

## การเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง

นภัสกร กรวยสวัสดิ์<sup>1</sup> และ แกมกาญจน์ สมประเสริฐศรี<sup>2</sup>

Received 04/08/2022 Revised 27/09/2022 Accepted 21/12/2022

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด เป็นการวิจัยเชิงคุณภาพ ใช้การวิเคราะห์เนื้อหาและการวิเคราะห์เอกสาร เครื่องมือที่ใช้ในการรวบรวมข้อมูลคือ แบบวิเคราะห์ข้อมูล ผลการศึกษาพบว่า คำสำคัญและหัวเรื่องต่างก็มีความสำคัญกับการสืบค้นทรัพยากรสารสนเทศ แต่ผู้ใช้ส่วนใหญ่เลือกสืบค้นโดยใช้ทางเลือกคำสำคัญมากที่สุด แม้จะได้ผลลัพธ์มากแต่มีรายการที่ตรงกับความต้องการของผู้ใช้น้อยกว่าการสืบค้นจากหัวเรื่อง 2) พัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง เป็นการวิจัยเชิงทดลอง ใช้แบบจำลอง Word2Vec สร้างการฝังคำในการหาหัวเรื่องที่มีความคล้ายคลึงกับคำค้น ผลการวัดประสิทธิภาพของการสืบค้นโดยคำนึงถึงการจัดลำดับผลลัพธ์ พบว่า การสืบค้นโดยใช้คำสำคัญร่วมกับหัวเรื่องด้วยแบบจำลอง Word2Vec ให้ประสิทธิภาพสูงกว่าการสืบค้นโดยใช้คำสำคัญ ซึ่งสามารถนำไปใช้เป็นแนวทางในการพัฒนาระบบสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่องแบบอัตโนมัติ เพื่อช่วยให้ได้ผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้มากที่สุด

**คำสำคัญ:** การสืบค้นทรัพยากรสารสนเทศ; คำสำคัญ; หัวเรื่อง; การฝังคำ; Word2Vec

<sup>1</sup> อาจารย์ หน่วยงานวิจัยสร้างสรรค์สารสนเทศดิจิทัลและนวัตกรรม คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม  
อีเมล napassakorn.k@msu.ac.th

<sup>2</sup> ผู้ช่วยศาสตราจารย์ หน่วยงานวิจัยสร้างสรรค์สารสนเทศดิจิทัลและนวัตกรรม คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม  
อีเมล gamgarns@msu.ac.th, ผู้ประพันธ์บรรณกิจ

# Enhancing Information Search Results Using Both Keywords & Subject Headings

Napassakorn Kruaysawat<sup>1</sup> and Gamgarn Somprasertsri<sup>2</sup>

Received 04/08/2022 Revised 27/09/2022 Accepted 21/12/2022

## Abstract

This research aims to investigate the importance associated with the use of both keywords and subject headings when searching library resources for information, using a Qualitative Research method. For the purposes of this study, content- and document analysis tools were used to collect data and compile data analysis forms. The research results revealed that keywords and subject headings are indeed important when searching library resources for information. However, most users tend to only use keywords in their searches. Despite the high number of search results, there are rather fewer items that match the user's needs due to the absence of subject headings. An experimental study was conducted to investigate whether the use of both keywords and subject headings would enhance the quality of search results. In the process, the Word2Vec model was used to generate embedded words to find subject headings that were similar to respective keywords. The results of the measurement for the efficiency of search results by ranking found that the results produced from using both keywords and subject headings, with the help of Word2Vec model, were more efficient than keyword-based searches alone. This outcome can be used as a guideline for developing an automated information search system that incorporates both keywords and subject headings, which could generate higher accuracy of search results that users often demand.

**Keywords:** Information resources searching, Keywords, Subject headings, Word Embedding, Word2Vec

---

<sup>1</sup> Lecturer, Creativity in Digital Information and Innovation Research Unit, Faculty of Informatics, Mahasarakham University, Thailand. Email: napassakorn.k@msu.ac.th

<sup>2</sup> Assistant Professor, Creativity in Digital Information and Innovation Research Unit, Faculty of Informatics, Mahasarakham University. Email: gamgarns@msu.ac.th, Corresponding author.

## บทนำ

ห้องสมุดเป็นแหล่งสารสนเทศที่มีความสำคัญ เนื่องจากช่วยให้ผู้ใช้สามารถศึกษาค้นคว้า วิจัย และช่วยสนับสนุนการเรียนรู้ในทุกระดับ รวมทั้งเป็นแหล่งที่ช่วยส่งเสริมให้เกิดการเรียนรู้ตลอดชีวิต อีกทั้งห้องสมุดยังเป็นแหล่งสารสนเทศที่ให้บริการทรัพยากรสารสนเทศทุกประเภท ได้แก่ ทรัพยากรสารสนเทศสิ่งตีพิมพ์ เช่น หนังสือ วารสาร นิตยสาร หนังสือพิมพ์ ทรัพยากรสารสนเทศประเภทสิ่งไม่ตีพิมพ์ เช่น รูปภาพ แผนที่ แผนภูมิ และทรัพยากรอิเล็กทรอนิกส์ เช่น หนังสืออิเล็กทรอนิกส์ วารสารอิเล็กทรอนิกส์ เป็นต้น โดยห้องสมุดมีเครื่องมือที่ช่วยในการค้นหาทรัพยากรสารสนเทศที่สำคัญ คือ ฐานข้อมูลรายการบรรณานุกรมออนไลน์ (Online Public Access Catalog : OPAC) โดยผู้ใช้สามารถสืบค้นได้จากหลายทางเลือก เช่น ชื่อผู้แต่ง ชื่อเรื่อง หัวเรื่อง คำสำคัญ เลขหมู่หนังสือ เลข ISBN เป็นต้น จากการศึกษางานวิจัยที่เกี่ยวกับการสืบค้นทรัพยากรสารสนเทศในห้องสมุด ในส่วนของปัญหาการสืบค้นทรัพยากรสารสนเทศโดยการใช้ OPAC พบว่า ผู้ใช้ประสบปัญหาในการกำหนดคำค้นมากที่สุด ผู้ใช้ไม่ทราบว่าจะกำหนดคำค้นใดในการสืบค้นสารสนเทศเพื่อให้ได้ข้อมูลที่ตรงความต้องการ ผู้ใช้ส่วนใหญ่มักกำหนดคำค้นจากคำสำคัญ (Keyword) ศัพท์อิสระ (Free term) มากกว่าหัวเรื่อง (Subject) (Panommit, Arayaphan & Julrode, 2017 ; Kruaysawat, Manmart & Kabmala, 2018) เนื่องจาก การสืบค้นด้วยคำสำคัญเป็นคำที่ผู้ใช้คุ้นเคย มีความทันสมัย มักเป็นคำศัพท์ที่เกิดขึ้นใหม่เสมอ ในขณะที่ห้องสมุดให้ความสำคัญกับหัวเรื่องมากกว่า เพราะนอกจากจะช่วยในการจัดหมวดหมู่ทรัพยากรสารสนเทศแล้วยังช่วยในการสืบค้นทรัพยากรสารสนเทศด้วย แต่เนื่องจากคำที่ใช้เป็นหัวเรื่องเป็นคำควบคุมหรือเป็นศัพท์บัญญัติ กำหนดโดยบรรณารักษ์หรือผู้เชี่ยวชาญแต่ละสาขา ซึ่งเป็นคำศัพท์วิชาการส่งผลให้ผู้ใช้ไม่รู้จักหรือไม่คุ้นเคย (Kruaysawat, Manmart & Kabmala, 2018)

นอกจากนั้น จากการศึกษางานวิจัยต่าง ๆ ทั้งในประเทศและต่างประเทศ พบว่า มีงานวิจัยจำนวนมากที่ศึกษาการพัฒนาประสิทธิภาพการค้นคืนสารสนเทศ โดยใช้แนวคิด ทฤษฎีต่าง ๆ เช่น การใช้ ThaiWordNet การนำแนวคิด Folksonomy และ Ontology มาช่วยในการเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศ การพัฒนานวัตกรรมการค้นคืนสารสนเทศแบบสหความสัมพันธ์ เทคนิคการค้นคืนสารสนเทศข้ามภาษา (ไทย-อังกฤษ) โดยใช้ออนโทโลยี เป็นต้น (Intraprapan, 2014 ; Intraprapan & Kesorn, 2014 ; Kruaysawat, 2016 ; Lertmahakiat & Mingkwan, 2010 ; Yu & Chen, 2020) นอกจากนั้นจากการศึกษาปัญหาของการเพิ่มประสิทธิภาพการสืบค้นด้วยคำสำคัญร่วมกับหัวเรื่อง พบปัญหาที่สำคัญ คือการจับคู่ (Matching) ระหว่างคำสำคัญที่ผู้ใช้กำหนดเป็นคำสืบค้นกับหัวเรื่องที่เป็นตัวแทนเนื้อหาของทรัพยากรสารสนเทศที่บรรณารักษ์หรือนักสารสนเทศกำหนดขึ้น ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด และนำเสนอแนวทางการเพิ่มประสิทธิภาพของการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้แบบจำลอง Word2Vec สร้างการฝังคำ (Word embedding) ในการขยายคำค้นหรือหัวเรื่องที่เกี่ยวข้องกับคำสำคัญที่ผู้ใช้กำหนดในการสืบค้น และใช้วิธีการวัดความคล้ายคลึงโคไซน์ (Cosine similarity) ระหว่างคำค้นและหัวเรื่อง เพื่อช่วยให้ผู้ใช้สามารถสืบค้นทรัพยากรสารสนเทศได้ตรงตามความต้องการมากที่สุด

## การทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การสืบค้นทรัพยากรสารสนเทศ หมายถึง กระบวนการค้นหาทรัพยากรสารสนเทศที่ต้องการ โดยใช้เครื่องมือสืบค้นทรัพยากรสารสนเทศที่สถาบันบริการสารสนเทศจัดเตรียมไว้ให้ ซึ่งแบ่งออกเป็น 2 วิธี คือ (1) การสืบค้นทรัพยากรสารสนเทศด้วยระบบมือ (Manual system) สามารถกระทำได้โดยผ่านเครื่องมือหลายประเภท เช่น บัตรรายการ บัตรบรรณานุกรม บรรณานุกรม (2) การสืบค้นทรัพยากรสารสนเทศด้วยระบบคอมพิวเตอร์ (Computer system) สามารถกระทำได้โดยผ่านอุปกรณ์คอมพิวเตอร์ ในการค้นหาข้อมูลจากฐานข้อมูลต่าง ๆ ได้แก่ ฐานข้อมูล OPAC ฐานข้อมูลซีดีรอม ฐานข้อมูลออนไลน์ ฐานข้อมูลบนอินเทอร์เน็ต (Information Technology for Works, 2017) โดยงานวิจัยนี้จะมุ่งเน้นการสืบค้นทรัพยากรสารสนเทศผ่าน OPAC เท่านั้น

ในขณะที่การสืบค้นทรัพยากรสารสนเทศผ่านฐานข้อมูลรายการบรรณานุกรมออนไลน์ หรือ OPAC เป็นเครื่องมือที่เป็นตัวช่วยในการสืบค้นรายการทรัพยากรสารสนเทศของห้องสมุด เช่น หนังสือทั่วไป สิ่งพิมพ์ต่อเนื่อง สื่อทัศนวัสดุ วิทยานิพนธ์ ทำให้ผู้ใช้บริการสามารถสืบค้นทรัพยากรสารสนเทศได้ตามที่ต้องการ ได้อย่างสะดวกและรวดเร็ว (Informatic Innovation Center of Excellence, Walailak University, 2019) ซึ่งในการสืบค้นทรัพยากรสารสนเทศผ่านฐานข้อมูลรายการบรรณานุกรมออนไลน์ ผู้ใช้สามารถเลือกใช้ทางเลือกในการสืบค้นได้จากหลายทางเลือก เช่น ชื่อผู้แต่ง ชื่อเรื่อง คำสำคัญ หัวเรื่อง เลขหมู่หนังสือ เป็นต้น ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้องกับการศึกษาพฤติกรรมของผู้ใช้ในการสืบค้นทรัพยากรสารสนเทศ พบว่า ผู้ใช้ส่วนมากเลือกใช้คำสำคัญในการสืบค้นมากที่สุด (TaKham, Kabmala & Manmart, 2016 ; Panommit, Arayaphan & Julrode, 2017 ; Kruaysawat, Manmart & Kabmala, 2018 ; Sriwanta & Arayaphan, 2022) เนื่องจาก คำสำคัญซึ่งหมายถึง คำหรือวลีสำคัญในชื่อเรื่อง หัวเรื่อง เนื้อหา และสาระสังเขป เป็นภาษาอิสระที่ทำให้สามารถสืบค้นสารสนเทศที่เกี่ยวข้องจากฐานข้อมูลได้ เป็นคำหรือวลีที่มีความสำคัญหรือมีนัยสำคัญต่อระบบการสืบค้นสารสนเทศ เป็นคำที่มีสาระหรือประโยชน์ ในชื่อเรื่องหรือเนื้อหาจากเอกสาร นอกจากนั้น การใช้คำสำคัญเพื่อการสืบค้นจะมีประโยชน์เมื่อหัวเรื่องของเอกสารนั้นไม่ชัดเจนหรือไม่เป็นที่รับทราบโดยทั่วกัน (Nimsomboon, 2018)

การกำหนดหัวเรื่องมีวัตถุประสงค์เพื่อให้สามารถค้นหาทรัพยากรสารสนเทศที่เกี่ยวข้องได้ทั้งหมด ช่วยให้ทรัพยากรสารสนเทศที่มีเนื้อหาวิชาเดียวกันมารวมอยู่ในที่เดียวกัน โดยมีได้คำนึงถึงความแตกต่างในด้านคำศัพท์ ภาษา ผู้แต่งหรืออื่น ๆ และช่วยอำนวยความสะดวกให้ผู้เชี่ยวชาญและผู้ทั่วไปในการค้นหาทรัพยากรสารสนเทศ ให้เข้าใจความหมายได้ เป็นการกำหนดคำบรรยายเนื้อหาทรัพยากรสารสนเทศอย่างเป็นสากล โดยการใช้คำหรือสำนวนที่สั้นและเฉพาะเจาะจง รวมทั้งเป็นเครื่องมือสำหรับผู้ใช้อาคารในการเลือกทรัพยากรสารสนเทศได้ตรงตามความต้องการ (Shera & Egan, 1956)

งานวิจัยนี้ได้นำเสนอวิธีการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้แบบจำลอง Word2Vec ในการสร้างการฝังคำสำหรับขยายคำค้น คือ หัวเรื่องที่เกี่ยวข้องกับคำสำคัญที่ผู้กำหนดในการสืบค้น ซึ่งการฝังคำ หรือ Word Embedding เป็นแนวคิดของการสร้างแบบจำลองทางภาษาสำหรับการประมวลผลภาษาธรรมชาติ ที่เป็นการแทนคำคำศัพท์ด้วยเวกเตอร์ที่มีจำนวนมิติลดลง

โดยอาศัยการจับคู่คำกับเวกเตอร์ของจำนวนจริงในปริภูมิเวกเตอร์ (Vector space) การจับคู่ดังกล่าวจะอาศัยการพิจารณาการปรากฏของคำนั้น ๆ กับคำอื่นที่ปรากฏแวดล้อม เนื่องจากคำที่ถูกแวดล้อมด้วยบริบทของคำที่คล้ายกันจะมีความหมายใกล้เคียงกัน แบบจำลองที่นิยมใช้สร้างการฝังคำ คือ Word2Vec ซึ่งเป็นการสร้างเวกเตอร์ของแต่ละคำขึ้นมา โดยดูจากความสัมพันธ์ระหว่างคำ ๆ นั้นกับคำรอบข้างที่เป็นบริบท (Context) ของคำ แบบจำลองนี้สามารถกำหนดหรือลดจำนวนมิติ (Dimension) ของข้อมูลได้ โดยไม่ต้องใช้จำนวนมิติเท่ากับจำนวนคำเสมอไป (Mikolov et al., 2013) สำหรับการจับคู่ระหว่างคำสำคัญซึ่งเป็นคำค้นกับหัวเรื่องนั้น งานวิจัยนี้ใช้วิธีการวัดความคล้ายคลึงโคซายน์ ซึ่งเป็นวิธีการวัดความคล้ายคลึงที่นิยมใช้ในการค้นคืนสารสนเทศ (Han, Kamber & Pei, 2012) โดยการหาความคล้ายคลึงจากค่าความต่างของข้อมูล 2 ชุด ที่เกิดขึ้นบนปริภูมิเวกเตอร์ ค่าความคล้ายคลึงโคซายน์มีค่าอยู่ระหว่าง 0-1 โดยค่ายิ่งเข้าใกล้ 1 หมายถึงข้อมูลนั้นมีความคล้ายคลึงกันสูง

สำหรับการวัดประสิทธิภาพการค้นคืนสารสนเทศ สามารถดำเนินการได้ 2 รูปแบบ ได้แก่ (1) การวัดประสิทธิภาพของระบบค้นคืนสารสนเทศที่คำนึงถึงการจัดลำดับผลลัพธ์ มักจะถูกวัดโดยการใช้ค่าความระลึก (Recall) และค่าความเที่ยงตรง (Precision) ซึ่งค่าความระลึกจะวัดความสามารถของระบบในการที่จะดึงเอกสารที่เกี่ยวข้องออกมา ในขณะที่ค่าความเที่ยงตรงจะวัดความสามารถในการดึงเอกสารที่เกี่ยวข้องได้ถูกต้องหรือสามารถที่จะจัดเอกสารที่ไม่เกี่ยวข้องออกไปได้ (Baeza-Yates & Ribeiro-Neto, 2011) (2) การวัดประสิทธิภาพของระบบค้นคืนสารสนเทศที่คำนึงถึงการจัดลำดับผลลัพธ์ (Ranking-based evaluation) ได้แก่ ค่าความเที่ยงตรงที่อันดับ k (Precision@k: P@K) คือค่าความเที่ยงตรงที่ตำแหน่ง k ใด ๆ ที่สนใจ และค่าเฉลี่ยความเที่ยงตรงที่อันดับ k (Average Precision@k: AP@k) คือค่าเฉลี่ยความเที่ยงตรงที่ตำแหน่ง k ใด ๆ ที่สนใจ (Buettcher, Clarke & Cormack, 2010) ซึ่งเป็นการวัดประสิทธิภาพที่นำมาใช้ในงานวิจัยนี้

### วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด
2. เพื่อพัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง

### วิธีการศึกษา

การวิจัยครั้งนี้ประกอบไปด้วยการวิจัยเชิงคุณภาพและการวิจัยเชิงทดลอง โดยมีรายละเอียดดังนี้

1. การศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด เป็นการวิจัยเชิงคุณภาพ เพื่อทราบสภาพการใช้งานของฐานข้อมูลรายการบรรณานุกรมออนไลน์ (OPAC) ในส่วนที่เกี่ยวข้องกับการสืบค้นข้อมูลโดยการใช้คำสำคัญและหัวเรื่องในการสืบค้น โดยผู้วิจัยใช้การวิเคราะห์จากฐานข้อมูล OPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม สำนักหอสมุด มหาวิทยาลัยขอนแก่น การใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC ซึ่งผู้วิจัยเลือกศึกษาจากเว็บไซต์ของทั้ง 2 แห่งนี้ เนื่องจากมีการใช้ระบบห้องสมุดอัตโนมัติแตกต่างกัน เพื่อให้ทราบการใช้งานของระบบ OPAC และการใช้การวิเคราะห์เอกสาร (Document analysis) จากคู่มือการใช้งานระบบ OPAC ของระบบห้องสมุดอัตโนมัติ Walai Autolib และระบบห้องสมุดอัตโนมัติ Sierra รวมทั้งจากงานวิจัยต่าง ๆ ที่เกี่ยวกับการสืบค้น

ทรัพยากรสารสนเทศของห้องสมุดผ่านระบบ OPAC และงานวิจัยที่เกี่ยวกับการเพิ่มประสิทธิภาพการค้นคืนสารสนเทศ เครื่องมือที่ใช้ในการรวบรวมข้อมูลคือ แบบวิเคราะห์ข้อมูล ประกอบไปด้วยประเด็นที่ศึกษา ได้แก่ 1) เปรียบเทียบการสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม และสำนักหอสมุด มหาวิทยาลัยขอนแก่น 2) การสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC โดยการใช้คำสำคัญและหัวเรื่อง 3) พฤติกรรมการสืบค้นทรัพยากรสารสนเทศของผู้ใช้ผ่านระบบ OPAC โดยการใช้คำสำคัญและหัวเรื่อง ซึ่งผู้วิจัยดำเนินการรวบรวมข้อมูลด้วยตนเองและนำมาวิเคราะห์ข้อมูลเพื่อสรุปประเด็น

2. การพัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง เป็นการวิจัยเชิงทดลอง โดยใช้ Word Embedding ในการขยายคำค้นหรือหัวเรื่องที่เกี่ยวข้องกับคำสำคัญที่ผู้ใช้งานกำหนดในการสืบค้น แบ่งเป็น 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูลที่ใช้ในการทดลองในงานวิจัยนี้ แบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลทรัพยากรสารสนเทศ และข้อมูลหัวเรื่อง มีรายละเอียดในการเตรียมข้อมูลดังนี้

1.1 การเตรียมข้อมูลทรัพยากรสารสนเทศ งานวิจัยนี้ใช้ข้อมูลทรัพยากรสารสนเทศภาษาอังกฤษที่สืบค้นได้จาก WebOPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม โดยใช้ในการค้นคืนด้วยคำสำคัญ “database system” ซึ่งพบรายการทั้งหมด 24 รายการ และให้ผู้ใช้จัดลำดับความเกี่ยวข้อง พบว่ามีหนังสือที่เกี่ยวข้องและตรงกับที่ผู้ใช้งานต้องการ จำนวน 7 รายการ มีหนังสือที่ไม่เกี่ยวข้องและไม่ตรงกับความต้องการของผู้ใช้ จำนวน 17 รายการ

1.2 การเตรียมข้อมูลหัวเรื่อง ใช้ข้อมูลหัวเรื่องจากเว็บไซต์ หัวเรื่องหอสมุดรัฐสภาอเมริกัน (Library of Congress Subject Heading : LCSH) โดยใช้คำค้น “database system”

ขั้นตอนที่ 2 การหาหัวเรื่องที่สอดคล้องกับคำสำคัญ (Mapping to Subject Heading) ขั้นตอนนี้จะสร้างการฝังคำด้วยแบบจำลอง Word2Vec ในการหาหัวเรื่อง (subject heading) ที่มีค่าความคล้ายคลึงกับคำสำคัญที่ผู้ใช้ใช้ในการสืบค้น โดยใช้ข้อมูลจากหัวเรื่องหอสมุดรัฐสภาอเมริกัน (LCSH) ที่เตรียมในขั้นตอนที่ผ่านมา การหาหัวเรื่องที่สอดคล้องกับคำสำคัญพิจารณาจากค่าความคล้ายคลึงโคซายน์ของหัวเรื่องกับคำสำคัญ โดยหัวเรื่องที่ใกล้เคียงคำสำคัญนั้นจะต้องมีค่าความคล้ายคลึงโคซายน์มากที่สุด ซึ่งถ้าคะแนนค่าความคล้ายคลึงมีค่าเข้าใกล้ 1 หมายถึงคำทั้งสองคำ มีความหมายคล้ายคลึงกัน ในขณะเดียวกันถ้าคะแนนค่าความคล้ายคลึงมีค่าเข้าใกล้ 0 หมายถึงทั้งสองคำไม่มีความคล้ายคลึงกันเชิงความหมาย งานวิจัยนี้ใช้ spaCy ในการทำ Word2Vec ซึ่งจะนำมาใช้ร่วมกับความคล้ายคลึงโคซายน์ในการหาหัวเรื่องจากคำค้นของผู้ใช้

ขั้นตอนที่ 3 การวัดประสิทธิภาพของการสืบค้น เป็นการเปรียบเทียบระหว่างผลลัพธ์การสืบค้นในการใช้ทางเลือกคำสำคัญ และหัวเรื่องด้วย Word2Vec โดยการใช้วัดประสิทธิภาพการค้นคืนที่คำนึงถึงการจัดลำดับผลลัพธ์ ได้แก่ ค่าความเที่ยงตรงที่อันดับ k ( $P@k$ ) และค่าเฉลี่ยความเที่ยงตรงที่อันดับ k ( $AP@k$ )

## ผลการศึกษา

ผู้วิจัยได้ดำเนินการศึกษาการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่องโดยแบ่งผลการศึกษาออกเป็น 2 ส่วน ได้แก่ 1) สภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้น

ทรัพยากรสารสนเทศของห้องสมุด 2) การพัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศ ด้วยคำสำคัญร่วมกับหัวเรื่อง สามารถแสดงรายละเอียดได้ดังนี้

1. สภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด

ผู้วิจัยใช้การวิเคราะห์ข้อมูลจากการศึกษาการสืบค้นทรัพยากรสารสนเทศของห้องสมุดจากเว็บไซต์ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม ที่ใช้ระบบห้องสมุดอัตโนมัติ Sierra และเว็บไซต์ของสำนักหอสมุด มหาวิทยาลัยขอนแก่น ที่ใช้ระบบห้องสมุดอัตโนมัติ Walai Autolib รวมทั้งศึกษาจากเอกสารคู่มือการใช้งานระบบห้องสมุดอัตโนมัติ งานวิจัยต่าง ๆ ที่เกี่ยวกับการสืบค้นทรัพยากรสารสนเทศผ่าน OPAC ซึ่งนำเสนอผลการศึกษา ดังนี้

1) เปรียบเทียบการสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม และสำนักหอสมุด มหาวิทยาลัยขอนแก่น ผลการศึกษาจำแนกออกเป็น 5 ประเด็น ดังนี้

(1) ระบบการสืบค้นทรัพยากรสารสนเทศของห้องสมุด สำนักวิทยบริการ มหาวิทยาลัยมหาสารคามใช้ระบบห้องสมุดอัตโนมัติ Sierra ส่วนสำนักหอสมุด มหาวิทยาลัยขอนแก่น ใช้ระบบห้องสมุดอัตโนมัติ Walai Autolib เดิมสำนักหอสมุด มหาวิทยาลัยขอนแก่นใช้งานระบบห้องสมุดอัตโนมัติ Sierra เช่นเดียวกัน แต่ได้ปรับเปลี่ยนมาใช้งานระบบห้องสมุดอัตโนมัติ Walai Autolib ในภาพรวมทั้งสองระบบสามารถใช้ประโยชน์ในการสืบค้นทรัพยากรสารสนเทศได้ดีเช่นเดียวกัน เพราะมีทางเลือกให้กับผู้ใช้ในการสืบค้น แต่ในส่วนองวิธีการสืบค้นของระบบห้องสมุดอัตโนมัติ Walai Autolib มีวิธีการสืบค้นที่หลากหลายแนวทางมากกว่าระบบห้องสมุดอัตโนมัติ Sierra เช่น สามารถสืบค้นได้จากการ Browse search หมวดหมู่ของทรัพยากร ระบบมีการพัฒนาขึ้นมาเพื่อรองรับระบบการทำงานของ web 2.0 ที่อนุญาตให้ผู้ใช้มีส่วนร่วมในการสร้างแท็ก (Tags) ให้กับรายการบรรณานุกรมและผู้ใช้สามารถสืบค้นผ่านแท็ก (Tags) ได้

(2) ทางเลือกในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด พบว่า สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม สามารถสืบค้นโดยใช้ทางเลือก คำสำคัญ (Keyword) ผู้แต่ง (Author) ชื่อเรื่อง (Title) หัวเรื่อง (Subject) เลขเรียกหนังสือ (Call Number) ในขณะที่สำนักหอสมุด มหาวิทยาลัยขอนแก่น สามารถสืบค้นโดยใช้ทางเลือก คำสำคัญ (Keyword) ชื่อเรื่อง (Title) ผู้แต่ง (Author) เลขมาตรฐานประจำหนังสือ/เลขมาตรฐานประจำวารสาร (ISBN/ISSN) ชื่อวารสาร (Journal title) ซึ่งในส่วนของการเลือกช่องทางสำหรับให้ผู้ใช้สืบค้นขึ้นอยู่กับนโยบายของห้องสมุดนั้น ๆ ว่าต้องการกำหนดทางเลือกใดให้กับผู้ใช้บ้าง ซึ่งทั้งสองห้องสมุดจะมีการกำหนดเหมือนกัน คือ การใช้ทางเลือกในการสืบค้นให้กับผู้ใช้ได้แก่ คำสำคัญ (Keyword) ผู้แต่ง (Author) ชื่อเรื่อง (Title)

(3) วิธีการสืบค้น ระบบห้องสมุดอัตโนมัติ Sierra ผู้ใช้สามารถสืบค้นได้ 2 วิธี คือ 1) การสืบค้นขั้นพื้นฐาน (Basic search) ได้แก่ การใช้คำสำคัญ หัวเรื่อง ชื่อผู้แต่ง ชื่อเรื่อง เลขเรียกหนังสือ 2) การสืบค้นขั้นสูง (Advanced search) ได้แก่ การจำกัดขอบเขตในการสืบค้นทรัพยากรสารสนเทศได้ เช่น การระบุปีพิมพ์ ประเภททรัพยากรสารสนเทศ ส่วนระบบห้องสมุดอัตโนมัติ Walai Autolib ผู้ใช้สามารถสืบค้นได้ 6 แบบ ได้แก่ 1) การสืบค้นแบบ Single search สามารถสืบค้นจากหน้าแรกของ Web OPAC ซึ่งสืบค้นได้ตาม คำสำคัญ ผู้แต่ง ชื่อวารสาร ชื่อหนังสือ ปีที่ ฉบับที่ 2) การสืบค้นขั้นพื้นฐาน (Basic search) เป็นการสืบค้นรายการทรัพยากร ที่สามารถสืบค้นได้เพียงเขตข้อมูลเดียว โดยสืบค้นได้จาก ชื่อเรื่อง (Title) ผู้แต่ง (Author) หัวเรื่อง (Subject) เป็นต้น 3) การสืบค้นขั้นสูง (Advanced search) เป็นการสืบค้นรายการทรัพยากรที่

เฉพาะเจาะจง สามารถจำกัดการสืบค้นและใช้เทคนิคตรรกบูลีน 4) การสืบค้นแบบเรียงลำดับตัวอักษร (Alphabetically search) เป็นการสืบค้นโดยใช้ตัวอักษรตัวแรกของคำค้น จากนั้นไล่เรียงตามลำดับตัวอักษร สามารถสืบค้นได้จากชื่อผู้แต่ง (Author) ชื่อเรื่อง (Title) หัวเรื่อง (Subject) เลขเรียกหนังสือ (Call Number) สำนักพิมพ์ (Publisher) เป็นต้น 5) การสืบค้นระเบียบรายการหลักฐาน (Authority search) เป็นการสืบค้นทรัพยากรอีกส่วนหนึ่งในการสืบค้น ทรัพยากรภายในห้องสมุด โดยสามารถสืบค้นทั้งในส่วน of ชื่อเรื่อง หัวเรื่อง และชื่อชุด 6) การสืบค้นจาก Browse search เป็นการไล่เรียงหมวดหมู่ของทรัพยากร ที่จัดการผ่านกระบวนการจัดหมวดหมู่จากระเบียบหลักฐานหัวเรื่องประกอบด้วยหมวดหมู่หลัก (Category) หมวดย่อย (Sub category) และรายการหัวเรื่อง ผู้ใช้สามารถเรียกดูรายการทรัพยากรจากหัวเรื่องที่จัดหมวดหมู่ไว้แล้ว

(4) การแสดงผลการสืบค้น ทั้งสองระบบสามารถแสดงหน้าจอการสืบค้นในรูปแบบรายการบรรณานุกรม และผู้ใช้สามารถเลือกดูการแสดงผลในรูปแบบการลงรายการ MARC ได้ แต่ในระบบห้องสมุดอัตโนมัติ Sierra มีการแสดงผลในส่วน of สารสังเขปในบางรายการ ทั้งสองระบบมีบางรายการแสดงผลในส่วนสารบัญของทรัพยากรสารสนเทศ นอกจากนี้ ระบบห้องสมุดอัตโนมัติ Walai Autolib ยังสามารถกรองผลการสืบค้นได้จากหลากหลายทางเลือก เช่น ผู้แต่ง หัวเรื่อง ปีพิมพ์ ประเภททรัพยากรสารสนเทศ สถานที่ เลขเรียกหนังสือ หน้าจอการแสดงผลการสืบค้นสามารถดูได้ง่าย และการใช้งานระบบง่ายไม่ซับซ้อน

(5) ในส่วนฟังก์ชันเสริมต่าง ๆ ระบบห้องสมุดอัตโนมัติ Walai Autolib ผู้ใช้สามารถแบ่งปันรายการผ่าน Social network ได้ ผู้ใช้สามารถแสดงความคิดเห็นในแต่ละรายการได้ (Review) ผู้ใช้สามารถเพิ่มแท็ก (Tags) ให้กับรายการทรัพยากรได้ และสามารถสืบค้นรายการทรัพยากรผ่านแท็ก (Tags) แท็กคลาวด์ (Tag cloud) ได้ และผู้ใช้สามารถเห็นสถิติการยืมคืนของทรัพยากรสารสนเทศในแต่ละรายการได้

## 2) การสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC โดยใช้คำสำคัญและหัวเรื่อง

จากการศึกษาการสืบค้นทรัพยากรสารสนเทศผ่านระบบ OPAC ของสำนักวิทยบริการ มหาวิทยาลัย

มหาสารคาม โดยใช้คำสำคัญและหัวเรื่อง ผ่านเว็บไซต์ <http://library.msu.ac.th> สามารถสรุปได้ดังนี้

คำสำคัญเป็นคำที่ผู้ใช้กำหนดขึ้นตามความต้องการสารสนเทศของผู้ใช้ ซึ่งหากผู้ใช้สืบค้นด้วยคำสำคัญ ระบบจะทำการค้นหาคำที่อยู่ในชื่อเรื่องและในส่วนต่าง ๆ ของรายการทรัพยากรสารสนเทศ ซึ่งระบบจะแสดงผลลัพธ์การสืบค้นจำนวนมาก ในขณะที่หัวเรื่อง เป็นคำหรือวลีที่ห้องสมุดกำหนดขึ้นมาแทนเนื้อหาทรัพยากรสารสนเทศ และมีความเป็นมาตรฐาน บรรณารักษ์เป็นผู้กำหนดหัวเรื่องแทนเนื้อหาทรัพยากรสารสนเทศนั้น ๆ โดยแต่ละรายการอาจมีหัวเรื่องได้มากกว่า 1 หัวเรื่อง ดังนั้น หากผู้ใช้กำหนดทางเลือกในการสืบค้นทรัพยากรสารสนเทศโดยใช้หัวเรื่อง ระบบจะทำการค้นคืนคำที่อยู่ในเขตข้อมูลหัวเรื่องที่มีความสัมพันธ์กับคำที่ผู้ใช้กำหนดและแสดงผลออกมาทั้งหมด ซึ่งรายการที่แสดงผลออกมาจะมีจำนวนน้อยกว่าการสืบค้นด้วยคำสำคัญ แต่มักจะตรงกับความต้องการของผู้ใช้มากกว่า รวมทั้งมีรายการโยงเพิ่มเติมให้กับผู้ใช้ด้วย แต่ในขณะเดียวกัน ผลลัพธ์การสืบค้นด้วยหัวเรื่องก็ขึ้นอยู่กับจำนวนทรัพยากรสารสนเทศของแต่ละห้องสมุดด้วย

นอกจากนั้น จากการเปรียบเทียบการสืบค้นโดยใช้คำสำคัญและหัวเรื่องของระบบ OPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม โดยผู้วิจัยได้กำหนดคำค้นที่มีทั้งคำศัพท์ภาษาไทย และ



ภาษาอังกฤษ ได้แก่ การจัดการความรู้ การพัฒนาเว็บไซต์ โปรแกรมคอมพิวเตอร์ ระบบฐานข้อมูล แอปพลิเคชัน Application, Computer program, Database, Database system, Knowledge management, Website สามารถแสดงได้ดังตาราง 1

ตาราง 1 เปรียบเทียบการสืบค้นโดยใช้คำสำคัญและหัวเรื่องของระบบ OPAC ของสำนักวิทยบริการมหาวิทยาลัยมหาสารคาม

คำค้น	การสืบค้น	ผลการสืบค้น	คำค้น	การสืบค้น	ผลการสืบค้น
ระบบฐานข้อมูล	คำสำคัญ	492	Database	คำสำคัญ	389
	หัวเรื่อง	1		หัวเรื่อง	243
การจัดการความรู้	คำสำคัญ	1,316	Database system	คำสำคัญ	47
	หัวเรื่อง	-		หัวเรื่อง	1
โปรแกรมคอมพิวเตอร์	คำสำคัญ	3,559	Knowledge management	คำสำคัญ	355
	หัวเรื่อง	1,413		หัวเรื่อง	113
แอปพลิเคชัน	คำสำคัญ	20	Computer program	คำสำคัญ	762
	หัวเรื่อง	-		หัวเรื่อง	277
การพัฒนาเว็บไซต์	คำสำคัญ	243	Application	คำสำคัญ	934
				หัวเรื่อง	62
	หัวเรื่อง	134	Website	คำสำคัญ	182
				หัวเรื่อง	-

จากตารางที่ 1 สามารถอธิบายได้ว่า การสืบค้นด้วยคำสำคัญจะได้ผลการสืบค้นจำนวนมาก และมีรายการที่ไม่ตรงกับความต้องการจำนวนมากเช่นเดียวกัน ในขณะที่การสืบค้นโดยใช้หัวเรื่อง ผลลัพธ์การสืบค้นน้อยกว่าการใช้คำสำคัญ แต่ผลการสืบค้นตรงกับความต้องการมากกว่า และพบว่ามีคำศัพท์บางคำไม่ได้ถูกกำหนดให้ใช้เป็นหัวเรื่อง ได้แก่ คำว่า การจัดการความรู้ แอปพลิเคชัน Website ดังนั้น ผู้ใช้สามารถสืบค้นและได้ผลลัพธ์จากการสืบค้นโดยใช้ทางเลือกคำสำคัญ และมีบางคำเมื่อผู้ใช้สืบค้นจากหัวเรื่อง ระบบมีการทำรายการโยงให้กับผู้ใช้ เพื่อให้ทราบว่าห้องสมุดไม่ได้กำหนดให้เป็นหัวเรื่อง คือคำว่า การจัดการความรู้ แต่ห้องสมุดใช้คำว่า การบริหารองค์ความรู้ เป็นหัวเรื่อง ดังนั้น เมื่อเลือกไปที่รายการโยงคำว่า การบริหารองค์ความรู้ จะได้ผลลัพธ์การค้นจำนวน 326 รายการ คำว่า ระบบฐานข้อมูล ผู้ใช้สามารถสืบค้นได้จากทางเลือกคำสำคัญและหัวเรื่อง โดยสืบค้นจากคำสำคัญจะได้ผลลัพธ์การค้นจำนวน 492 รายการ เมื่อใช้ทางเลือก หัวเรื่อง ในการสืบค้น พบว่า มีเพียงรายการเดียว ซึ่งสาเหตุที่พบจำนวนน้อย เนื่องจากห้องสมุดมีหนังสือที่กำหนดหัวเรื่องว่า ระบบฐานข้อมูล จำนวน 1 รายการ แต่เมื่อผู้วิจัยใช้คำค้น ฐานข้อมูล ในการสืบค้นจากทางเลือก หัวเรื่อง พบว่า ได้ผลลัพธ์การสืบค้นจำนวน 125 รายการ

### 3) พฤติกรรมการสืบค้นทรัพยากรสารสนเทศของผู้ใช้ผ่าน OPAC โดยการใช้คำสำคัญและหัวเรื่อง

จากการศึกษางานวิจัยที่เกี่ยวกับพฤติกรรมของผู้ใช้ในการสืบค้นทรัพยากรสารสนเทศผ่าน OPAC ในอดีต พบว่าผู้ใช้ส่วนใหญ่เลือกใช้ทางเลือก ชื่อเรื่อง ในการสืบค้นมากที่สุด (Cheysiri, 2000 ; Jutopama, 2007 ; Ouantrai, 2008) ในขณะที่ผลการศึกษางานวิจัยในปัจจุบันพบว่า ผู้ใช้เลือกใช้ทางเลือก คำสำคัญมากที่สุด (TaKham, Kabmala & Manmart, 2016 ; Panommit, Arayaphan & Julrode, 2017 ; Kruaysawat, Manmart & Kabmala, 2018) ส่วนปัญหาการกำหนดคำค้นที่ผู้ใช้ไม่ทราบจะกำหนดคำค้นใด ยังคงเป็นปัญหาของผู้ใช้ตั้งแต่อดีตจนถึงปัจจุบัน ดังนั้น จึงอาจกล่าวได้ว่าเหตุผลที่ผู้ใช้ในปัจจุบันเลือกใช้ทางเลือก คำสำคัญในการสืบค้นมากกว่าการใช้ ชื่อเรื่องหรือหัวเรื่อง เนื่องจากการพัฒนาเทคโนโลยีทางการสืบค้นมากขึ้น รวมทั้ง OPAC ก็มีความสามารถในการสืบค้นจากคำสำคัญได้มากขึ้นหรือมีการนำแท็ก (Tags) มาใช้ประโยชน์ในการสืบค้นทรัพยากรสารสนเทศ เช่น OPAC ของสำนักหอสมุด มหาวิทยาลัยขอนแก่น ซึ่งเป็นการเพิ่มทางเลือกให้กับผู้ใช้ในการสืบค้นได้

จากการศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุดดังกล่าวข้างต้น สามารถสรุปได้ว่าระบบ OPAC ของแต่ละโปรแกรมมีฟังก์ชันการทำงานและมีความสามารถในการใช้งานที่แตกต่างกัน แต่เป็นเครื่องมือที่จะช่วยให้ผู้ใช้สามารถสืบค้นทรัพยากรสารสนเทศของห้องสมุดได้ในการสืบค้นทรัพยากรสารสนเทศของห้องสมุดผู้ใช้ส่วนใหญ่เริ่มต้นการสืบค้นโดยใช้ทางเลือก คำสำคัญ ในขณะที่ห้องสมุดก็มีคำที่กำหนดแทนเนื้อหาของทรัพยากรสารสนเทศคือหัวเรื่อง ดังนั้น จากการศึกษาวิจัยจะพบว่ามีการศึกษาแนวทางการเพิ่มประสิทธิภาพการค้นคืนสารสนเทศหลากหลายแนวทาง รวมทั้งมีการเสนอวิธีการเพิ่มประสิทธิภาพการสืบค้นด้วยหัวเรื่องโดยใช้ออนโทโลยีที่สร้างจากหัวเรื่องหอสมุดรัฐสภาอเมริกัน (Library of Congress Subject Heading: LCSH) กำหนดความสัมพันธ์ระหว่างแนวคิดของ LCSH และแสดงหัวเรื่องในรูปแบบของคลาส ความสัมพันธ์แบบลำดับชั้นย่อย ประกอบด้วยคำที่กว้างกว่า (BT) คำที่แคบกว่า (NT) คำที่เกี่ยวข้อง (RT) ซึ่งจะช่วยให้การขยายคำในการค้นหาที่ผู้ใช้สามารถเลือกค้นเพิ่มเติมจากคำสำคัญที่ใช้ในการค้นได้ อย่างไรก็ตามปัญหาที่สำคัญของการเพิ่มประสิทธิภาพของการสืบค้นด้วยคำสำคัญร่วมกับหัวเรื่อง คือการจับคู่ (match) ระหว่างคำสำคัญที่ผู้ใช้กำหนดเป็นคำสืบค้นกับหัวเรื่องที่เป็นตัวแทนเนื้อหาของทรัพยากรสารสนเทศที่บรรณารักษ์หรือนักสารสนเทศกำหนดขึ้น งานวิจัยนี้จึงนำเสนอวิธีการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้การฝังคำในการขยายคำค้นหรือหัวเรื่องที่เกี่ยวข้องกับคำสำคัญที่ผู้ใช้กำหนดในการสืบค้น

### 2. ผลการพัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง

ผู้วิจัยได้พัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้แบบจำลอง Wrod2Vec ในการขยายคำค้นหรือหัวเรื่องที่เกี่ยวข้องกับคำสำคัญที่ผู้ใช้กำหนดในการสืบค้น ซึ่งผู้วิจัยได้ทดลองกับรายการทรัพยากรสารสนเทศที่สืบค้นโดยใช้คำว่า “database system” ด้วยทางเลือกคำสำคัญจาก WebOPAC ของสำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม และวัดประสิทธิภาพ

ผลลัพธ์การสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec โดยใช้ค่า P@k และ AP@k ผลการทดลองนำเสนอได้ดังนี้

1) การหาหัวเรื่องที่สอดคล้องกับคำสำคัญ (Mapping to Subject Heading)

ผลการหาหัวเรื่องที่สอดคล้องกับคำสำคัญซึ่งเป็นคำค้น “database system” โดยใช้แบบจำลอง Word2Vec ในการสร้างการฝังคำและค่าความคล้ายคลึงโคซายน์ในการหาหัวเรื่องของหอสมุดรัฐสภาอเมริกัน (LCSH) สามารถแสดงค่าความคล้ายคลึงโคซายน์ระหว่างคำสำคัญ “database system” และหัวเรื่องของหอสมุดรัฐสภาอเมริกัน ได้ดังตาราง 2

ตาราง 2 ความคล้ายคลึงโคซายน์ระหว่างคำสำคัญ “database system” และหัวเรื่องของหอสมุดรัฐสภาอเมริกัน

หัวเรื่องของหอสมุดรัฐสภาอเมริกัน (LCSH)	ค่าความคล้ายคลึงโคซายน์
Database management	0.8573
Database security	0.8258
Database management in libraries	0.8063
Database design	0.7936
Database industry	0.7934
Database selection	0.7805
Database searching	0.7655
Web databases	0.7647
Database marketing	0.7471
Portable databases	0.7416
Distributed databases	0.7402
Cataloging of databases	0.7242
Online databases	0.7177
HIBROWSE (Information retrieval system)	0.7068
ICONDA (Information retrieval system)	0.7068
Acquisition of databases	0.7024
Constraint databases	0.7007
Probabilistic databases	0.7004
Temporal databases	0.6898
Object-oriented databases	0.6822
FRED (Information retrieval system)	0.6813
Deductive databases	0.6738
Multidimensional databases	0.6726
Graph databases	0.6721

ผลจากการหาหัวเรื่องที่สอดคล้องกับคำสำคัญผู้วิจัยเลือกใช้หัวเรื่องที่มีค่าความคล้ายคลึงโคซายน์มากที่สุดกับคำค้น “database system” คือคำว่า “database management” ในการสืบค้นโดยใช้หัวเรื่อง และมีการทดลองเปรียบเทียบการสืบค้นด้วยคำสำคัญและหัวเรื่องด้วย Word2Vec โดยแสดงผลการทดลองได้ดังข้อ 2

2) ผลการเปรียบเทียบการจัดลำดับผลลัพธ์การสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec แสดงดังตาราง 3

ตาราง 3 เปรียบเทียบการจัดลำดับผลลัพธ์การสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec

รายการทรัพยากรสารสนเทศ	การจัดลำดับผลลัพธ์การสืบค้น		
	โดยผู้ใช้	คำสำคัญ	Word2Vec
1. Database systems: management and design	1	20	<b>1</b>
2. Database system concepts (2011)	2	1	<b>2</b>
3. Database system concepts (1997)	3	2	<b>3</b>
4. Database system concepts (1991)	4	3	<b>4</b>
5. Designing relational database systems	5	4	8
6. Database management systems: designing and building business applications (2002)	6	15	<b>6</b>
7. Database management systems: designing and building business applications (1999)	7	18	5
8. Development of GIS based traffic accident database through trauma management system: a case study of Khon Kaen, Thailand	ไม่เกี่ยวข้อง	5	16
9. IT development of soil database and land use information system [electronic resource]	ไม่เกี่ยวข้อง	6	13
10. Accounting information systems	ไม่เกี่ยวข้อง	7	9
11. Bioinformatics: high performance parallel computer architectures	ไม่เกี่ยวข้อง	8	21
12. Principles of accounting information systems	ไม่เกี่ยวข้อง	9	14
13. SQL, the complete reference	ไม่เกี่ยวข้อง	10	18
14. The Role of intellectual property rights in biotechnology innovation	ไม่เกี่ยวข้อง	11	22
15. Encyclopedia of geographic information science	ไม่เกี่ยวข้อง	12	20

รายการทรัพยากรสารสนเทศ	การจัดลำดับผลลัพธ์การสืบค้น		
	โดยผู้ใช้	คำสำคัญ	Word2Vec
16. Contemporary issues in database design and information systems development	ไม่เกี่ยวข้อง	13	10
17. SQL bible	ไม่เกี่ยวข้อง	14	24
18. Subject analysis in online catalogs	ไม่เกี่ยวข้อง	16	19
19. Databases: design, development, and deployment: using microsoft access	ไม่เกี่ยวข้อง	17	12
20. Exploring poetry: user's manual	ไม่เกี่ยวข้อง	19	23
21. Enterprise database connectivity: the key to enterprise applications on the desktop	ไม่เกี่ยวข้อง	21	11
22. The database book	ไม่เกี่ยวข้อง	22	15
23. Data processing in the UNIX environment: with Informix-SQL, Embedded-SQL, C-ISAM, and Turbo	ไม่เกี่ยวข้อง	23	17
24. UNIX relational database management application development in the UNIX environment	ไม่เกี่ยวข้อง	24	7

จากตาราง 3 ผลการเปรียบเทียบการจัดลำดับผลลัพธ์การสืบค้นโดยผู้ใช้งานกับลำดับผลลัพธ์การสืบค้นในการใช้ทางเลือกคำสำคัญผ่าน WebOPAC พบว่า ลำดับผลลัพธ์การสืบค้นในการใช้ทางเลือกคำสำคัญผ่าน WebOPAC ตรงกับความต้องการของผู้ใช้และถูกจัดลำดับผลลัพธ์ใน 7 ลำดับแรก จำนวน 4 รายการ ซึ่งได้แก่ รายการทรัพยากรสารสนเทศรายการที่ 2-5 ในขณะที่รายการที่ถูกจัดลำดับผลลัพธ์โดยผู้ใช้งานในลำดับที่ 1 กลับไปปรากฏอยู่ในลำดับที่ 20 ของลำดับผลลัพธ์จากการสืบค้นโดยใช้ทางเลือกคำสำคัญ ซึ่งผลลัพธ์ส่วนใหญ่ไม่เกี่ยวข้องกับความต้องการของผู้ใช้ สำหรับการเปรียบเทียบการจัดลำดับผลลัพธ์การสืบค้นด้วย Word2Vec กับการจัดลำดับผลลัพธ์การสืบค้นโดยผู้ใช้งาน พบว่า การจัดลำดับผลลัพธ์การสืบค้น Word2Vec ที่ได้ มีรายการที่ตรงกับลำดับผลลัพธ์การสืบค้นโดยผู้ใช้งาน จำนวน 5 รายการ คือ รายการทรัพยากรสารสนเทศรายการที่ 1-4 และรายการที่ 6 นอกจากนี้เมื่อพิจารณารายการในลำดับที่ 1-7 พบว่าการสืบค้นด้วย Word2Vec ให้ผลลัพธ์การสืบค้น จำนวน 6 รายการ ที่มีความเกี่ยวข้องและตรงกับความต้องการของผู้ใช้ ซึ่งมีจำนวนมากกว่าการสืบค้นในการใช้คำสำคัญ

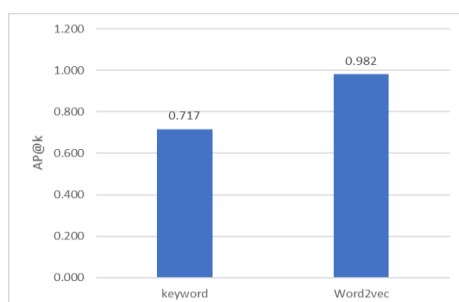
3) การวัดประสิทธิภาพของการสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec ด้วยค่า P@k และค่า AP@k แสดงดังตาราง 4 และภาพ 1

ตาราง 4 ค่า P@K ของการสืบค้นด้วยคำสำคัญ และการสืบค้นหัวเรื่องด้วย Word2Vec

อันดับที่ (K)	คำสำคัญ		Word2Vec	
	R/N	P@K	R/N	P@K
1	R	1.000	R	1.000
2	R	1.000	R	1.000
3	R	1.000	R	1.000
4	R	1.000	R	1.000
5	N	0.800	R	1.000
6	N	0.667	R	1.000
7	N	0.571	N	0.857
8	N	0.500	R	0.875
9	N	0.444	N	0.778
10	N	0.400	N	0.700
11	N	0.364	N	0.636
12	N	0.333	N	0.583
13	N	0.308	N	0.538
14	N	0.286	N	0.500
15	R	0.333	N	0.467
16	N	0.313	N	0.438
17	N	0.294	N	0.412
18	R	0.333	N	0.389
19	N	0.316	N	0.368
20	R	0.350	N	0.350
21	N	0.333	N	0.333
22	N	0.318	N	0.318
23	N	0.304	N	0.304
24	N	0.292	N	0.292

หมายเหตุ: R หมายถึงเกี่ยวข้องกับความต้องการของผู้ใช้ และ N หมายถึงไม่เกี่ยวข้องกับความต้องการของผู้ใช้

จากตาราง 4 เมื่อพิจารณาค่า จากค่า P@K พบว่าการสืบค้นหัวเรื่องด้วย Word2Vec ผลลัพธ์ที่ได้จะให้ค่าความเที่ยงตรงที่อันดับ k สูงกว่าการสืบค้นด้วยคำสำคัญ



ภาพ 1 ค่า AP@k ของการสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec

นอกจากนั้นเมื่อพิจารณาค่า AP@k ของการสืบค้นด้วยคำสำคัญ และหัวเรื่องด้วย Word2Vec จากกราฟในภาพ 1 พบว่าการสืบค้นโดยใช้หัวเรื่องด้วย Word2Vec ให้ค่า AP@k สูงสุด คือ 0.982 ซึ่งมากกว่าการสืบค้นโดยใช้คำสำคัญ แสดงให้เห็นว่าการสืบค้นโดยใช้คำสำคัญร่วมกับหัวเรื่องด้วย Word2Vec ให้ประสิทธิภาพสูงสุด

จากผลการทดลอง สามารถสรุปแนวทางการเพิ่มประสิทธิภาพการสืบค้นด้วยคำสำคัญร่วมกับหัวเรื่องได้ดังนี้ เริ่มจากการนำคำสำคัญที่ผู้ใช้ระบุในการสืบค้นมาขยายคำค้นด้วยหัวเรื่องที่เกี่ยวข้องกับคำสำคัญ ซึ่งการขยายคำค้นด้วยหัวเรื่องสามารถทำได้โดยใช้แบบจำลอง Word2Vec ในการสร้างการฝังคำและค่าความคล้ายคลึง โคชายนีในการหาหัวเรื่องของหัวเรื่องหอสมุดรัฐสภาอเมริกัน (LCSH) แล้วนำหัวเรื่องที่ได้ไปสืบค้นร่วมกับคำสำคัญ ซึ่งพบว่าผลลัพธ์จากการสืบค้นมีความเกี่ยวข้องหรือตรงกับความต้องการของผู้ใช้มากกว่าการสืบค้นโดยใช้คำสำคัญเพียงอย่างเดียว

### สรุปและอภิปรายผล

จากการศึกษาสภาพการใช้คำสำคัญและหัวเรื่องในการสืบค้นทรัพยากรสารสนเทศของห้องสมุด และแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง สามารถสรุปได้ว่าคำสำคัญและหัวเรื่องต่างก็มีความสำคัญ แต่ผู้ใช้เลือกสืบค้นทรัพยากรสารสนเทศจากทางเลือกคำสำคัญมากที่สุด เนื่องจากคำสำคัญหรือศัพท์อิสระนั้นมีลักษณะใกล้เคียงกับภาษาธรรมชาติมากที่สุด (Cabanac et al., 2016) ในขณะที่หัวเรื่องก็มีความสำคัญกับห้องสมุด นอกจากช่วยในการจัดหมวดหมู่ทรัพยากรสารสนเทศยังช่วยในการสืบค้นทรัพยากรสารสนเทศด้วย แต่เนื่องจากคำที่ห้องสมุดกำหนดเป็นหัวเรื่อง ถูกกำหนดโดยบรรณารักษ์และผู้เชี่ยวชาญในแต่ละสาขา ผู้ใช้จึงไม่คุ้นเคยและอาจจะไม่รู้จัก และพบว่ามีคำศัพท์บางคำไม่ได้ถูกกำหนดให้ใช้เป็นหัวเรื่อง ได้แก่ คำว่า การจัดการความรู้ แอปพลิเคชัน website ดังนั้น ผู้ใช้สามารถสืบค้นและได้ผลลัพธ์จากการสืบค้นโดยใช้ทางเลือกคำสำคัญ ซึ่งสอดคล้องกับงานวิจัยของ Halder (2021) ที่เปรียบเทียบการสืบค้นระหว่างคำสำคัญและหัวเรื่องพบว่า การสืบค้นโดยใช้หัวเรื่องมีโอกาสที่ไม่ได้ผลลัพธ์การสืบค้นเช่นเดียวกัน ส่วนระบบห้องสมุดอัตโนมัติมีการพัฒนามากขึ้น เช่นเดียวกับระบบห้องสมุดอัตโนมัติ Walai Autolib ที่มีการเพิ่มทางเลือกในการสืบค้นให้กับผู้ใช้ ช่วยให้ผู้ใช้สามารถสืบค้นได้ง่ายขึ้น ระบบมีการ

พัฒนาขึ้นมาเพื่อรองรับการทำงานของ web 2.0 ที่อนุญาตให้ผู้ใช้งานมีส่วนร่วมในการสร้างแท็ก (Tags) ให้กับรายการบรรณานุกรมและผู้ใช้สามารถสืบค้นผ่านแท็ก (Tags) ได้ ซึ่งแท็กเป็นคำที่ผู้ใช้งานกำหนดขึ้นมาอย่างอิสระเป็นคำศัพท์ใหม่ ๆ ที่มีความทันสมัย อีกทั้ง ในปัจจุบันมีเครื่องมือช่วยค้นที่คนนิยมใช้ คือ Google ที่ผู้ใช้สามารถกำหนดคำค้นหรือคำศัพท์อิสระ (Free term) ในการสืบค้นได้อย่างสะดวกรวดเร็ว ทำให้ผู้ใช้เกิดความคุ้นชินในการใช้คำสำคัญหรือคำศัพท์อิสระในการสืบค้นสารสนเทศ ดังนั้น จึงอาจเป็นเหตุผลหนึ่งที่ส่งผลให้ผู้ใช้ในปัจจุบันส่วนใหญ่รู้จักและเลือกใช้คำสำคัญในการสืบค้นมากกว่าการใช้หัวเรื่องหรือชื่อเรื่อง

นอกจากนั้น อาจกล่าวได้ว่า การพัฒนาเครื่องมือสำหรับช่วยผู้ใช้ในการสืบค้นในรูปแบบต่างๆ ยังคงมีความสำคัญและมีการพัฒนาอยู่เสมอ ทั้งในส่วนของพัฒนาโปรแกรมช่วยสืบค้น การนำเทคโนโลยีใหม่ ๆ มาประยุกต์ใช้ รวมทั้งการนำเทคนิคหรือแนวทางต่าง ๆ เข้ามาช่วยเพิ่มประสิทธิภาพการค้นคืนสารสนเทศ เช่นเดียวกับงานวิจัยนี้ได้พัฒนาแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้แบบจำลอง Word2vec โดยผลการสืบค้นในการใช้ทางเลือกคำสำคัญจาก WebOPAC และการสืบค้นโดยใช้หัวเรื่องด้วย Word2Vec พบว่าการสืบค้นโดยใช้หัวเรื่องด้วย Word2Vec ให้ค่า AP@k เท่ากับ 0.982 ซึ่งมากกว่าการสืบค้นโดยใช้คำสำคัญ แสดงให้เห็นว่าการสืบค้นโดยใช้คำสำคัญร่วมกับหัวเรื่องด้วย Word2Vec ให้ประสิทธิภาพสูงสุดในการสืบค้นทรัพยากรสารสนเทศ ซึ่งสอดคล้องกับงานวิจัยของ Aklouche, Bounhas & Slimani (2018) ที่ขยายคำค้นโดยใช้การประมวลผลภาษาธรรมชาติร่วมกับแบบจำลอง Word2vec ในการฝังคำ พบว่าช่วยเพิ่มประสิทธิภาพในการสืบค้นที่ดีขึ้น ดังนั้น จึงกล่าวได้ว่าแนวทางการเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศด้วยคำสำคัญร่วมกับหัวเรื่อง โดยใช้แบบจำลอง Word2vec ในการสร้างการฝังคำถือเป็นอีกแนวทางที่ช่วยเพิ่มประสิทธิภาพการสืบค้นทรัพยากรสารสนเทศ

### ข้อเสนอแนะ

งานวิจัยนี้สามารถนำไปประยุกต์ใช้กับการค้นคืนทรัพยากรสารสนเทศในห้องสมุดได้ โดยนำไปใช้ในการพัฒนาเครื่องมือช่วยค้น ซึ่งจะช่วยเพิ่มประสิทธิภาพในการสืบค้นทรัพยากรสารสนเทศ โดยเริ่มกำหนดคำค้นจาก คำสำคัญที่ผู้ใช้ส่วนใหญ่เลือกใช้ มีระบบที่ช่วยเปรียบเทียบค่าความคล้ายคลึงระหว่างคำสำคัญและหัวเรื่องที่ห้องสมุดกำหนด และระบบสามารถเสนอแนะหัวเรื่องที่มีความคล้ายคลึงกับคำสำคัญที่ผู้ใช้กำหนดขึ้นมา ซึ่งถือเป็นแนวทางที่จะช่วยให้ผู้ใช้สืบค้นทรัพยากรสารสนเทศได้ตรงตามความต้องการมากยิ่งขึ้น

### เอกสารอ้างอิง

- Aklouche, B., Bounhas, I., & Slimani, Y. (2018). *Query Expansion Based on NLP and Word Embeddings* (pp.1-7). The 27th Text REtrieval Conference (TREC 2018). Gaithersburg, Maryland, USA.
- Baeza-Yates, R. & Ribeiro-Neto, B. (2011). *Modern information retrieval the concepts and technology behind search* (2<sup>nd</sup> ed.). Harlow: Pearson.
- Buettcher, S., Clarke, C.L.A. & Cormack, G.V. (2010). *Information retrieval: implementing and evaluation search engines*. Cambridge: MIT Press.



- Cabanac, G, et al., (2016). *Joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL 2016)*(pp.299-300). 2016 IEEE/ACM JointConference on Digital Libraries (JCDL).
- Cheysiri, P. (2000). *Student Usage of Online Public Access Catalog Service : A Study at Ramkhamhaeng University* (Master's thesis). Ramkhamhaeng University, Thailand. [In Thai]
- Halder, S. N. (2021). Empirical Study on Subject-Based Information Retrieval with Special Reference to Bengali Documents. *Journal of Library and Information Science*, 15(2), 130-139.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Amsterdam: Morgan Kautmann.
- Informatic Innovation Center of Excellence. Walailak University. (2019). *OPAC&Utility*. Retrieved from <https://www.lib.buu.ac.th/KM/wp-content/uploads/2020/06/User-Doc-OPAC-Manual-Ultimate-1-May-2019.pdf> [In Thai]
- [Information Technology for Works] เทคโนโลยีสารสนเทศเพื่อการจัดการอาชีพ. (2017). Retrieved from <http://avangerthefor.blogspot.com/2017/06/33.html> [In Thai]
- Intraprapan, C. (2014). *Optimization of Information Retrieval Using ThaiWordNet*. Retrieved from <https://research.kpru.ac.th/sac/fileconference/7672018-05-11.pdf> [In Thai]
- Intraprapan, C. & Kesorn, K. (2014). Ontology-based Technique for Cross-Language (Thai-English) Information Retrieval. *Information Technology Journal*, 10(1), 21-30. [In Thai]
- Jutopama, S. (2007). *Using Service on Information Retrieval Through OPAC System of Buriram Rajabhat University Students*. Retrieved from <http://dspace.bru.ac.th/xmlui/bitstream/handle/123456789/3284/2549-research-Siranee.J.pdf?sequence=2&isAllowed=y> [In Thai]
- Kruaysawat, N. (2016). *Integrating folksonomy concept in building subject heading for information resources in the fields of Business administration and economics*. (Doctoral dissertation). Khon Kaen University, Thailand. [In Thai]
- Kruaysawat, N., Manmart, L., & Kabmala, M. (2018). Users's Behavior In Accessing and Using Key Words to Retrive Business and Economic Information. *Journal of Information Science*, 36(1), 69–108. [In Thai]
- Lertmahakiat, W. & Mingkwan, A. (2010). The Innovation of Multiple Relations Information Retrieval. *The Journal of KMUTNB*, 20(3), 514-523. [In Thai]
- Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (pp.1-12). Proceedings of the International Conference on Learning Representations.

- Nimsomboon, N. (2018). การกำหนดคำสำคัญ [Using keyword]. Retrieved from <http://203.131.219.167/km2559/2018/06/12/การกำหนดคำสำคัญ-keyword/> [In Thai]
- Ouantrai, S. (2008). *The Use of Web Online Public Access Catalog of Undergraduate Students, Khon Kaen University* (Master's thesis). Khon Kaen University, Thailand. [In Thai]
- Panommit, P., Arayaphan, W., & Julrode, P. (2017). Retrieving and Accessing to the Knowledge of Lanna Textiles from Information Resources in the Upper Northern Thailand. *Veridian E-Journal*, Silapakorn, 10(3), 1545-1559. [In Thai]
- Shera, J.H. & Egan, M. E. (1956). *The classified catalog : basic principles and practices*. Chicago: American Library Association.
- Sriwantan, P. & Arayaphan, W. (2022). Storage and Retrieval of Lanna Information Resources at the Institutes of Information Services in the Upper North of Thailand. *Journal of LiberalArts Maejo University*, 10(1), 42-66. [In Thai]
- TaKham, T., Kabmala, M., & Manmart, L. (2016). Scholar and Researchers' Behavior in Retrieving. *Journal of Information Science*, 34(1), 54-93. [In Thai]
- Yu, W. & Chen, J.(2020). Enriching the library subject headings with folksonomy. *The Electronic Library*, 38(2), 297-315.