การวิเคราะห์วัตถุประสงค์เชิงวิจัยและการขยายขอบเขตโครงการ

Analysis of Research Purpose and Project Scope Expansion

หัวข้อวิทยานิพนธ์	การสร้างคำอธิบายรูปภาพและกราฟในเอกสารวิชาการโดยใช้บริบทจากเนื้อหาเอกสาร
ผู้เขียน	ผู้นำเสนอโครงการวิจัย
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. หัชทัย ชาญเลขา
วัตถุประสงค์เอกสาร	บันทึกแนวคิด ทิศทางการวิจัย และความจำเป็นทางวิทยาศาสตร์ในการขยายขอบเขตระบบสืบค้น
วันที่บันทึก	June 21, 2026

บทสรุปการวิเคราะห์เชิงยุทธศาสตร์

จากประเด็นคำถามเชิงโครงสร้างและการจัดทำคลังเอกสารวิชาการที่สามารถประมวลผลรูปภาพประเภทอื่น ๆ นอกเหนือจากชาร์ตหรือกราฟทั่วไป (เช่น ภาพถ่ายสัตว์ ป่าไม้ เซลล์ทดลอง) และมีความสามารถในการสืบค้นข้อมูลเชิงโต้ตอบผ่านภาพประกอบได้นั้น ทางคณะผู้วิจัยได้ทำการจำแนกเจตจำนงเชิงยุทธศาสตร์การวิจัยออกเป็น 4 มิติมูลฐาน ซึ่งจะส่งผลโดยตรงต่อการเขียนข้อเสนอโครงการวิจัย (Thesis Proposal) และการออกแบบสถาปัตยกรรมทางเทคนิค:

1. การขยายขอบเขตงานวิจัยเพื่อแก้ปัญหาในโลกจริง (Research Scope Expansion)

- **บริบทเดิม:** ตามกรอบการวิจัยแบบดั้งเดิม (Baseline) และเปเปอร์กลุ่มแรกที่เตรียมไว้ (เช่น SciCap, DePlot, MatCha) จะจำกัดการทดลองอยู่บนรูปภาพประเภทแผนภูมิข้อมูลทางสถิติเชิงเดี่ยว (Single-panel Graph Plots/Charts) เท่านั้น
- **ทิศทางใหม่เพื่อแก้ปัญหา:** ในการใช้งานจริง เอกสารวิชาการอุดมไปด้วยภาพถ่ายธรรมชาติและภาพถ่ายเชิงทดลองที่หลากหลาย (Natural/Generic Images) เช่น ภาพถ่ายสัตว์ (สุนัข, แมว) สภาพพื้นที่ (ป่าไม้) หรือสไลด์เนื้อเยื่อ ซึ่งโมเดลประเภทดึงความสัมพันธ์ตัวเลขของชาร์ต (เช่น DePlot) ไม่สามารถทำงานได้ การสร้างระบบที่ยืดหยุ่นพอจะเข้าใจภาพทั่วไปเหล่านี้จึงสร้างผลกระทบและคุณค่าเชิงวิชาการในโลกความเป็นจริงได้มากกว่า

2. การเปลี่ยนผ่านจาก Figure Captioning สู่ Multimodal RAG

- **บริบทเดิม:** วัตถุประสงค์ขั้นพื้นฐานเดิมมุ่งเป้าไปที่ "Figure Captioning" หรือการสร้างคำบรรยายใต้ภาพให้ใกล้เคียงกับรูปแบบการเขียนของมนุษย์ เพื่อวัดคะแนนทางภาษา (Language Metrics เช่น BLEU, ROUGE)
- **ทิศทางใหม่เพื่อแก้ปัญหา:** ขยับเป้าหมาย (Milestone) สู่ระดับระบบงานประยุกต์ คือการพัฒนา **ระบบสืบค้นข้อมูลเชิงความหมายหลายโหมด (Multimodal Retrieval-Augmented Generation)** ซึ่งไม่เพียงแต่เขียนแคปชันภาพเพื่อส่งมอบงานวิจัย แต่เป็นการทำ Indexing รูปภาพเพื่อนำกลับมาสืบค้นและดึงข้อมูลมาตอบคิวรีของผู้ใช้ ส่งผลให้เกณฑ์การวัดผลขยายขอบเขตไปสู่ **ประสิทธิภาพในการสืบค้น (Retrieval/Search Performance เช่น MRR, Recall)**

3. การคัดกรองข้อมูลเพื่อประหยัดเวลาศึกษา (Focused Literature Selection)

- **บริบทเดิม:** คลังข้อมูลมีเปเปอร์เกี่ยวกับการวิเคราะห์รูปภาพและแผนภูมิวิชาการจำนวนมาก (ระบุไว้ในเล่มจัดลำดับการอ่าน) ซึ่งมีเนื้อหาและเป้าหมายย่อยที่กระจัดกระจายกันไป

- **ทิศทางใหม่เพื่อแก้ปัญหา:** ผู้วิจัยมุ่งเป้าคัดแยกและสกัดกลุ่มเปเปอร์ที่เป็น "เสียงรบกวน (Noise)" ออกไป เช่น การข้ามเปเปอร์ในสายแปลงชาร์ตเป็นตาราง (Chart-to-Table) หรือการแปลงแผนภูมิเป็นสมการคณิตศาสตร์ที่มีความซับซ้อนสูง และหันมาให้ความสนใจอย่างลึกซึ้งต่อแนวคิดการยึดโยงภาพเข้ากับข้อความอ้างอิงในระดับทั้งเอกสาร (Document-level Multimodal Fusion) เพื่อนำสถาปัตยกรรมทางทฤษฎีมาวางโครงสร้างระบบทันที

4. การหาหลักฐานทางวิชาการมารองรับสถาปัตยกรรมระบบ (Methodology & Thesis Justification)

- **แนวคิดการทำดัชนี (Indexing Concept):** การเก็บและดึงข้อมูลของรูปภาพธรรมชาติ (ภาพถ่ายสัตว์, ป่าไม้) ในฐานข้อมูลจะใช้แนวคิดของ **MIND-RAG (2025)** ที่ทำการวิเคราะห์ภาพร่วมกับย่อหน้าที่กล่าวอ้างถึงภาพ (Surrounding context) แล้วนำมาสรุปเป็นคุณลักษณะคำอธิบายเชิงข้อความความหนาแน่นสูง (Semantic Summary) เพื่อแปลงโจทย์ให้อยู่ในรูปเวกเตอร์ภาษาธรรมชาติที่ค้นหาได้ง่ายและแม่นยำ
- **ประโยชน์ในการทำรูปเล่ม:** ข้อมูลวิจัยเหล่านี้จะทำหน้าที่เป็นแกนหลักในบทที่ 2 (วรรณกรรมวิจัยที่เกี่ยวข้อง) และบทที่ 3 (ระเบียบวิธีวิจัย) เพื่อใช้ชี้แจงกับคณะกรรมการประเมินวิทยานิพนธ์ว่า ระบบสืบค้นภาพถ่ายทั่วไปบนเอกสารวิชาการไม่ได้สร้างขึ้นมาจากสมมติฐานส่วนบุคคล แต่พัฒนาต่อยอดจากหลักฐานทางวิทยาศาสตร์และเปเปอร์ที่ได้รับการยอมรับในระดับสากล

> **บันทึกทิศทางการดำเนินงานถัดไป (Next Steps):**
> 1. จัดเตรียมโมดูลสำหรับแบ่งแยกรูปภาพ (Layout Analysis) โดยคัดเลือกว่ารูปใดเป็น Graph/Plot และรูปใดเป็น Natural/Generic Image (หมา, แมว, ป่าไม้)
> 2. พัฒนาสคริปต์สกัดข้อความแวดล้อมรอบรูปภาพ (Mention Text/Captions)
> 3. นำข้อมูลข้อความแวดล้อมผสานเข้ากับตัวภาพเพื่อสร้างดัชนีข้อความอธิบายภาพผ่าน API ของ 'latex-server' และ MLLM ต่อไป

บันทึกเชิงวิเคราะห์โดย Antigravity (Gemini 3.5 Flash) · June 21, 2026