเอกสารหลายฉบับได้อย่างชัดเจน การวิเคราะห์การปรากฏร่วม (Co-occurrence Analysis) แบบดั้งเดิมใช้เวลามากและซับซ้อน ดังนั้น จึงจำเป็นต้องมีระบบที่สามารถแปลงเอกสารให้เป็นกราฟเครือข่ายความรู้ที่เข้าใจได้ง่าย และสามารถผสมผสานข้อมูลจากหลายแหล่งเพื่อสร้างความรู้ใหม่

## ทฤษฎีพื้นฐานที่ใช้ในการพัฒนา

### 1. ทฤษฎีการปรากฏร่วม (Co-occurrence Theory)

หลักการพื้นฐานของการวิเคราะห์การปรากฏร่วมอิงตามสมมติฐานที่ว่า คำหรือแนวคิดที่ปรากฏใกล้กันในข้อความมักจะมีความสัมพันธ์หรือความเชื่อมโยงกัน ทฤษฎีนี้ได้รับการสนับสนุนจากงานวิจัยของ Mario M. Kubek et al. (2016, 2017) ที่แสดงให้เห็นว่าการวิเคราะห์ Centroid Terms และ Spreading Activation Algorithm สามารถช่วยในการระบุความสัมพันธ์เชิงความหมายได้อย่างมีประสิทธิภาพ

### 2. ทฤษฎีกราฟและเครือข่าย (Graph Theory and Network Theory)

การแสดงความรู้ในรูปแบบกราฟเครือข่ายอิงตามทฤษฎีกราฟ ซึ่งประกอบด้วยโหนด (Nodes) แทนแนวคิดหรือเอนทิตี และขอบเชื่อม (Edges) แทนความสัมพันธ์ ทฤษฎีนี้ช่วยให้สามารถคำนวณคุณสมบัติต่าง ๆ ของเครือข่าย เช่น ความหนาแน่น (Density) ความเป็นศูนย์กลาง (Centrality) และการจัดกลุ่ม (Clustering) ซึ่งมีความสำคัญต่อการวิเคราะห์และการค้นพบความรู้

### 3. ทฤษฎีการประมวลผลภาษาธรรมชาติ (Natural Language Processing Theory)

การสกัดข้อความ และการวิเคราะห์เนื้อหาจากเอกสาร PDF อาศัยหลักการของ NLP รวมถึงเทคนิค Tokenization, Part-of-Speech Tagging, Named Entity Recognition และ Semantic Analysis เพื่อให้สามารถระบุและแยกแยะแนวคิดสำคัญได้อย่างแม่นยำ

### 4. ทฤษฎีการเรียนรู้ของเครื่อง (Machine Learning & Deep Learning Theory)

การบูรณาการกับ Large Language Models (LLM) อาศัยหลักการของ Deep Learning และ Transformer Architecture เพื่อปรับปรุงความแม่นยำในการระบุความสัมพันธ์และการทำ Semantic Reasoning งานวิจัยของ Yanakorn Ruamsuk et al. (2025) แสดงให้เห็นว่าการใช้ Text-Representing Centroid ร่วมกับ LLM สามารถบรรลุอัตราความสำเร็จ 75

### 5. ทฤษฎีการจัดการฐานข้อมูล (Database Management Theory)

การจัดเก็บและจัดการข้อมูลกราฟอาศัยหลักการของ Graph Database และ NoSQL Database เพื่อรองรับการประมวลผลข้อมูลเชิงสัมพันธ์ที่ซับซ้อนและการ Query แบบ Real-time

**แนวทางการประยุกต์ใช้** โครงการนี้นำทฤษฎีต่าง ๆ มาประยุกต์ใช้ในการสร้างระบบที่ครอบคลุมตั้งแต่การประมวลผลเอกสาร การวิเคราะห์การปรากฏร่วม การสร้างกราฟเครือข่าย การแสดงผลแบบโต้ตอบ จนถึงการบูรณาการกับ AI เพื่อสร้างแพลตฟอร์มที่สามารถจัดการความรู้ได้อย่างมีประสิทธิภาพ และสร้างคุณค่าใหม่จากการเชื่อมโยงข้อมูลจากแหล่งต่าง ๆ ซึ่งสอดคล้องกับแนวโน้มการวิจัยในปัจจุบันที่เน้นการใช้เทคโนโลยี AI และ Knowledge Engineering ในการแก้ปัญหการจัดการความรู้ขนาดใหญ่

## 2.1. รายละเอียดทฤษฎีที่ใช้ในการจัดทำปริญญานิพนธ์

### 2.1.1 ทฤษฎีที่เกี่ยวกับ Web Application

1. Client-Server Architectures เป็นรูปแบบของการออกแบบระบบคอมพิวเตอร์ที่แบ่งหน้าที่และภาระงานระหว่างคอมพิวเตอร์ที่เรียกว่าไคลเอนท์ (Clients) และคอมพิวเตอร์ที่เรียกว่าเซิร์ฟเวอร์ (Servers) เพื่อให้การทำงานและการประมวลผลเกิดขึ้นอย่างมีประสิทธิภาพและมีความสมดุล สิ่งที่ทำให้แบบไคลเอนท์เซิร์ฟเวอร์มีประโยชน์หลัก 4 ประการได้แก่:

1.1 สามารถปรับเพิ่มลดได้ (Scalable)

1.2 สนับสนุนความหลากหลายรูปแบบ

1.3 การแยกแยะและการปรับปรุงง่าย

1.4 ความเสถียรและความยืดหยุ่นในการเปลี่ยนแปลง

2. HTTP (Hypertext Transfer Protocol) คือโพรโตคอลการสื่อสารระหว่างคอมพิวเตอร์ที่ใช้กันบนเว็บไซต์และบนอินเทอร์เน็ตเพื่อแลกเปลี่ยนข้อมูลและแสดงผลในรูปแบบของเอกสาร...

3. REST (Representational State Transfer) เป็นแนวคิดทางสถาปัตยกรรมในการสื่อสารระหว่างระบบคอมพิวเตอร์บนเว็บ...

4. การใช้ HTTP Verbs ในการกระทำ: REST ใช้เมธอด (HTTP Verbs) เพื่อให้กำหนดการกระทำต่าง ๆ กับทรัพยากร เช่น

4.1 GET: ดึงข้อมูลทรัพยากรหรือคอลเลกชันของทรัพยากร

4.2 POST: สร้างทรัพยากรใหม่

4.3 PUT: อัปเดตข้อมูลทรัพยากรเฉพาะ

4.4 DELETE: ลบทรัพยากรที่ระบุ

### 2.1.2 รายงานการค้นคว้า การศึกษา หรือการวิจัยที่เกี่ยวข้อง

\*\*\*\* สอง - สาม งาน (กำลังดูและเลือกรายงานที่เหมาะสมกับงานมากที่สุด)

## 2.2. วิธีการดำเนินงานจัดทำโครงการพิเศษ

ภาคการศึกษาที่ 1/2566

ภาคการศึกษาที่ 2/2566

## 2.3. แผนกิจกรรมและตารางเวลาในการจัดทำ

### 2.3.1 แผนกิจกรรมหลักและระยะเวลา

ตารางที่ 1: แผนการดำเนินงานภาคการศึกษาที่ 1

ขั้นตอนการดำเนินงาน	กรกฎาคม				สิงหาคม				กันยายน				ตุลาคม			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1																
2																
3																
4																
5																

ภาคการศึกษาที่ 1

ตารางที่ 2: แผนการดำเนินงานภาคการศึกษาที่ 2

ขั้นตอนการดำเนินงาน	ธันวาคม				มกราคม				กุมภาพันธ์				มีนาคม			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1																
2																
3																
4																
5																
6																
7																

ภาคการศึกษาที่ 2

## 2.4. ทรัพยากรที่ต้องใช้ในการจัดทำโครงการพิเศษ

### 2.4.1 เครื่องมือในการจัดทำโครงการพิเศษ

Software

2.7.1.1.1 MySQL

2.7.1.1.2 NLP



#### 2.7.1.1.3 Python

#### 2.7.1.1.4 Power BI

### Hardware

#### 2.7.1.2.1 คอมพิวเตอร์/โน้ตบุ๊ก

#### 2.7.1.2.2 โทรศัพท์

### 2.4.2 งบประมาณที่ใช้ในการจัดทำ

- ค่าจัดทำปฏิญญาพันธ 1,000 บาท
- ค่าใช้จ่ายอื่น ๆ 500 บาท
- รวมเป็นเงิน 1,500 บาท

### 2.5. ผลที่คาดว่าจะได้รับ (ปรับตามวัตถุประสงค์)

#### 2.8.1 ผู้ใช้สามารถหาเมนูอาหารที่ตรงตามความต้องการได้

#### 2.8.2 Chatbot สามารถที่จะตอบโต้และแนะนำเมนูอาหารได้

#### 2.8.3 ผู้ใช้ได้รับคำแนะนำอย่างเหมาะสมในด้านการรับประทานเมนูอาหาร

### 2.6. เอกสารอ้างอิง

\*\* ใช้ Zotero เลือกเป็น APA7th

### 2.7. ภาคผนวก

\*\* ใช้ figma หรือ wireframe

ลงชื่อ.....ผู้เสนอโครงการ  
(นายพรเทพ )

ลงชื่อ.....ผู้เสนอโครงการ  
(นางสาวลลนา สุขรักษ์)

วันที่ยื่นเสนอโครงการ...../...../.....

ความเห็นอาจารย์ที่ปรึกษาโครงการ

.....

ลงชื่อ.....อาจารย์ที่ปรึกษา  
(.....)  
วันที่...../...../.....

สาขาวิชา / ภาควิชาที่ได้รับแบบเสนอโครงการวันที่ .....

ผลการพิจารณา

.....

ลงชื่อ.....ประธาน  
(.....)  
วันที่...../...../.....

ลงชื่อ.....กรรมการ  
(.....)  
วันที่...../...../.....

ลงชื่อ.....กรรมการ  
(.....)  
วันที่...../...../.....

ลงชื่อ.....กรรมการ  
(.....)  
วันที่...../...../.....