

การพัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

Development of Keyword Suggestion for Thai Document Search

ปกป้อง ส่องเมือง^{1*}, รัฐภูมิ เกียรติวิระศักดิ์², รัชฎา คงคะจันทร์³,

ศรัณย์ กุลยานนท์⁴ และ วสิศ ลิ้มประเสริฐ⁵

Pokpong Songmuang^{1*}, Rattapoom Kedtiwerasak²,

Rachada Kongkachandra³, Sarun Gulyanon⁴ and Wasit Limprasert⁵

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์^{1*,2}

และวิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์^{3,4,5}

Faculty of Science and Technology, Thammasat University^{1*,2} and

College of Interdisciplinary Studies, Thammasat University^{3,4,5}

E-Mail: pokpongs@tu.ac.th

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย 2) เปรียบเทียบประสิทธิภาพของพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย และ 3) ศึกษาผลการทดลองใช้พีเจอร์การแนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย โดยกลุ่มเป้าหมายในการวิจัยครั้งนี้ คือ บุคคลที่มีพื้นฐานในการใช้ระบบค้นหา จำนวน 21 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ 1) ระบบค้นหา 3 แบบ ได้แก่ ระบบค้นหาที่ไม่มีพีเจอร์แนะนำคำสำคัญ, ระบบค้นหาที่มีพีเจอร์แนะนำคำสำคัญจากพจนานุกรม และ ระบบค้นหาที่มีพีเจอร์แนะนำคำสำคัญจากออนโทโลยี 2) แบบสัมภาษณ์พื้นหลังของผู้ใช้ระบบ สถิติที่ใช้ในการวิจัย คือ ค่าเฉลี่ย

ผลการวิจัยพบว่า 1) ผลการพัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย ประกอบด้วย 2 ส่วน ได้แก่ แบบสัมภาษณ์เพื่อผู้พื้นหลังของผู้ทดลอง และแบบทดลองค้นหา 2) ผลการเปรียบเทียบประสิทธิภาพของการแนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย พบว่า ประสิทธิภาพการค้นหาเอกสารของผู้ใช้ได้ใช้ค่าที่แนะนำจากระบบที่มีการแนะนำคำสำคัญจากออนโทโลยีมีประสิทธิภาพสูงสุดจากระบบทั้งหมดที่ใช้ทดสอบ และ 3) ผลการทดลองใช้พีเจอร์การแนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย พบว่า ผู้ใช้ค้นหาเอกสารได้ตรงตามที่ต้องการมากขึ้นจากระบบค้นหาที่ได้รับคำแนะนำจากพีเจอร์แนะนำคำสำคัญจากออนโทโลยี

คำสำคัญ: ระบบค้นหา, ออนโทโลยี, พจนานุกรม, การแนะนำคำสำคัญ

Abstract

The purposes of the research were: 1) to develop keyword suggestion feature for searching Thai document based on dictionary and ontology keyword suggestion. 2) to compare the effectiveness of keyword suggestions for searching Thai documents. 3) to study keyword suggestion feature result for searching Thai document from users in a real trial. The target audience for this research was 21 people who had experience from search engine. The research tools were: 1) 3 types of search engines: search without keyword suggestions, search with a dictionary suggestion,

and search with an ontology suggestion. 2) User background interview form. The statistic used in the research is an average.

Research results were: 1) The result of the search engine development consisted of 2 parts; user background interview form and searching test form 2) The result of the performance comparison of keyword suggestions for Thai document search was found the performance of search with the ontology keyword suggestion that was the most efficient across all systems used for testing. 3) The result of the search engine testing from users in a real trial was found users can find exactly documents that they want with the search engine using the ontology keyword suggestion

Keywords: Search Engine, Ontology, Dictionary, Keyword Suggestion

บทนำ

ในปัจจุบัน ข้อมูลที่ถูกจัดเก็บอยู่บนอินเทอร์เน็ตนั้นมีปริมาณมากและมีเพิ่มขึ้นตลอดเวลา [1] ซึ่งทำให้การหาความรู้จากอินเทอร์เน็ตสามารถเข้าถึงข้อมูลที่หลากหลายและเป็นประโยชน์ได้มากขึ้น แต่การได้มาซึ่งความรู้จากอินเทอร์เน็ต ผู้ใช้จำเป็นต้องค้นหาข้อมูลผ่านระบบค้นหา (Search Engine) ซึ่งระบบค้นหาที่มีความนิยมในปัจจุบันมีอยู่หลายระบบ เช่น www.google.com, www.bing.com, www.yahoo.com, www.ask.com

เมื่อผู้ใช้ระบบค้นหาใส่คำค้นเข้าสู่ระบบค้นหา คำค้นจะถูกนำไปเปรียบเทียบกับ index ของข้อมูลที่อยู่ในฐานข้อมูล [1],[2] ซึ่งการเปรียบเทียบส่วนใหญ่เป็นการเปรียบเทียบตามตัวอักษร (String Matching) หลังจากนั้นระบบค้นหา จัดลำดับผลลัพธ์การค้นหาแตกต่างกันตามอัลกอริทึมของแต่ละระบบ ส่วนมากระบบค้นหา เรียงลำดับผลลัพธ์ตามความสัมพันธ์ของหน้าเว็บ (Web Page) กับคำค้น ซึ่งอาจมีปัจจัยอื่นถูกนำมาเป็นปัจจัยร่วมในการจัดลำดับ เช่น ประเทศของผู้ใช้ ภาษาที่ใช้ค้น ขนาดของไฟล์ จำนวนผู้เข้าชมของหน้าเว็บ ความถี่ในการอัปเดตข้อมูล หรือจำนวนลิงก์ที่เชื่อมโยงกับหน้าเว็บ บางครั้งผู้ใช้อาจต้องการค้นหาข้อมูลเฉพาะทาง แต่ผู้ที่ไม่มีความรู้ในข้อมูลเฉพาะทางนั้น ผู้ใช้อาจทราบเพียงบริบทของข้อมูลเฉพาะทาง เช่น ผู้ใช้ต้องการค้นหาข้อมูลของประเพณีสงกรานต์ แต่ผู้ใช้ทราบเพียงบริบทของประเพณี เช่น เล่นน้ำ ทำบุญ และวันหยุดยาว จึงใช้คำค้นที่ไม่มีความสัมพันธ์หรือมีความสัมพันธ์น้อยกับหน้าเว็บที่เกี่ยวข้องกับข้อมูลเฉพาะทาง ระบบค้นหาจึงไม่แสดงผลที่ต้องการ จากปัญหาข้างต้นได้มีหลายงานวิจัยศึกษาวิธีการแก้ปัญหาของการใช้คำค้นที่ไม่สัมพันธ์กับข้อมูลเฉพาะทาง

ผู้วิจัยได้พัฒนาระบบค้นหาและแนะนำคำสำคัญโดยใช้เทคนิคการแนะนำคำสำคัญจากพจนานุกรมและออนโทโลยี และเปรียบเทียบประสิทธิภาพของการแนะนำคำสำคัญทั้ง 2 วิธีข้างต้น

1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อพัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย
- 1.2 เพื่อเปรียบเทียบประสิทธิภาพของพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย
- 1.3 เพื่อศึกษาผลการทดลองพีเจอร์การแนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

Bharathi, G. and Venkatesan, D. [3] ศึกษาการขยายคำค้นด้วยคำพ้องความหมาย (Synonym) ร่วมกับ การจัดกลุ่มเอกสาร (clustering) ด้วยเทคนิค K-mean ผลการทดลองแสดงให้เห็นว่าการขยายคำค้นด้วยคำพ้องความหมายร่วมกับวิธีจัดกลุ่มเอกสารนั้นเพิ่มประสิทธิภาพของระบบการสืบค้นได้ดีกว่าการใช้วิธีจัดกลุ่มเอกสารเพียงอย่างเดียว ในแง่ของการเพิ่มขึ้นของผลลัพธ์การสืบค้นแต่อาจไม่ตรงกับความต้องการของผู้ใช้เนื่องจากผู้วิจัยต้องการเพียงเพิ่มปริมาณผลลัพธ์แต่ไม่ได้วัดความถูกต้องของผลลัพธ์

Gan, L. and Tu, W. [4] ศึกษาเกี่ยวกับการขยายแบบสอบถาม (Query Expansion) เพื่อเพิ่มประสิทธิภาพในการค้นหาข้อมูลและลดความผิดพลาดในการค้นหา โดยในการทดลอง Markov semantic network ถูกนำมาสร้างเครือข่าย Markov 2 รูปแบบ คือ W-Markov network ที่สร้างจากคลังข้อมูลของ Wikipedia และ B-Markov network ที่สร้างจากคลังข้อมูลส่วนตัว แล้วนำทั้ง 2 เครือข่ายมาซ้อนทับกันเพื่อให้ได้ C-Markov network ที่เป็นเครือข่ายที่ใหญ่ขึ้น ซึ่งเครือข่ายที่ใหญ่ขึ้นนี้ทำให้ได้ความสัมพันธ์ที่มากขึ้นและสมบูรณ์ยิ่งขึ้นสำหรับคำที่ไม่เชื่อมโยงรวมถึงคำทั่วไปด้วย วิธีการนี้เป็นประโยชน์ในการหาความสัมพันธ์ของคำและปรับปรุงประสิทธิภาพของการค้นหาข้อมูล แต่งานวิจัยนี้ไม่มีการวัดประสิทธิภาพของระบบค้นหา

Toman, S., Abed, M., & Toman, Z.H. [5] ศึกษาการปรับปรุงผลลัพธ์ของเอกสารจากการค้นคืนเพื่อลดเอกสารที่ไม่เกี่ยวข้องจากการค้นหาของผู้ใช้ ในบทความนี้ผู้วิจัยได้เสนอเทคนิค (K-mean) - hierarchical parallel genetic algorithms approach (HPGA) โดยการค้นคืนข้อมูลแบบคลัสเตอร์เป็นหนึ่งในเครื่องมือการค้นคืนที่จัดหมวดหมู่เว็บไซต์ตามความคล้ายคลึงกัน ผลการทดลองพบว่า (K-mean) -HPGA ได้รับความแม่นยำที่สูงขึ้นและคุณภาพที่ดีขึ้นในการดึงเอกสาร นอกจากนี้ยังพบการลดลงของผลลัพธ์ที่ไม่เกี่ยวข้องในการค้นหาของผู้ใช้

N. Lin, V. A. Kudinov, H. M. Zaw and S. Naing [6] ศึกษาวิธีการขยายคำค้นหาของระบบการค้นคืน Myanmar information retrieval (MIR) ซึ่งเขียนเป็นภาษาพม่าโดยใช้ WordNet เพื่อแก้ไขปัญหาคำที่คลุมเครือทางความหมายใน MIR ผลการทดลองแสดงให้เห็นถึงการปรับปรุงการดึงข้อมูลสำหรับภาษาพม่าด้วยแสดงด้วยค่า precision 65.4% recall 78.9% และ F-measure 70.96%

Kim, Y., & Croft, W. B. [7] ศึกษาเกี่ยวกับวิธีการปรับปรุงการค้นคืนสิทธิบัตร เนื่องจากตัวเอกสารสิทธิบัตรที่มีรูปแบบข้อมูลเฉพาะตัวคือมีคำอธิบายที่ยาวและซับซ้อนทำให้การค้นคืนแบบเดิมไม่ครอบคลุมทำให้เกิดปัญหาการจดสิทธิบัตรซ้อนซึ่งจะมีผลทำให้สิทธิบัตรที่จดใหม่นั้นเป็นโมฆะ ดังนั้นผู้วิจัยจึงปรับปรุงระบบการค้นคืนสิทธิบัตรด้วยวิธีการค้นหาแบบกระจาย โดยใช้เทคนิค DSPApprox (DSP) และ Learning-to-rank (LTR) ซึ่งผลการทดลองสรุปได้ว่าประสิทธิภาพของวิธีการ LTR นั้นดีกว่า DSP ในแง่ของการแสดงผลการค้นหาและจัดอันดับที่ใกล้เคียงกับคำสำคัญที่ใช้ค้นหาที่สุด

Rawat, R., Nayak, R., & Li, Y. [8] ศึกษาเกี่ยวกับการปรับปรุงการจัดอันดับผลการสืบค้นสำหรับฐานข้อมูลเว็บไซต์เนื่องจากบางครั้งผู้ใช้ทำการสืบค้นแล้วใส่คำค้นที่เฉพาะเจาะจงเกินไปทำให้ผลลัพธ์ออกมาน้อยและไม่เจอข้อมูลที่ต้องการหรือในบางครั้งก็ใส่คำค้นที่กว้างเกินไปทำให้ผลสืบค้นออกมามากเกินไปไม่พบสิ่งที่ต้องการ

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 ระบบค้นหา 3 แบบ ได้แก่

- 1.1.1 ระบบค้นหาที่ไม่มีการแนะนำคำสำคัญ
- 1.1.2 ระบบค้นหาที่มีการแนะนำคำสำคัญจากพจนานุกรม
- 1.1.3 ระบบค้นหาที่มีการแนะนำคำสำคัญจากออนโทโลยี

1.2 แบบสัมภาษณ์พื้นหลังและประสบการณ์การใช้งานของผู้ใช้ระบบ

2. กลุ่มเป้าหมาย

กลุ่มเป้าหมายในการวิจัยครั้งนี้ถูกเลือกโดยการสุ่มกลุ่มตัวอย่างจากประชากรผู้ใช้ทั่วไปจำนวน 21 คน ในเดือนกรกฎาคม 2563

3. ขั้นตอนการดำเนินการวิจัย

3.1 การออกแบบและพัฒนาระบบค้นหาข้อมูล

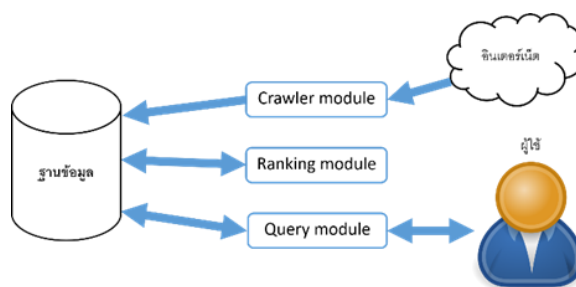
ในงานวิจัยครั้งนี้ได้รวบรวมเว็บไซต์จากกระทรวงวัฒนธรรมภายใต้โดเมน culture.go.th มีจำนวนหน้าเว็บทั้งสิ้น 103,126 หน้า และมีวิธีการแนะนำคำสำคัญจากการใช้พจนานุกรม และออนโทโลยี เพื่อแนะนำคำสำคัญสำหรับใช้ค้นหา ในระบบแนะนำคำสำคัญมีออนโทโลยีทั้งหมด 70 ชุด โดยหนึ่งออนโทโลยีจะอธิบายองค์ความรู้ของหนึ่งวัฒนธรรม และคำในพจนานุกรมทั้งหมด 40,226 คำ

โดยงานวิจัยนี้ผู้วิจัยได้ทำการออกแบบและพัฒนาระบบค้นหาข้อมูลซึ่งจะอธิบายตามลำดับดังนี้

3.1.1 องค์ประกอบของระบบค้นหาข้อมูล

3.1.1.1 ระบบค้นหาข้อมูลบนเว็บไซต์ได้เว็บ

การค้นคืนเว็บไซต์ (Web Information Retrieval) [11], [12] ใช้การเก็บข้อมูลสำหรับค้นหาเว็บไซต์ซึ่งแสดงดังภาพที่ 2 และมีส่วนประกอบดังนี้



ภาพที่ 2 ภาพรวมของระบบค้นหา

จากภาพที่ 1 พบว่า ภาพรวมของระบบค้นหา ดังนี้

1) Crawler Module: ทำหน้าที่ในการเก็บข้อมูลจากเว็บไซต์ โดยจะสร้างโปรแกรมสำหรับเก็บข้อมูลหน้าเว็บที่มีชื่อเรียกว่า ‘spider’ ซึ่งโปรแกรมนี้อ่านข้อมูลในหน้าเว็บเพื่อหาข้อมูลและลิงก์เชื่อมโยงไปหน้าเว็บใหม่ และใช้ลิงก์เชื่อมโยงเก็บข้อมูลหน้าเว็บเพิ่มเติม [1]

2) Page Repository: เป็นส่วนเก็บหน้าเว็บ ทำหน้าที่เก็บข้อมูลจากโปรแกรมที่ได้จากส่วนของ Crawler Module ซึ่งเก็บข้อมูลทั้งหมดของหน้าเว็บ และนำข้อมูลทั้งหมดไปทำการประมวลผลต่อในส่วน Indexes

3) Indexes: เก็บข้อมูลสำคัญของแต่ละหน้าเว็บ และทำส่วนอ้างอิงจากคำสำคัญกลับไปยังหน้าเว็บจริง เพื่ออ้างอิงกลับไปหาหน้าเว็บที่แท้จริงให้เร็วขึ้น โดยในส่วน Indexes จะทำงานภายในฐานข้อมูล

4) Query Module: เป็นส่วนที่แปลงคำค้นของผู้ใช้ให้เป็นรูปแบบที่เหมาะสมสำหรับการค้นหาในฐานข้อมูลและดึงหน้าเว็บที่เกี่ยวข้องขึ้นมาเป็นคำตอบ

5) Ranking Module: เป็นส่วนที่นำหน้าเว็บที่เกี่ยวข้องมาจัดเรียงใหม่ให้ตรงกับเงื่อนไขมากยิ่งขึ้น โดยผลลัพธ์ที่ออกมาจะเป็นรายการหน้าเว็บที่ถูกเรียงไว้ ซึ่งหน้าเว็บที่อยู่ด้านบนจะเป็นหน้าเว็บที่ตรงกับคำค้นหาที่ผู้ใช้ต้องการมากที่สุด

3.1.1.2 การเตรียมข้อมูลลงฐานข้อมูล

การเตรียมข้อมูลจะใช้ Crawler ในการเก็บข้อมูล และทำการตัดคำข้อมูลและคำนวณ TF-IDF (Term Frequency - Inverse Document Frequency) สำหรับค้นหาเอกสารขั้นพื้นฐานในระบบค้นหา โดยที่ TF-IDF สามารถคำนวณได้จากผลคูณของ TF และ IDF [13] การตัดคำที่ใช้ในการทดลองนี้จะใช้พื้นฐานจากการตัดคำโดยใช้ maximum matching ในการค้นหาคำที่ยาวที่สุด โดยใช้คำจากพจนานุกรมในการตัดคำ ซึ่งใช้ Thai Character Cluster (TCC) เป็นเงื่อนไขในการตัดคำที่ไม่ได้อยู่ในพจนานุกรม ในรายละเอียดเพิ่มเติม การตัดคำ maximum matching โดยจะใช้เทคนิคที่เรียกว่า TRIE [14] ในการอ่านและตัดคำให้เร็วขึ้น

ส่วนการตัดคำที่ไม่ได้มาจากพจนานุกรมจะใช้เทคนิคที่เรียกว่า TCC ในการตัดคำ โดยวิเคราะห์จากการแยกหน่วยย่อยของอักขรก่อนตัดสนใจขอบเขตของคำ ซึ่งรูปแบบของ TCC จะถูกกำหนดด้วยกฎ ซึ่งกฎที่ใช้จะเป็นการรวมรูปแบบของคำของภาษาไทยให้ถูกต้อง รวมถึงใช้รูปแบบการเรียนรู้จากคลังข้อมูล [15]

3.1.2 การแนะนำคำสำคัญ

ในงานวิจัยนี้ได้สร้างพีเจอรแนะนำคำสำคัญเพื่อนำมาช่วยในการค้นหา โดยระบบแบ่งเป็น 2 วิธี ดังนี้

3.1.2.1 การแนะนำคำสำคัญจากพจนานุกรม

การแนะนำคำสำคัญโดยใช้พจนานุกรมในงานวิจัยนี้ได้ใช้คำศัพท์และความหมายของคำตามพจนานุกรมจากสำนักงานราชบัณฑิตยสภา โดยนำคำค้นหาจากการกรอกเข้ามาของผู้ใช้งานเทียบกับความหมายของคำที่มีอยู่ในพจนานุกรม และแนะนำคำสำคัญจากคะแนนของคำที่มีค่าเป็นส่วนหนึ่งของความหมายจากคำในพจนานุกรม

การทำงานของพีเจอรแนะนำคำสำคัญเริ่มด้วยการรับคำที่จะใช้ค้นหาจากผู้ใช้งาน โดยนำคำที่ได้จากผู้ใช้งานลงในตัวแปร query_data แล้วนำคำที่ได้มาเชื่อมกันด้วย AND ไปค้นหาในคอลัมน์ meaning ผลลัพธ์ที่ได้เก็บไว้ในตัวแปร m_query ถ้าหากไม่มีผลลัพธ์จะเปลี่ยนตัวเชื่อมเป็น OR แล้วทำการค้นหาใหม่อีกครั้ง จากนั้นนำคำจากตัว

แปร query_data ไปค้นหาในคอลัมน์ word ด้วยตัวเชื่อม OR เก็บผลลัพธ์ที่ได้ไว้ใน ตัวแปร w_query นำผลลัพธ์จากตัวแปร m_query และ w_query มารวมกันแล้วเก็บไว้ในตัวแปร merge_query นำค่าจากตัวแปร merge_query ไปค้นหาคะแนน IDF ที่ได้คำนวณเก็บไว้จากฐานข้อมูลในข้อ 3.1.1.2 แล้วเก็บไว้ในตัวแปร word_result เรียงลำดับของค่าจากมากไปน้อยด้วยค่า IDF เก็บผลลัพธ์ไว้ในตัวแปร sorted_result แล้วส่งค่าคืนผู้ใช้ดังภาพที่ 3

```

Search_Dictionary(query_data)
If query_data length > 1
    m_query = Query query_data with
        "AND" operation from "meaning" in
        Database.
Else
    m_query = Query query_data from
        "meaning" in Database.
If Data from "meaning" is length == 0 AND query_data length > 1
    m_query = Query query_data with
        "OR" operation from "meaning" in
        Database.
If query_data length > 1
    w_query = Query query_data with
        "OR" operation from "word" in
        Database.
Else
    w_query = Query query_data from
        "meaning" in Database.
merge_query = merge m_query and w_query
If merge_query == 0
    Return No result"
For "word" in merge_query
    idf_query = Query "IDF" score by "word"
    from "term" in database.
    If idf_query length != 0
        word_result append array
        ["word", "meaning", "IDF"]
sorted_result = Sort word_result DESC with "IDF".
Return sorted_result
END

```

ภาพที่ 3 ชุดโค้ดของการแนะนำคำโดยใช้พจนานุกรม

3.1.2.2 การแนะนำคำสำคัญจากออนโทโลยี

การแนะนำคำสำคัญจากออนโทโลยีในงานวิจัยนี้ได้ใช้ข้อมูลจากเว็บไซต์กระทรวงวัฒนธรรม โดยเลือกประวัติของวัฒนธรรมและอาหารพื้นเมืองมาสร้างเป็นออนโทโลยี และใช้ความสัมพันธ์ของคำที่ปรากฏในประวัติเพื่อระบุว่าเป็นส่วนหนึ่งหรือเป็นสิ่งที่เกี่ยวข้องกับออนโทโลยี

การใช้ออนโทโลยีมาแนะนำเป็นคำสำคัญจะนำคำจากผู้ใช้งานที่ต้องการค้นหาหาคำที่เกี่ยวข้องจากทุกออนโทโลยี โดยมีวิธีการค้นหาคำแนะนำตามภาพที่ 4 โดยอัลกอริทึมของการค้นหาคำเพื่อมาใช้แนะนำในระบบ ตัวฟังก์ชันการทำงานรับข้อมูลขาเข้าเป็นลิสต์ของกราฟออนโทโลยี Gs และลิสต์ของคำที่จะใช้ค้นหา W เพื่อนำไปประมวลผลเพื่อให้ได้ผลลัพธ์ที่เป็นลิสต์ของคู่ลำดับที่ประกอบด้วย ชื่อของออนโทโลยี และคะแนนเฉลี่ยของค่า IDF คำนวณจากการหาคำและคำใกล้เคียงที่ปรากฏในลิสต์ W และนำค่าที่ได้ไปค้นหาค่า IDF จากฐานข้อมูลในข้อ 3.1.1.2 และนำคะแนนที่ได้มาหาค่าเฉลี่ยของออนโทโลยีเพื่อแนะนำคำสำคัญที่มีคะแนนสูงไปต่ำตามลำดับ

```

Search_Ontology(Gs, W)
Maximum_count = 0
Selected_graph = empty list to keep selected graph.
for G in Gs
    count = 0
    for w in W
        if G have w in node
            count = count + 1
    if Maximum count < count
        Selected_graph = empty list
        Maximum_count = count
    if Maximum_count == count
        append G in Selected_graph
Expanded_GW = empty list to keep tuple of graph and word in node where near w node in graph.
if length of Selected_graph > 0
    for G in Selected_graph
        Expanded_words = empty set to keep word where near w node in graph.
        for w in W
            n = node which keep word w in graph.
            parents = list of node which edge out from the node n.
            child = list of node which edge in to the node n.
            siblings = empty list to keep node which has the same parent.
            for p in parents
                c = list of child node of p node.
                append each node in c into siblings.
            append word in each node of n, parents, child, siblings into Expanded_words.
            append tuple of G and Expanded_words into Expanded_GW.
R = empty list to keep result of tuple of ontology name and average IDF score of ontology.
for G and Expanded_words in Expanded_GW
    sum = 0
    count = 0
    for epw in Expanded_words
        score = IDF score of epw word in database.
        sum = sum + score
        count = count + 1
    avg = sum / count is average IDF score.
    ontology_name = ontology name of G.
    append tuple of ontology_name and avg into R.
Return R
END

```

ภาพที่ 4 ชุดโค้ดของการแนะนำคำโดยใช้ออนโทโลยี

3.2 การทดลองใช้งานระบบค้นหาเพื่อทดสอบประสิทธิภาพ

รูปแบบการวัดประสิทธิภาพของระบบจะใช้การทดสอบจากผู้ทดลองจริง 21 คน โดยมีการทำแบบสัมภาษณ์เพื่อดูพื้นหลังของผู้ทดลอง โดยใช้คำถามหลักในแบบสัมภาษณ์ดังนี้

- 1) เคยใช้เว็บไซต์อะไรบ้างในการค้นหาข้อมูล เช่น Google.com, Bing.com, Yahoo.com และอื่น ๆ
- 2) ผู้ทดลองเคยใช้การค้นหาที่ครั้งถึงเจอข้อมูลที่ต้องการ
- 3) ผู้ทดลองเคยใช้คำที่คำถึงสามารถเจอข้อมูลที่ต้องการ

เมื่อผู้ทดลองกรอกข้อมูลครบแล้วพีเจอรแนะนำคำสำคัญสำหรับการทดลองจะแบ่งออกเป็น 3 ระบบในการทดสอบ คือ ระบบที่ไม่มีการแนะนำคำสำคัญ, พีเจอรแนะนำคำสำคัญจากพจนานุกรม และพีเจอรแนะนำคำสำคัญจากออนโทโลยี โดยแต่ละพีเจอรแนะนำคำสำคัญจะมีคำถามที่เป็นข้อความทั้งสิ้น 5 ข้อความ สำหรับให้ผู้ทดลองอ่านแล้วตอบคำถาม 2 ข้อ คือ ชื่อวัฒนธรรม และ ลิงก์ที่เลือก จากทั้ง 3 ระบบ รวมมีคำถามทั้งหมด 15 คำถาม โดยจะเวียนคำถามให้แต่ละระบบมีคำถามไม่ซ้ำกันในแต่ละผู้ทดลอง โดยมีข้อกำหนดคือ

- 1) ผู้ทดลองจะสามารถใช้หรือไม่ใช้คำแนะนำจากระบบได้
- 2) ในการค้นหา 1 ครั้ง สามารถใส่คำค้นได้ไม่เกิน 5 คำ
- 3) ในแต่ละข้อค้นหาได้ไม่เกิน 5 ครั้ง ถ้าครบ 5 ครั้งแล้วไม่พบผลลัพธ์ที่ต้องการให้ใส่คำตอบว่า “ไม่พบ”
- 4) คำค้นผู้ทดลองสามารถใส่คำอะไรก็ได้

ซึ่งผู้ทดลองจะถูกบันทึกเวลาในการทำแต่ละข้อของการทดสอบ และผู้ทดลองไม่ทราบว่าพีเจอร์แนะนำคำสำคัญที่ใช้ทำการทดลองทำงานอย่างไร

4. สถิติที่ใช้ในการวิจัย คือ ค่าเฉลี่ยของผลลัพธ์ในการค้นคืนข้อมูล [7]

ผลการวิจัย

1. พัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย โดยใช้เทคนิคพจนานุกรม และออนโทโลยีในการแนะนำคำสำคัญ

1.1 องค์ประกอบของระบบพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

ระบบพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทยประกอบด้วย 2 ส่วน คือ 1) ส่วนค้นคืนเว็บไซต์สำหรับใช้กรอกคำค้นหาจากผู้ใช้งาน และค้นคืนเว็บไซต์ที่เกี่ยวข้องกับคำค้นหา 2) ส่วนแนะนำคำสำคัญที่เกี่ยวข้องกับคำค้นหาที่ได้จากผู้ใช้งาน โดยแบ่งเป็น 3 พีเจอร์ คือ 1) ไม่ใช้การแนะนำคำสำคัญ 2) ใช้การแนะนำคำสำคัญจากพจนานุกรม และ 3) ใช้การแนะนำคำสำคัญจากออนโทโลยี ซึ่งในส่วนการแนะนำคำสำคัญสามารถเลือกพีเจอร์ได้ทีละหนึ่งพีเจอร์เท่านั้น แสดงดังภาพที่ 5

การแนะนำคำสำคัญ

☒ ไม่มีการแนะนำ
 ☐ แนะนำด้วยพจนานุกรม
 ☐ แนะนำด้วยออนโทโลยี

คำลัดคัด ผ่าป่า

ค้นหา

ผลการค้นหา

https://www.m-culture.go.th/tak/ewt_news.php?nid=476&filename=index

- คำว่า “ตานตอด” ภาษากลาง คือ ทานทอด เป็นคำประสมระหว่างคำว่า ทาน กับ ทอด ความหมายคือ “วางของให้” ก็คือการให้ทาน ซึ่งเป็นประเพณีที่แสดงถึงความเอื้ออาทรของชาวบ้านที่มีต่อคนทุกข์ยาก เป็นการทำบุญด้วยใจไม่หวังผลตอบแทน ประเพณีตานตอดมีต้นกำเนิดมาจากล้านนา ดังนั้นจะเห็นได้ว่าการทำบุญประเภทนี้จึงมีแต่เฉพาะทางภาคเหนือของประเทศไทยเท่านั้น และ ...

ภาพที่ 5 ระบบพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

1.2 ผลการพัฒนาระบบพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

1.1 ผลการออกแบบและพัฒนาระบบค้นหา พบว่าต้องเก็บข้อมูลเว็บไซต์โดย Crawler Module ซึ่งทำหน้าที่เก็บข้อมูลจากอินเทอร์เน็ตลงใน ฐานข้อมูล ซึ่งจะมี Page Repository เป็นส่วนเก็บข้อมูลที่เก็บไฟล์ต้นฉบับของหน้าเว็บและมี Index Module ทำหน้าที่เก็บข้อมูลคำสำคัญของแต่ละหน้าเว็บเอาไว้ และทำส่วนอ้างอิงจากคำสำคัญกลับไปยังหน้าเว็บ ในส่วนของระบบค้นหาจะประกอบด้วย Query Module ทำการแปลงคำค้นของผู้ใช้ให้เป็นรูปแบบที่เหมาะสมสำหรับการค้นหาในฐานข้อมูลและดึงหน้าเว็บที่เกี่ยวข้องขึ้นมาเป็นคำตอบ โดยจะมีการ

แนะนำค่าสำคัญให้ผู้ใช้งานเลือกตามฟีเจอร์ที่ระบบให้เลือก และ Ranking Module ทำหน้าที่จัดเรียงและแสดงผลให้กับผู้ใช้ รายละเอียดของแต่ละส่วน

2. ผลการเปรียบเทียบประสิทธิภาพของฟีเจอร์แนะนำค่าสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

2.1 ผลการทดลองใช้ระบบค้นหาเพื่อเปรียบเทียบประสิทธิภาพการใช้งานเมื่อมี/ไม่มีฟีเจอร์แนะนำค่าสำคัญ

ตารางที่ 1 ผลลัพธ์คำตอบของผู้ทดลอง

คำตอบ	ระบบที่ไม่มีฟีเจอร์แนะนำค่าสำคัญ	ระบบที่ใช้ฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม	ระบบที่ใช้ฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยี
คำตอบถูก (ข้อ)	80	87	88
คำตอบผิด (ข้อ)	6	5	6
ไม่พบคำตอบ (ข้อ)	19	13	11

ผลลัพธ์คำตอบของผู้ทดลองแสดงในตารางที่ 1 จะเห็นว่าผลลัพธ์ที่มีการใช้ฟีเจอร์แนะนำค่าสำคัญมีคำตอบถูกจำนวนสูงกว่าที่ไม่มีการใช้ฟีเจอร์แนะนำค่าสำคัญ การใช้งานของผู้ทดลองของฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยีมีจำนวนคำตอบถูกที่ 88 ข้อ ซึ่งสูงกว่าการใช้งานฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม และมีจำนวนการไม่พบคำตอบที่ 11 ข้อ ซึ่งน้อยกว่าการใช้งานฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม ส่วนฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม ผู้ทดลองมีจำนวนข้อที่ตอบผิดน้อยที่สุดคือ 5 ข้อ เมื่อเทียบกับฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยี สังเกตได้ว่าเมื่อผู้ทดลองใช้ฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยีจำนวนข้อที่ตอบถูกมีเพิ่มขึ้นและการไม่พบคำตอบลดลง ในส่วนคำตอบที่ผิดมีความใกล้เคียงกันในฟีเจอร์แนะนำค่าสำคัญ

จากตารางที่ 1 ผลคำตอบถูก ผู้ทดลองฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยีมีคำตอบถูกมากที่สุด เมื่อสังเกตคำค้นที่ใช้มีจำนวนใกล้เคียงกันในทุกฟีเจอร์แนะนำค่าสำคัญ แต่เมื่อดูการใช้งานคำค้น มีคำค้นที่มาจากการแนะนำของฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยีคือ 0.31 คำ/ข้อ ทำให้ผู้ทดลองที่ได้รับคำแนะนำค่าสำคัญจากฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยี ตอบถูกมากกว่าฟีเจอร์แนะนำค่าสำคัญอื่น

2.2 ผลการเปรียบเทียบประสิทธิภาพของการแนะนำค่าสำคัญสำหรับการค้นหาข้อมูลภาษาไทย

ตารางที่ 2 ผลลัพธ์ค่าเฉลี่ยจำนวนคำค้นหาของคำตอบที่ได้ต่อข้อของแต่ละผลลัพธ์การตอบ

	ฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม	ฟีเจอร์แนะนำค่าสำคัญจากออนโทโลยี
ค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ถูก/ครั้ง/ข้อ	2.63	2.49
ค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ผิด/ครั้ง/ข้อ	3.00	2.33
ค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ไม่พบคำตอบ/ครั้ง/ข้อ	2.15	2.55

จากตารางที่ 2 แสดงผลลัพธ์ค่าเฉลี่ยจำนวนคำค้นหาของคำตอบที่ได้ต่อข้อของแต่ละผลลัพธ์การตอบ เพื่อเทียบประสิทธิภาพฟีเจอร์แนะนำค่าสำคัญ ระหว่างฟีเจอร์แนะนำค่าสำคัญจากพจนานุกรม และฟีเจอร์แนะนำค่า

สำคัญจากออนไลน์ จากตารางที่ 2 ค่าตอบที่ผิดจำนวนสูงสุดเกิดจากการใช้พีเจอร์แนะนำ คำสำคัญจากพจนานุกรม โดยที่พีเจอร์แนะนำคำสำคัญจากออนไลน์ ค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ถูกมีค่าคือ 2.49 คำ/ ครั้ง/ข้อ ซึ่งต่ำที่สุด และพีเจอร์แนะนำคำสำคัญจากพจนานุกรม มีค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ไม่พบคำตอบมีค่าคือ 2.15 คำ/ครั้ง/ข้อ ซึ่งต่ำที่สุด

ตารางที่ 3 ผลลัพธ์ค่าเฉลี่ยจำนวนการใช้คำแนะนำของคำตอบที่ได้ต่อข้อของแต่ละผลลัพธ์การตอบ

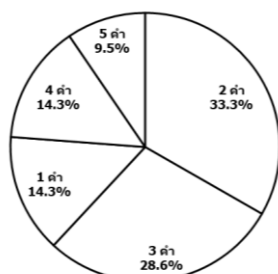
	พีเจอร์แนะนำคำสำคัญจาก พจนานุกรม	พีเจอร์แนะนำคำสำคัญจาก ออนไลน์
ค่าเฉลี่ยจำนวนการใช้คำแนะนำและคำตอบถูก/ข้อ	0.00	0.31
ค่าเฉลี่ยจำนวนการใช้คำแนะนำและคำตอบผิด/ข้อ	0.00	0.33
ค่าเฉลี่ยจำนวนการใช้คำแนะนำและไม่พบคำตอบ/ข้อ	0.00	0.36

จากตารางที่ 3 พีเจอร์แนะนำคำสำคัญจากออนไลน์ มีค่าเฉลี่ยจำนวนในการใช้คำแนะนำและตอบถูก, ตอบผิด และไม่พบคำตอบ สูงที่สุดคือ 0.31 0.33 และ 0.36 คำ/ข้อ ตามลำดับ ซึ่งจากผลสัมภาษณ์ของผู้ทดลอง พบว่าผู้ทดลองไม่เลือกคำแนะนำจากพีเจอร์แนะนำคำสำคัญจากพจนานุกรม ซึ่งทำให้พีเจอร์แนะนำคำสำคัญจากพจนานุกรม เทียบเคียงได้กับระบบที่ไม่มีการแนะนำคำสำคัญ

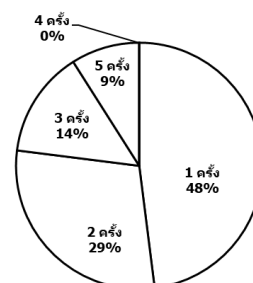
3. ศึกษาผลการทดลองใช้พีเจอร์การแนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทยกับผู้ใช้งานจริง

จากแบบสัมภาษณ์เพื่อดูพื้นหลังของผู้ทดลองจาก คำถามในแบบสัมภาษณ์โดยมีผลดังนี้ ภาพที่ 6 กราฟแสดงสัดส่วนของผู้ทดลองแสดงจำนวนคำที่เคยใช้เพื่อค้นหาข้อมูล ซึ่งจากกราฟผู้ทดลองส่วนใหญ่ใช้ประมาณ 2 - 3 คำ ในการค้นหาข้อมูลแต่ละครั้งตามลำดับ ภาพที่ 7 กราฟแสดงสัดส่วนจำนวนครั้งการค้นหาของผู้ทดลอง ซึ่งส่วนใหญ่ใช้การค้นหาประมาณ 1 - 2 ครั้ง ในการค้นหาข้อมูล และ ภาพที่ 8 กราฟแสดงสัดส่วนเว็บไซต์ที่ผู้ทดลองเคยใช้ค้นหาของผู้ทดลอง ส่วนใหญ่เป็น Google.com และ Yahoo.com ตามลำดับ พบว่า ผลการทดลองระบบค้นหาจากการแนะนำคำสำคัญโดยใช้ออนไลน์สามารถช่วยให้การค้นหาได้ผลลัพธ์ที่ตรงตามต้องการมากที่สุดเมื่อเทียบกับการใช้ระบบค้นหาอื่นจากออนไลน์

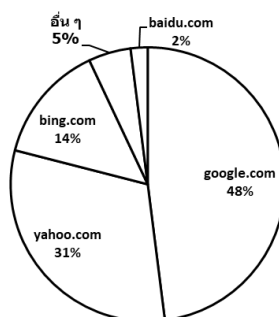
อภิปรายผลการวิจัย



ภาพที่ 6 กราฟแสดงจำนวนคำที่ใช้ค้นหาของผู้ทดลอง



ภาพที่ 7 กราฟแสดงจำนวนคำที่ใช้ค้นหาและผลได้ตามที่ต้องการของผู้ทดลอง



ภาพที่ 8 กราฟแสดงอัตราส่วนการใช้เว็บไซต์สำหรับการค้นหาของผู้ทดลอง

1. ผลการพัฒนาพีเจอร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย เป็นเครื่องมือที่ช่วยแนะนำคำสำคัญเพื่อนำมาใช้เป็นคำค้นหา เพื่อช่วยให้ผลการค้นหาตรงตามความต้องการของผู้ใช้ โดยใช้เทคนิคการแนะนำคำสำคัญจากพจนานุกรมและออนโทโลยี พบว่า การใช้คำสำคัญจากระบบที่มีพีเจอร์การแนะนำคำสำคัญ ช่วยให้ในการค้นหาช่วยให้ผลลัพธ์การค้นหาถูกต้องมากขึ้น ซึ่งสอดคล้องกับ N. Lin, V. A. Kudinov, H. M. Zaw and S. Naing [6] ได้ศึกษาวิธีการขยายคำค้นหาของระบบการค้นคืน Myanmar information retrieval (MIR) ซึ่งเขียนเป็นภาษาเมียนมาร์ พบว่า การพัฒนา WordNet ที่สำหรับขยายคำสำหรับค้นหา เพื่อแก้ไขปัญหาคำที่คลุมเครือทางความหมาย ทำให้ผลลัพธ์การค้นหามีคำตอบมากขึ้น

2. ผลการเปรียบเทียบพีเจอร์แนะนำคำสำคัญระหว่างคำแนะนำคำสำคัญจากพจนานุกรมกับการแนะนำคำสำคัญจากออนโทโลยี เพื่อดูผลลัพธ์คำตอบจากการค้นหาของผู้ทดลอง พบว่า เมื่อผู้ทดลองใช้คำค้นหาที่ได้จากพีเจอร์แนะนำคำสำคัญจากออนโทโลยีส่งผลให้ผลการค้นหาได้คำตอบมากที่สุด โดยมีการใช้คำค้นที่มาจากการแนะนำของพีเจอร์แนะนำคำสำคัญจากออนโทโลยีคือ 0.31 คำ/ข้อ และมีค่าเฉลี่ยจำนวนคำค้นของคำตอบที่ถูกของพีเจอร์แนะนำคำสำคัญจากออนโทโลยี มีค่าสูงที่สุด 2.49 คำ/ครั้ง/ข้อ ซึ่งอธิบายได้ว่าการใช้พีเจอร์แนะนำคำสำคัญจากออนโทโลยีช่วยให้การใช้จำนวนคำในการค้นหาต่อครั้งน้อยลงทำให้ผู้ใช้ระบบที่ได้รับคำแนะนำคำสำคัญจากพีเจอร์แนะนำคำสำคัญจากออนโทโลยี ตอบถูกมากกว่าพีเจอร์แนะนำคำสำคัญอื่น เหตุที่เป็นเช่นนี้อาจเนื่องมาจาก

การใช้พีเจอาร์การแนะนำคำสำคัญจากออนโทโลยีจะแนะนำคำสำคัญที่มีบริบทที่เกี่ยวข้องกับคำที่ผู้ใช้ระบบใส่เข้ามา สอดคล้องกับ N. Lin, V. A. Kudinov, H. M. Zaw and S. Naing [6] ที่ใช้ WordNet สำหรับให้คำขยายสำหรับเป็นคำในการค้นหาเพิ่มทำให้ได้คำตอบที่ถูกต้องมากยิ่งขึ้น

3. ผลการศึกษาจากการทดลองใช้พีเจอาร์แนะนำคำสำคัญสำหรับการค้นหาข้อมูลภาษาไทย พบว่า ผู้ใช้ระบบใช้พีเจอาร์แนะนำคำสำคัญจากออนโทโลยีช่วยให้เวลาที่ใช้ค้นหาน้อยลงเมื่อเทียบกับระบบอื่นที่นำมาใช้ทดสอบ และเป็นพีเจอาร์แนะนำคำสำคัญเดียวที่มีผู้ทดลองเลือกใช้คำสำคัญที่พีเจอาร์แนะนำมาเป็นคำค้น ซึ่งเมื่อใช้พีเจอาร์แนะนำคำสำคัญจากออนโทโลยีทำให้การค้นหาคำตอบที่ไม่พบผลลัพธ์ที่ต้องการลดลง และเจอคำตอบที่ถูกต้องมากขึ้นอย่างมีนัยสำคัญ เหตุที่เป็นเช่นนี้อาจเนื่องมาจากการนำคำที่ผู้ใช้ระบบใส่เข้ามา ซึ่งเป็นคำที่ผู้ใช้ต้องการค้นหาแต่ยังไม่มากพอที่จะใช้สำหรับนำมาใช้ค้นหา และผู้ใช้ได้ผลคำตอบที่ยังไม่ถูกต้อง ทำให้การใช้คำสำคัญที่แนะนำจากออนโทโลยีจะช่วยให้ได้คำค้นหาที่ตรงเพิ่มขึ้น และทำให้ใช้เวลาค้นหาน้อยลง สอดคล้องกับ Gan, L. and Tu, W. [4] ที่ใช้การขยายคำค้นหาจากคำที่ผู้ใช้ใส่เข้ามา กับ Markov semantic network ทำให้ได้คำขยายในสิ่งที่ต้องการค้นหาได้ถูกต้องยิ่งขึ้น

ข้อเสนอแนะ

การแนะนำคำสำคัญสามารถนำไปช่วยในการปรับปรุงการจัดลำดับเว็บไซต์ให้มีประสิทธิภาพในการค้นหาข้อมูลจากคำค้นของผู้ทดลองที่รู้เพียงคำในบริบทของเอกสารที่ต้องการค้นหา แต่ไม่รู้จักคำสำคัญที่ต้องใช้ในการค้นหาที่ตรงกับเอกสารหรือไม่มีความสามารถในการหาข้อมูล ให้สามารถพบเอกสารที่ถูกต้องและรวดเร็ว

กิตติกรรมประกาศ

งานวิจัยนี้ประสบความสำเร็จได้ต้องขอขอบพระคุณทุนวิจัยจาก สำนักงานการวิจัยแห่งชาติ (วช.) เลขที่ RDG61Q0020 ขอขอบพระคุณ ผศ.ดร.ชัยณรงค์ เกษามูล ในการเป็นที่ปรึกษาของงานวิจัยชิ้นนี้ และขอขอบพระคุณ นางสาว วรวรรณ วันดี และ นาย วรพล ล้วนตั้งศรีสุข ผู้ช่วยพัฒนาระบบของงานวิจัยนี้

เอกสารอ้างอิง

- [1] Cho, J. and Garcia-Molina, H., 2000, The Evolution of the Web and Implications for an Incremental Crawler, pp. 200-209, Proceedings of the 26th International Conference on Very Large Data Bases, San Francisco.
- [2] Gillies, J. and Cailliau, R., 2000, How the Web Was Born: The Story of the World Wide Web. Oxford University Press, United Kingdom, 392 p.
- [3] Bharathi, G. and Venkatesan, D., 2012, Improving Information Retrieval Using Document Clusters and Semantic Synonym Extraction, J. Theor. Appl. Inf. Tech. 36: 167-173.
- [4] Gan, L. and Tu, W., 2014, Improving Query Expansion Using Wikipedia, pp. 143-146, 2014 International Conference on Management of e-Commerce and e-Government, Shanghai.

- [5] Toman, S., Abed, M., & Toman, Z.H., 2020, Cluster-Based Information Retrieval by using (K-means)- Hierarchical Parallel Genetic Algorithms Approach, TELKOMNIKA Telecommunication, Computing, Electronics and Control Vol. 19, No. 1, February 2021, pp. 349-356.
- [6] N. Lin, V. A. Kudinov, H. M. Zaw and S. Naing, 2020, Query Expansion for Myanmar Information Retrieval Used by WordNet, 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg and Moscow, Russia , pp. 395-399
- [7] Kim, Y., & Croft, W. B., 2015, Improving Patent Search by Search Result Diversification. Paper presented at the Proceedings of the 2015 International Conference on The Theory of Information Retrieval, Northampton, Massachusetts, USA.
- [8] Rawat, R., Nayak, R., & Li, Y., 2011, Improving web database search incorporating users query information. Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS 2011, Sogndal, Norway, May 25 - 27, 2011.
- [9] Saravanan, M., Ravindran, B. & Raman, S., 2009, Improving legal information retrieval using an ontological framework, Artif Intell Law 17, 101-124.
- [10] Noy, N. F. and McGuinness, D. L., 2001, Ontology Development 101: A Guide to Creating Your First Ontology, Available Source: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf, August 3, 2020.
- [11] Langville, A. N. and Meyer, C. D., 2006, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, United Kingdom, 240 p.
- [12] Lyman, P. and Varian, H. R., 2003, How Much Information 2003?, Available Source: <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>, June 14, 2020.
- [13] Salton, G. and Buckley, C., 1988, Term-weighting Approaches in Automatic Text Retrieval, Inf. Process. Manage. 24(5): 513-523.
- [14] Fredkin, E., 1960, Trie Memory, Commun. ACM 3(9): 490-499.
- [15] Teeramunkong, T., Sornlertlamvanich, V., Tanhermhong T. and Chinnan, W., 2000, Character Cluster Based Thai Information Retrieval, pp. 75-80, Proceedings of the fifth international workshop on Information retrieval with Asian languages (IRAL '00), Assoc. Comp. Mach., New York.