

ระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในสื่อสิ่งพิมพ์

A Knowledge Navigation System for Accessing Contents in Printed Materials

สมจิน เปียโคกสูง¹
ดร. นิสัชล จำนงศรี²

Somjin PhiaKoksong
Dr. Nisachol Chamnongsri

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนากระบวนการและระบบนำทางความรู้ และประเมินประสิทธิภาพของระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์ โดยใช้แหล่งข้อมูลจากรายการสารบัญและดัชนีท้ายเล่มของสื่อสิ่งพิมพ์เป็นตัวแทนของประเด็นเนื้อหา ประยุกต์ใช้เทคนิคเอ็นแกรมและแบบจำลองปริภูมิเวกเตอร์สเปซโมเดลเพื่อสืบค้นข้อความในโมดูลการสืบค้น ใช้โครงสร้างกราฟเอ็มแอลนำเสนอและเชื่อมโยงผลการค้น ผลการวิจัย พบว่า ระบบที่พัฒนาขึ้นมีประสิทธิภาพในการนำทางความรู้โดยมีค่าความแม่นยำเฉลี่ยที่ 0.81 ค่าความระลึกเฉลี่ยที่ 0.64 และค่าเอฟเมเชอร์เฉลี่ยที่ 0.71 การประเมินระดับความพึงพอใจของผู้ใช้งานที่มีต่อระบบ พบว่า โดยภาพรวมผู้ใช้มีความพึงพอใจต่อระบบในระดับมาก ที่ค่าเฉลี่ยที่ 3.72 โดยด้านประโยชน์ในการส่งเสริมการเรียนรู้และด้านคุณภาพความรู้และสารสนเทศที่ได้รับผู้ใช้มีความพึงพอใจในระดับมาก ส่วนด้านสมรรถนะในการให้บริการของระบบผู้ใช้มีความพึงพอใจในระดับปานกลาง

Abstract

The purposes of this research were to design, develop and evaluate the efficiency of a knowledge navigation process and system for accessing contents in printed materials. This study used table of contents and book indexes as the resources of content representation, applied a n-gram technique and a vector space model to query in search module, and used GraphML structure to present and link between the search results. The results of this research found that the developed system has an efficiency in knowledge navigation, the average of precision is 0.81, recall is 0.64 and F-measure is 0.71. The evaluation of user satisfaction found that, over all users satisfy with the system at high level, average was 3.72. The dimensions of "The useful of learning support" and "The quality of knowledge and information" were rated at high level while "The performance of system service" was rated at moderate.

คำสำคัญ: การนำทางความรู้ ระบบนำทางความรู้ แบบจำลองการค้นคืนสารสนเทศ เทคนิคเอ็นแกรม
แบบจำลองปริภูมิเวกเตอร์ การสืบค้นสารสนเทศ

Keywords: Knowledge navigation, Knowledge navigation system, Information retrieval model,
N-gram technique, Vector space mode, Information retrieval

¹ นักศึกษาระดับบัณฑิตศึกษา สาขาวิชาวิทยาการสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

² อาจารย์ประจำสาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

บทนำ

ปัจจุบันรูปแบบและพัฒนาการของศาสตร์แขนงต่างๆ เกิดการผสมผสานและเชื่อมโยงความรู้จากศาสตร์หลากหลายแขนงเข้าด้วยกัน ในลักษณะของสหวิทยาการ (Interdisciplinary) ก่อให้เกิดวิทยาการแขนงใหม่ๆ ที่มีได้ตั้งอยู่บนฐานความรู้จากศาสตร์เพียงแขนงใดแขนงหนึ่ง เช่น การศึกษาด้านสิ่งแวดล้อม ซึ่งมีความเกี่ยวข้องกับทั้งศาสตร์ด้านธรรมชาติวิทยา มานุษยวิทยา ศิลปะและสังคมวิทยา (Caviglia-Harris and Hatley, 2004) เป็นต้น

จากลักษณะของความรู้ที่เปลี่ยนแปลงเข้าสู่รูปแบบของสหวิทยาการ ส่งผลให้รูปแบบและเนื้อหาสาระของทรัพยากรสารสนเทศทั้งที่อยู่ในรูปของสื่อสิ่งพิมพ์หรือสื่อดิจิทัลมีความซับซ้อนในด้านประเด็นเนื้อหา ทำให้ผู้ที่ต้องการศึกษาค้นคว้าจำเป็นต้องคำนึงถึงจากทรัพยากรสารสนเทศในศาสตร์หลากหลายแขนง เพื่อให้สามารถเข้าถึงเนื้อหาในแขนงวิชานั้นๆ ได้อย่างถูกต้องแท้ ถูกต้อง และครอบคลุม อันจะส่งเสริมการเรียนรู้ของแต่ละบุคคลให้เกิดประสิทธิภาพสูงสุด

ปัจจุบันสถาบันที่ให้บริการสารสนเทศได้พยายามพัฒนาเครื่องมือช่วยค้นสารสนเทศให้มีความสามารถมากขึ้น ทั้งในด้านการสืบค้นและแสดงผล อย่างไรก็ตาม เครื่องมือช่วยค้นดังกล่าวยังไม่สามารถตอบสนองความต้องการสืบค้นทรัพยากรสารสนเทศที่มีลักษณะสหวิทยาการได้อย่างมีประสิทธิภาพ ทั้งนี้เนื่องจาก

1. คำสำคัญและหัวเรื่องที่ห้องสมุดกำหนดให้เพื่อเป็นคำค้น เป็นคำในระดับกว้าง (General terms) ที่ทำหน้าที่แทนเฉพาะในระดับตัวเล่ม ทำให้ผู้ใช้ไม่สามารถสืบค้นไปในระดับประเด็นย่อยๆ เช่น บท หรือ ตอน ที่มีอยู่ภายในเล่มได้
2. คำสืบค้นในระดับกว้างมักเป็นคำศัพท์ควบคุม (Controlled vocabulary) ซึ่งผู้ใช้งานส่วนใหญ่ไม่คุ้นเคยกับคำศัพท์ดังกล่าว ส่งผลให้เกิดความผิดพลาดในการสืบค้น (Borgman, 1996; Sridhar, 2004 และ

ศุภรินทร์ วิมุกตายน, 2548) ทั้งนี้ ผู้ใช้งานส่วนใหญ่มักสืบค้นด้วยคำศัพท์อิสระซึ่งเป็นคำที่มักปรากฏอยู่ในเนื้อหา (รีนฤดี ไชยวิชิตกุล, 2550) และเป็นศัพท์ที่มีความทันสมัยกว่าศัพท์ควบคุม

3. การกำหนดค่าแทนสาระในระดับกว้างส่งผลให้ได้รับรายการสืบค้นจำนวนมาก ซึ่งผู้ใช้งานต้องเสียเวลาในการพิจารณาว่าหนังสือเล่มใดมีเนื้อหาที่ตรงตามความต้องการ และจากการศึกษางานวิจัยด้านพฤติกรรมกรสืบค้นสารสนเทศของผู้ใช้ พบว่า ผู้ใช้งานส่วนใหญ่พิจารณาผลการสืบค้นไม่เกิน 5 รายการแรก และร้อยละ 66 พิจารณาไม่เกิน 2-3 รายการของผลการสืบค้น (Jansen and Spink, 2003) ร้อยละ 81 พิจารณาเฉพาะหน้าจอแรก (Beitzel, Jensen, Chowdhury, Grossman and Frieder, 2004) และร้อยละ 90 พิจารณาไม่เกิน 3 หน้าจอแรก (iProspect, 2006)

4. รูปแบบการนำเสนอผลการสืบค้นของเครื่องมือช่วยค้นโดยทั่วไปจะแสดงเฉพาะชื่อเรื่อง ชื่อผู้แต่ง และเลขเรียกหนังสือ แม้จะมีการแสดงหัวเรื่องแต่ยังคงไม่สามารถแสดงให้เห็นถึงความเชื่อมโยงไปสู่เนื้อหาประเด็นย่อยระหว่างสื่อสิ่งพิมพ์ อีกทั้งยังไม่สามารถแสดงความสัมพันธ์ระหว่างประเด็นย่อยภายในเล่มเดียวกันได้ ซึ่งไม่สนับสนุนให้เกิดการศึกษาค้นคว้าอย่างกว้างขวางและมีประสิทธิภาพ

จากปัญหาที่ปรากฏอยู่ของระบบการสืบค้นสารสนเทศประเภทสื่อสิ่งพิมพ์ที่มีในปัจจุบันเกี่ยวกับการสืบค้นสื่อสิ่งพิมพ์ซึ่งมีลักษณะเนื้อหาเป็นสหวิทยาการดังที่กล่าวมาข้างต้น กอปรกับการศึกษางานวิจัยซึ่งปรากฏแนวคิดเกี่ยวกับการนำทางความรู้ (Knowledge navigation) ที่จะช่วยสนับสนุนให้ผู้ที่ต้องการความรู้สามารถรู้และเข้าถึงแหล่งความรู้ได้อย่างถูกต้อง โดยการเชื่อมโยงระหว่างทรัพยากรสารสนเทศหนึ่งไปยังทรัพยากรสารสนเทศอื่นๆ ผู้วิจัยจึงมุ่งศึกษาและพัฒนากระบวนการที่จะช่วยในการสืบค้น เพื่อให้สามารถเข้าถึงสื่อสิ่งพิมพ์ที่มีลักษณะสหวิทยาการได้อย่างมีประสิทธิภาพสูงสุด

วัตถุประสงค์การวิจัย

1. เพื่อออกแบบและพัฒนากระบวนการนำทางความรู้เพื่อการเข้าถึงเนื้อหาในทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์
2. เพื่อพัฒนาระบบการนำทางความรู้เพื่อการเข้าถึงเนื้อหาในทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์
3. เพื่อประเมินประสิทธิภาพของระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์

ขอบเขตงานวิจัย

1. การวิจัยครั้งนี้ใช้ทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์ที่เป็นตำราภาษาอังกฤษ ในสามหมวดหมู่ คือ คณิตศาสตร์ ฟิสิกส์ และวิศวกรรมไฟฟ้า ในหัวข้อ สนามแม่เหล็กไฟฟ้า (Electromagnetic fields)
2. ข้อมูลที่ใช้เป็นตัวแทนเนื้อหาของสื่อสิ่งพิมพ์ในการวิจัยครั้งนี้ ได้แก่ รายการสารบัญ (Table of contents) และดัชนีท้ายเล่ม (Book indexes) ของหนังสือ ซึ่งจะใช้เป็นแหล่งข้อมูลเพื่อการจัดทำฐานข้อมูลของระบบ

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การนำทางความรู้

Matsumura, Ohsawa and Ishizuka (2001) ได้ให้คำจำกัดความของการนำทางความรู้ โดยกล่าวว่า ระบบนำทางความรู้เป็นส่วนหนึ่งของซอฟต์แวร์ที่ช่วยในการตอบคำถามของผู้ใช้งาน ซึ่งคำถามนั้นสามารถเป็นได้ทั้งในลักษณะของชุดคำถาม หรือประโยคคำถาม

Qiu, Yao, Wang and Wang (2006) ระบุถึงลักษณะของการนำทางความรู้ โดยกล่าวว่า การนำทางความรู้ เป็นการนำไปสู่ความรู้ด้วยการสร้างเครือข่ายของความรู้ นำเสนอความรู้ที่มีและเชื่อมโยงไปยังแหล่งทรัพยากรสารสนเทศที่จะช่วยให้ผู้ใช้ได้รับความรู้นั้นๆ

Le-Khac, Aouad and Kechadi (2007) กล่าวว่า การนำทางความรู้ เป็นกระบวนการซึ่งทำหน้าที่

ในการชี้แนะให้ผู้ใช้งานที่ยังไม่สามารถระบุถึงความรู้ที่ต้องการได้อย่างแท้จริงให้สามารถกำหนดหรือเข้าถึงความรู้ที่ต้องการได้

จากคำจำกัดความข้างต้น สามารถกล่าวได้ว่าการนำทางความรู้ คือ กระบวนการหรือซอฟต์แวร์ลักษณะหนึ่งที่ทำหน้าที่ในการค้นคืน ชี้นำหรือหาคำตอบให้แก่ผู้ใช้งานโดยการสร้างเครือข่ายของความรู้ที่จะช่วยให้ผู้ใช้ระบุถึงความรู้ที่ต้องการ หรือเข้าถึงแหล่งทรัพยากรสารสนเทศที่สอดคล้องกับความรู้ที่ผู้ใช้งานต้องการได้

การกำหนดลำดับความสำคัญของเอกสาร

การจัดลำดับความสำคัญของเอกสารอาศัยกระบวนการสำคัญ ได้แก่ การกำหนดค่าน้ำหนักของคำตรรกะ (Term weight) และแบบจำลองการสืบค้นสารสนเทศ (Information retrieval model)

สุธรรม อูมาแสงทองกุล (2541) ได้ปริทัศน์วรรณกรรมเกี่ยวกับการจัดทำตรรกะแบบกำหนดค่าน้ำหนักและจำแนกการทำตรรกะแบบกำหนดค่าน้ำหนักออกเป็น 2 ประเภท คือ

- 1) การกำหนดค่าน้ำหนักในลักษณะการประเมินเชิงอัตวิสัย โดยใช้วิธีการวิเคราะห์เนื้อหาเอกสาร
- 2) การกำหนดค่าน้ำหนักโดยการคำนวณทางสถิติในลักษณะการแจกแจงทางสถิติ เช่น การนับ ความถี่ของคำ เป็นต้น

การใช้ตรรกะแบบกำหนดค่าน้ำหนักในระบบสืบค้นสารสนเทศ สามารถนำไปใช้ในขั้นตอน ดังนี้

- 1) การใช้ค่าน้ำหนักในขั้นตอนการทำตรรกะ
- 2) การใช้ค่าน้ำหนักในขั้นตอนการสืบค้นข้อมูล
- 3) การใช้ค่าน้ำหนักในขั้นตอนการแสดงผลข้อมูลด้วยการเรียงลำดับผลลัพธ์ตามค่าน้ำหนัก

ตัวแบบการสืบค้น

ศุภชัย ตั้งวงศ์ศานต์ (2551) ได้สรุปลักษณะของแบบจำลองการสืบค้นสารสนเทศ โดยแบ่งได้ 3 แบบหลัก ได้แก่ แบบจำลองบูลีน (Boolean model) แบบจำลองปริภูมิเวกเตอร์ (Vector space model) และแบบจำลองความน่าจะเป็น (Probabilistic relevance

model) โดยแบบจำลองปริภูมิเวกเตอร์เป็นแบบจำลองที่ได้รับความนิยมในทางปฏิบัติมากกว่าแบบจำลองลักษณะอื่นๆ ซึ่งมีลักษณะของแบบจำลองและการจัดเก็บข้อมูลให้อยู่ในรูปเมทริกซ์ของคำ (Term-document matrix) แต่ละหน่วยในเมทริกซ์จะมีค่าน้ำหนักที่ขึ้นอยู่กับความถี่ของคำที่ปรากฏในเอกสาร (Term frequency) และความถี่ผกผันของเอกสาร (Inverse document frequency) อัตราส่วนระหว่างจำนวนเอกสารทั้งหมดกับจำนวนเอกสารที่มีคำนั้นปรากฏ คำสอบถาม (Queries) จะเขียนอยู่ในรูปของเวกเตอร์ที่สอดคล้องกับเมทริกซ์ของเอกสาร คำที่ปรากฏจะมีค่าน้ำหนักแตกต่างกัน ขึ้นอยู่กับความสำคัญ หรือเป็น 0 หากไม่ปรากฏ ในการประมวลผลจะทำการเปรียบเทียบการทำมุมระหว่างเวกเตอร์ของคำสอบถามกับเวกเตอร์ของเอกสารแต่ละฉบับ เป็นลักษณะการคำนวณทางคณิตศาสตร์โดยการหาผลคูณภายใน (Inner product) ระหว่างเวกเตอร์ ซึ่งผลลัพธ์ดังกล่าวจะมีค่าอยู่ระหว่าง 0 ถึง 1 และสามารถใช้ในการจัดลำดับความสำคัญของเอกสารที่เป็นคำตอบตามค่าที่คำนวณได้

เทคนิคเอ็นแกรม

เทคนิคเอ็นแกรมเป็นเทคนิคหนึ่งในการเปรียบเทียบและจับคู่ข้อมูลที่ได้รับคามนิยมมาก เทคนิคเอ็นแกรมอาศัยกระบวนการประมวลผลตรรกะจากการแบ่งกลุ่มของคำหรือข้อความตามขนาดที่กำหนด ($n = 1, 2, 3, \dots$) ซึ่งเทคนิคเอ็นแกรมนั้นสามารถจัดกระทำได้ใน 2 ลักษณะ คือ แบ่งกลุ่มคำเป็นฐาน (Term-based index) และแบ่งกลุ่มอักษรเป็นฐาน (Character-based index) (Chu, 2003; Yamamoto, Ohmi and Tsuji, 2003) โดยมีลักษณะการจัดทำชุดข้อมูลที่มีการซ้อนทับกัน (Overlapping) จากชุดข้อมูลหลักที่ต้องการ เช่น

$n = 2$ คำว่า method = {me, et, th, ho, od}

$n = 3$ คำว่า method = {met, eth, tho, hod}

และทำการเปรียบเทียบความคล้ายคลึงกันของข้อมูล โดยใช้สูตรต่างๆ อาทิ Jaccard-Needham, Dice, Correlation, Yule, และ Rogers-Tanimoto เป็นต้น

ซึ่งสูตรต่างๆ ที่สามารถใช้ในการเปรียบเทียบข้อมูลได้ถูกนำเสนอในงานวิจัยของ Choi, Cha and Charles (2010) Lourenco, Lobo and Bacao (2004), Monev (2004) และ Zhang and Sargur (n.d.)

จากการศึกษาของ Salton (1988) ได้ระบุว่า การเลือกใช้เทคนิคเอ็นแกรมในลักษณะของไบนแกรม (Bigrams) ($n = 2$) และไตรแกรม (Trigrams) ($n = 3$) เป็นขนาดที่เหมาะสมที่สุดสำหรับใช้เพื่อการค้นคืนคำที่มีความคล้ายคลึง ส่วนการใช้ยูนิแกรม (Unigrams) ($n = 1$) จะให้ผลการค้นคืนมากเกินไปและไม่สามารถแยกแยะความแตกต่างได้ ในขณะที่การกำหนดให้ n มีขนาดที่ยาวเกินไป ($n \geq 4$) จะส่งผลให้เกิดความผิดพลาดได้มากกว่าสำหรับคำที่มีขนาดสั้น

เทคนิคการตรวจสอบความสัมพันธ์เชิงความหมาย

การวัดความสัมพันธ์เชิงความหมาย (Semantic relatedness) เป็นวิธีการหนึ่งซึ่งช่วยให้การสกัดข้อมูลมีความถูกต้องและแม่นยำมากยิ่งขึ้น โดยพิจารณาจากความสัมพันธ์ทางความหมายที่แฝงอยู่ระหว่างข้อมูลหรือแนวคิดการหาความสัมพันธ์เชิงความหมายโดยวิธีทางสถิติ อาทิ วิธีมิวซวลอินฟอร์เมชัน วิธีพอยน์ไวซ์มิวซวลอินฟอร์เมชัน หรือวิธีไค-สแควร์ เป็นต้น โดยวิธีพอยน์ไวซ์มิวซวลอินฟอร์เมชันได้มีการประยุกต์ใช้กับข้อมูลจากโปรแกรมค้นหา (Search engine) (Bollegala, Matsuo and Ishizuka, 2007) เรียกว่า วิธีเว็บพอยน์ไวซ์มิวซวลอินฟอร์เมชัน (Web-pointwise mutual information)

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงประยุกต์ (Applied research) เพื่อศึกษาและพัฒนาระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในสื่อสิ่งพิมพ์ โดยใช้แนวทางของวงจรการพัฒนาระบบ (System Development Life Cycle/SDLC) ใช้แบบจำลองปริภูมิเวกเตอร์ร่วมกับเทคนิคเอ็นแกรมในการค้นคืนสารสนเทศและใช้โครงสร้างกราฟเอ็มแอลนำเสนอข้อมูลในรูปของแผนภาพ แบ่งการดำเนินการวิจัย 3 ระยะ ดังนี้

ระยะที่ 1 การศึกษาและวิเคราะห์ความต้องการของระบบ เพื่อกำหนดความต้องการของระบบ (System requirements) โดยศึกษาในประเด็นพฤติกรรมการสืบค้นสารสนเทศ โครงสร้างทรัพยากรสารสนเทศประเภทสื่อสิ่งพิมพ์ การกำหนดค่าแทนสาระของทรัพยากรสารสนเทศ และขั้นตอนวิธีต่างๆที่เกี่ยวข้องกับการนำทางความรู้ และเทคนิคเอ็นแกรม

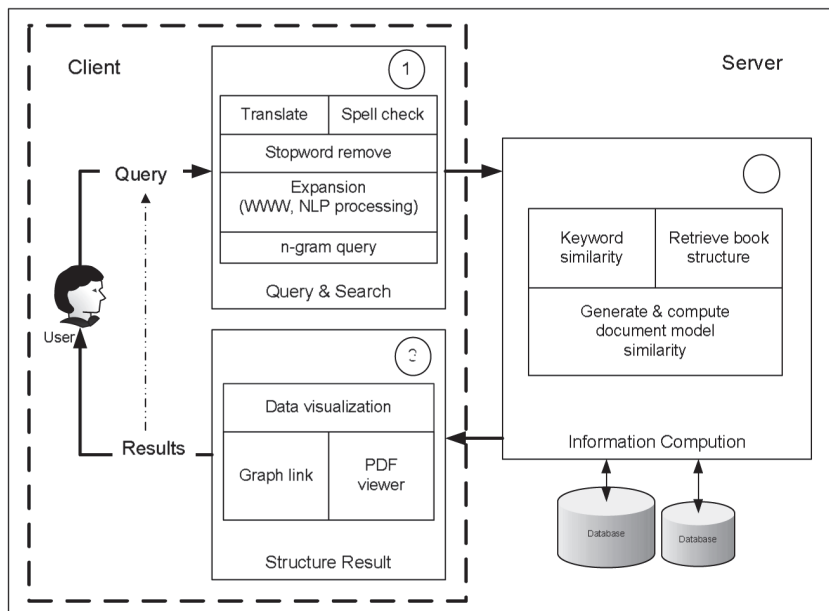
ระยะที่ 2 การออกแบบและพัฒนาระบบ จากผลการศึกษาความต้องการของระบบจากการดำเนินการตามระยะที่ 1 ผู้วิจัยได้นำมาเป็นข้อมูลพื้นฐานในการออกแบบและพัฒนาระบบ ประกอบด้วย การออกแบบโครงสร้างและวิธีการจัดเก็บข้อมูล การออกแบบและพัฒนาระบบการค้นคืนและการเชื่อมโยงข้อมูล และการออกแบบส่วนต่อประสานกับผู้ใช้และการแสดงผล (ภาพที่ 1) ออกแบบระบบให้สามารถทำงานผ่านทางเครือข่ายอินเทอร์เน็ตโดยมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 ผู้ใช้งานป้อนคำสืบค้นผ่านทางเครื่องลูกข่าย (Client)

ขั้นตอนที่ 2 ระบบนำคำสืบค้นเข้าสู่กระบวนการในการสืบค้นข้อมูลและนำทางความรู้ ซึ่งแบ่งเป็น 3 ส่วน ดังนี้

2.1) ส่วนข้อคำถามและการค้นหา (Query & Search) ทำหน้าที่ แปลภาษา (Translation) ตรวจสอบการสะกดคำ (Spell check) กำจัดคำหยุด (Stopword remove) ขยายคำสืบค้น (Expansion) และสร้างชุดคำค้นในลักษณะของชุดข้อมูลเอ็นแกรม (N-gram query)

2.2) ส่วนการคำนวณสารสนเทศ (Information computation) ทำหน้าที่ ค้นหาและเปรียบเทียบคำค้นกับคำสำคัญที่มีในระบบ (Keyword similarity) สร้างตัวแบบการสืบค้นของเอกสารเพื่อการเปรียบเทียบและหาระดับความคล้ายคลึงกับคำค้น (Generate & compute document model similarity) โดยใช้ค่าน้ำหนักคำสำคัญ ดังตารางที่ 1 และค้นคืนข้อมูลและโครงสร้างของเอกสารที่สอดคล้องกับคำค้น (Retrieve book structure)



ภาพที่ 1 โครงสร้างการทำงานของระบบนำทางความรู้

ตารางที่ 1 ค่าน้ำหนักตามระดับความสำคัญที่ปรากฏของคำตรรกะนี้

ระดับที่ปรากฏของคำตรรกะนี้	ค่าน้ำหนัก
ชื่อหนังสือ (BT)	30
ชื่อบทที่ (CH)	25
ชื่อหัวข้อย่อย (SC)	20
ชื่อหัวข้อย่อยระดับที่ 1 (SSC)	15
ชื่อหัวข้อย่อยระดับที่ 2 (SSSC)	10
คำตรรกะนี้ท้ายเล่ม (BI)	1
ชื่อตาราง (TN)	1
ชื่อภาพ (FN)	1

2.3) ส่วนโครงสร้างการแสดงผล (Structure result) ทำหน้าที่จัดทำโครงสร้างการเชื่อมโยงคำสำคัญระหว่างเอกสารกับคำค้น (Graph link) แสดงข้อมูลแผนภาพการเชื่อมโยงต่างๆ (Data visualization) และแสดงไฟล์หนังสือที่ต้องการ (PDF viewer)

ขั้นตอนที่ 3 ส่งผลการสืบค้นและโครงสร้างข้อมูลไปยังเครื่องลูกข่าย

ขั้นตอนที่ 4 เครื่องลูกข่ายแสดงผลการสืบค้นข้อมูล

ระยะที่ 3 การประเมินประสิทธิภาพของระบบ ผู้วิจัยกำหนดข้อมูลที่จะใช้ในการทดสอบระบบในเรื่องสนามแม่เหล็กไฟฟ้า และประเมินประสิทธิภาพของระบบใน 2 ลักษณะ ได้แก่

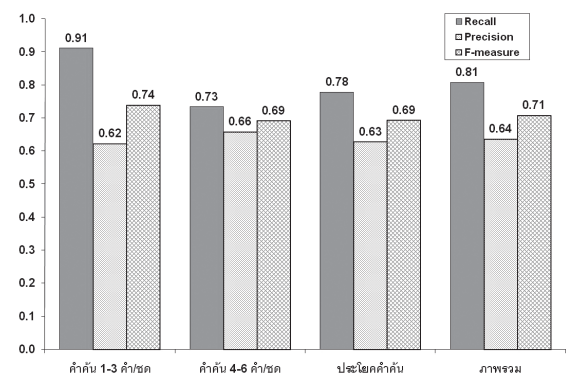
1) การประเมินประสิทธิภาพการทำงานด้านค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเอฟเมเชอร์ (F-measure) ของระบบ ซึ่งดำเนินการโดยผู้พัฒนาระบบและผู้เชี่ยวชาญจำนวน 4 ท่าน ทำการทดสอบสืบค้นด้วยคำค้น 3 ลักษณะ ได้แก่ ชุดคำค้นขนาด 1-3 คำต่อชุด ชุดคำค้นขนาด 4-6 คำต่อชุด และประโยคคำค้น

2) การประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบ กลุ่มตัวอย่าง ได้แก่ นักศึกษามหาวิทยาลัยเทคโนโลยีสุรนารี ระดับปริญญาตรี สาขาวิศวกรรมไฟฟ้า ชั้นปีที่ 3 จำนวน 32 คน ผู้วิจัยได้ทำการติดตั้ง

ระบบที่พัฒนาขึ้น บนเครื่องแม่ข่ายจำลอง กำหนดให้กลุ่มตัวอย่างทำการทดสอบเข้าใช้งานระบบเป็นระยะเวลา 1 สัปดาห์ โดยผู้ใช้งานสามารถสืบค้นด้วยคำค้นต่างๆ ที่เกี่ยวข้องกับเนื้อหาเรื่อง สนามแม่เหล็กไฟฟ้า ได้อย่างอิสระ เก็บรวบรวมข้อมูลโดยใช้แบบสอบถามมาตราส่วนประมาณค่า 5 ระดับ สอบถามใน 3 ประเด็น ได้แก่ ด้านคุณภาพความรู้และสารสนเทศที่ได้รับ ด้านประโยชน์ในการส่งเสริมการเรียนรู้ และด้านสมรรถนะในการให้บริการของระบบ วิเคราะห์ข้อมูลจากแบบสอบถามโดยใช้สถิติพรรณนา ได้แก่ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย

ผลการประเมินประสิทธิภาพของระบบ พบว่าระบบมีประสิทธิภาพการทำงานในระดับปานกลาง โดยมีค่าความแม่นยำเฉลี่ยที่ 0.64 ค่าความระลึกเฉลี่ยที่ 0.81 และค่าเอฟเมเชอร์ที่ 0.71 เมื่อพิจารณาจำแนกตามลักษณะของชุดคำค้น พบว่า ระบบมีประสิทธิภาพการทำงานด้านค่าความแม่นยำในระดับปานกลางทุกลักษณะคำค้น ขณะที่มีความระลึกในระดับดีมากในลักษณะคำค้น 1-3 คำต่อชุด ส่วนการค้นในลักษณะคำค้น 4-6 คำต่อชุดและประโยคคำค้น ระบบมีค่าความระลึกในระดับดี สำหรับค่าเอฟเมเชอร์ พบว่าทุกลักษณะคำค้นมีค่าเอฟเมเชอร์อยู่ในระดับปานกลาง (ภาพที่ 2)



ภาพที่ 2 ประสิทธิภาพการทำงานของระบบนำทางความรู้

ส่วนผลการประเมินความพึงพอใจของผู้ใช้ มีความพึงพอใจในระดับมาก โดยมีค่าเฉลี่ยที่ 3.87 และพบว่า ผู้ใช้มีความพึงพอใจต่อระบบโดยภาพรวมอยู่ในระดับมาก โดยมีค่าเฉลี่ยที่ 3.72 เมื่อพิจารณาในแต่ละด้าน พบว่า ด้านประโยชน์ในการส่งเสริมการเรียนรู้และด้านคุณภาพความรู้และสารสนเทศที่ได้รับ ผู้ใช้งานมีความพึงพอใจในระดับมาก โดยมีค่าเฉลี่ยที่ 3.75 ตามลำดับ ส่วนด้านสมรรถนะในการให้บริการของระบบ ผู้ใช้มีความพึงพอใจในระดับปานกลาง โดยมีค่าเฉลี่ยที่ 3.36 ดังตารางที่ 2

ตารางที่ 2 ความพึงพอใจของผู้ใช้งานที่มีต่อระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในสื่อสิ่งพิมพ์โดยภาพรวม

ข้อคำถาม	n	$\bar{X}$	S.D.	ระดับ
คุณภาพความรู้และสารสนเทศที่ได้รับ				
1. ความรู้/สารสนเทศที่ได้จากระบบตรงตามความต้องการที่จะใช้ในการศึกษาค้นคว้า	32	3.88	0.66	มาก
2. ความรู้/สารสนเทศที่ได้จากระบบมีความละเอียดเพียงพอต่อการนำไปใช้เพื่อการศึกษา ค้นคว้า	302	3.63	0.75	มาก
ภาพรวมด้านคุณภาพความรู้และสารสนเทศที่ได้รับ	32	3.75	0.64	มาก
ประโยชน์ในการส่งเสริมการเรียนรู้				
3. ระบบช่วยผู้ใช้งานให้สามารถสืบค้นและเข้าถึงคำสำคัญในระดับประเด็นย่อยๆ ภายในสื่อสิ่งพิมพ์ได้ง่าย	32	3.75	1.02	มาก
4. ระบบช่วยผู้ใช้งานให้สามารถขยายขอบเขตของเรื่องที่ต้องการศึกษาค้นคว้าได้เป็นอย่างดี	32	3.94	0.84	มาก
5. ระบบช่วยผู้ใช้งานให้สามารถจำกัดขอบเขต หรือเจาะลึกเรื่องที่ต้องการศึกษาค้นคว้าได้เป็นอย่างดี	32	3.75	0.80	มาก
6. ระบบช่วยผู้ใช้งานให้มองเห็นถึงความเชื่อมโยงของความรู้เรื่องที่ต้องการศึกษา ที่บันทึกไว้ในสื่อสิ่งพิมพ์ต่างเล่มกันได้อย่างดี	31	3.87	0.76	มาก
7. ระบบช่วยผู้ใช้งานให้มองเห็นถึงความเชื่อมโยงของความรู้เรื่องที่ต้องการศึกษา ที่บันทึกไว้ในสื่อสิ่งพิมพ์แต่ละเล่มได้เป็นอย่างดี	32	4.00	0.72	มาก
8. ระบบช่วยผู้ใช้งานในการคัดกรองและนำเสนอสื่อสิ่งพิมพ์ให้ตรงตามความต้องการของผู้ใช้งานได้อย่างดี	32	3.75	0.84	มาก
9. ระบบช่วยประหยัดเวลาในการพิจารณาคัดเลือกสื่อสิ่งพิมพ์ที่ต้องศึกษาได้เป็นอย่างดี	31	4.00	0.73	มาก
ภาพรวมด้านประโยชน์ในการส่งเสริมการเรียนรู้	32	3.87	0.58	มาก
สมรรถนะในการให้บริการของระบบ				
10. ระบบมีรูปแบบการใช้งานที่ง่ายต่อการเรียนรู้ (เช่น คำสั่งเข้าใจง่าย หน้าจอการแสดงผลเข้าใจง่าย เป็นต้น)	32	3.56	0.62	มาก
11. ระบบสามารถตอบสนองได้ทันระยะเวลาที่เหมาะสม	30	3.30	0.88	ปานกลาง
12. ระบบสามารถทำงานได้อย่างมีประสิทธิภาพ (เช่น ไม่ค้างบ่อย แสดงผลได้สมบูรณ์ เป็นต้น)	32	3.25	0.84	ปานกลาง
ภาพรวมด้านสมรรถนะในการให้บริการของระบบ	32	3.36	0.64	ปานกลาง
ภาพรวมทั้งหมด	32	3.72	0.52	มาก

สรุปและอภิปรายผล

ประสิทธิภาพในการสืบค้นข้อมูลของระบบ

1) จากการประเมินประสิทธิภาพด้านค่าความแม่นยำและค่าความระลึกโดยผู้เชี่ยวชาญ ซึ่งมีผลการประเมินในระดับปานกลาง และระดับความพึงพอใจของผู้ใช้งานที่มีต่อระบบในด้านคุณภาพความรู้และสารสนเทศที่ได้รับซึ่งมีระดับผลการประเมินในระดับมาก โดยเฉพาะรายข้อคำถามที่ 1) ความรู้/สารสนเทศที่ได้จากระบบตรงตามความต้องการที่จะใช้ในการศึกษาค้นคว้าแสดงให้เห็นถึงประสิทธิภาพในการสืบค้นข้อมูลของระบบที่สามารถทำงานได้ดี จากการประยุกต์ใช้เทคนิคเอ็นแกรมและตัวแบบปริภูมิเวกเตอร์ในการสร้างแบบจำลองการสืบค้น ซึ่งสอดคล้องกับผลการวิจัยของ Salton and Buckley (1988 อ้างถึงใน ศุภชัย ตั้งวงศ์ศานต์, 2551) ที่ระบุว่า ตัวแบบจำลองปริภูมิเวกเตอร์ สามารถทำงานได้มีประสิทธิภาพในกรณีโดยทั่วไปและเป็นที่ยอมรับในวงวิชาการ อีกทั้งประสิทธิภาพระบบที่พัฒนาขึ้นยังสอดคล้องกับ Mustafa and Al-Radaideh (2004) Barrón-Cedeño and Rosso (2009) และ สิทธิโชค ปัญญาฤกษ์ชัย และศิพาณี นุชิตประสิทธิ์ชัย (2552) ซึ่งได้ประยุกต์ใช้เทคนิคเอ็นแกรมในงานวิจัยและให้ค่าความแม่นยำและค่าความระลึกของระบบที่พัฒนาขึ้นในระดับที่สูงกว่า 0.60

เมื่อพิจารณาถึงประสิทธิภาพการค้นจากชุดคำค้นต่าง ๆ พบว่า ระบบให้ค่าความระลึกในระดับที่สูงและให้ค่าความแม่นยำในระดับปานกลางในทุกชุดคำค้น ทั้งนี้เนื่องจากการเลือกใช้ค่าแทนสาระจากรายการสารบัญหรือดรรชนีท้ายเล่ม แม้จะเป็นคำสืบค้นที่ผู้ใช้งานคุ้นเคยและระบุถึงเนื้อหาภายในสิ่งพิมพ์ได้เป็นอย่างดี แต่ค่าแทนสาระดังกล่าวมีระดับความครอบคลุม (Exhaustivity) ที่มากจนเกินไป ซึ่งส่งผลให้ได้รับค่าความระลึกในระดับที่สูงโดยมีค่าความแม่นยำที่ลดต่ำลงสอดคล้องกับ Maltby (1975 อ้างถึงใน Macgregor and McCulloch, 2006) และ Lancaster (1968 อ้างถึงใน สุธรรม อูมาแสงทองกุล, 2541) ซึ่งระบุว่า คำดรรชนีที่มีระดับความครอบคลุมสูงจะส่งผลให้ได้รับค่าความระลึกในระดับที่สูง ขณะที่ค่าความแม่นยำจะลดต่ำลงในทางกลับกันหากคำดรรชนีที่ใช้มีระดับความครอบคลุม

ที่น้อยลงหรือมีความเฉพาะเจาะจงมากขึ้นจะส่งผลให้ค่าความระลึกลดต่ำลงแต่จะให้ค่าความแม่นยำที่สูงขึ้น

2) การทดสอบประสิทธิภาพของระบบโดยใช้ชุดคำค้น 3 ลักษณะ ได้แก่ ชุดคำค้นขนาด 1-3 คำต่อชุด ชุดคำค้นขนาด 4-6 คำต่อชุด และประโยคคำค้นพบว่า โดยภาพรวมระบบมีประสิทธิภาพการทำงานในระดับปานกลางในทุกลักษณะคำค้น โดยทุกลักษณะคำค้นมีประสิทธิภาพในด้านความแม่นยำและค่าความระลึกใกล้เคียงกัน ยกเว้นการสืบค้นด้วยชุดคำค้น 1-3 คำค้นต่อชุดที่มีค่าความระลึกสูง แสดงให้เห็นว่าระบบซึ่งใช้แบบจำลองปริภูมิเวกเตอร์เป็นแบบจำลองในการสืบค้นสามารถดำเนินการได้อย่างมีประสิทธิภาพในระดับที่สูงเมื่อใช้งานร่วมกับการสืบค้นด้วยชุดคำค้นที่มีการระบุถึงคำสำคัญแต่ละคำอย่างชัดเจน ทั้งนี้เนื่องจากตัวแบบปริภูมิเวกเตอร์มีรูปแบบการจัดเก็บข้อมูลในลักษณะของรายการคำสำคัญและค่าน้ำหนักของคำสำคัญแต่ละคำ (Baeza-Yates and Ribeiro-Neto, 1999; Chu, 2003 และ ศุภชัย ตั้งวงศ์ศานต์, 2551) ส่วนประโยคคำค้นซึ่งเป็นลักษณะคำค้นที่มีความหมายแฝงของประโยคระบบอาจไม่สามารถทำได้อย่างมีประสิทธิภาพนัก

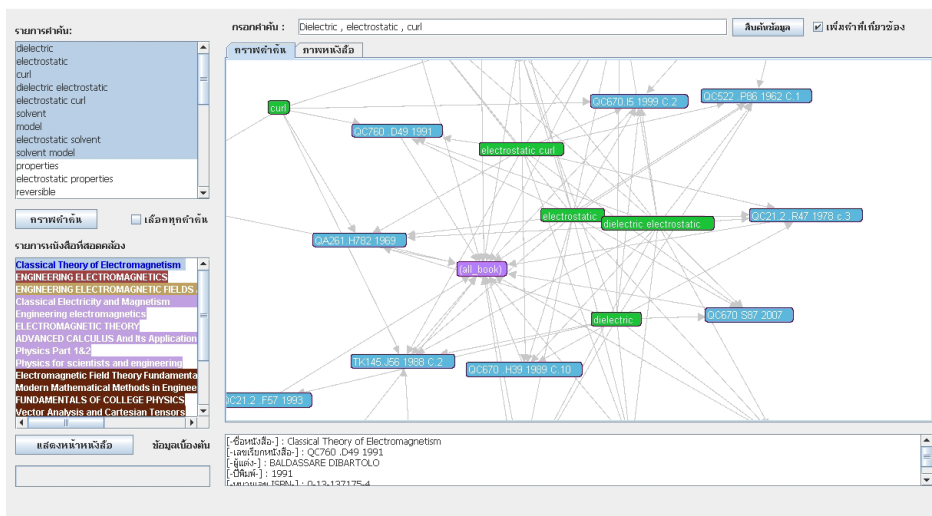
3) ประสิทธิภาพการให้บริการของระบบ เมื่อพิจารณาจากแบบประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบ ในด้านสมรรถนะการให้บริการของระบบ พบว่า ผู้ใช้งานโดยภาพรวมมีความพึงพอใจในระดับปานกลาง ทั้งนี้เนื่องจากระบบที่พัฒนาขึ้นอาศัยตัวแบบปริภูมิเวกเตอร์ในการสืบค้น เพื่อคัดเลือกสิ่งพิมพ์ที่มีความสอดคล้อง โดยดำเนินการในลักษณะของการสร้างเมทริกซ์ของคำดรรชนีของสิ่งพิมพ์ ซึ่งรูปแบบการคำนวณดังกล่าวจำเป็นต้องใช้หน่วยความจำที่สูงเพียงพอ อีกทั้งในกระบวนการทำงานของระบบมีการส่งผ่านข้อมูลขนาดใหญ่ระหว่างเครื่องแม่ข่ายกับเครื่องลูกข่ายผ่านทางเครือข่ายอินเทอร์เน็ต ซึ่งต้องใช้ระยะเวลาในการส่งผ่านข้อมูลและใช้หน่วยความจำที่สูงเช่นเดียวกันจึงส่งผลให้ระบบทดสอบตอบสนองการทำงานได้ช้า

ประสิทธิภาพด้านการนำทางความรู้

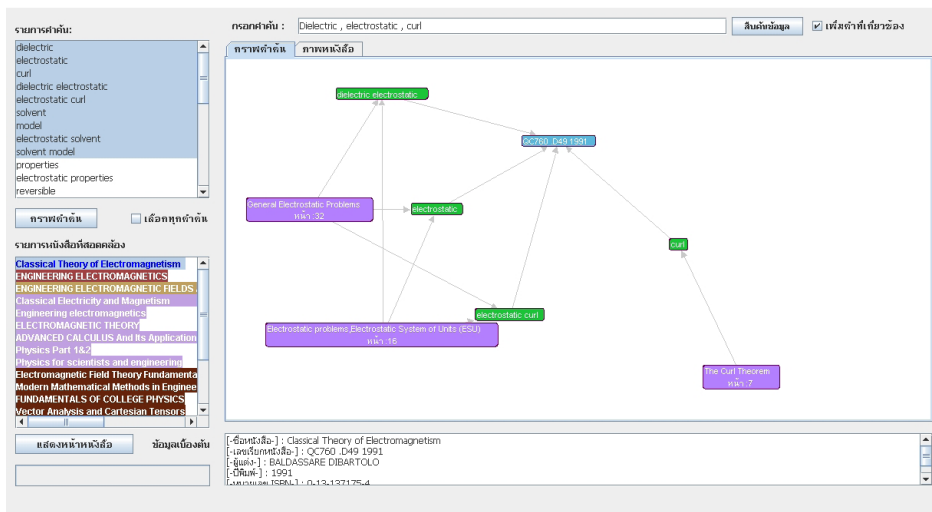
จากผลการประเมินความพึงพอใจของผู้ใช้งานที่มีต่อระบบในด้านประโยชน์ต่อการส่งเสริมการเรียนรู้พบว่า การนำทางความรู้โดยการนำเสนอแผนภาพ

ความเชื่อมโยงระหว่างคำค้นกับรายการสื่อสิ่งพิมพ์แต่ละเล่มจากรายการสารบัญและตรรกะที่ท้ายเล่ม โดยใช้โครงสร้างกราฟเอ็มแอลและนำเสนอรายการสื่อสิ่งพิมพ์ที่เกี่ยวข้องตามระดับความสอดคล้อง (ภาพที่ 3) ระบบสามารถทำงานได้อย่างมีประสิทธิภาพ สอดคล้องกับผลการประเมินความพึงพอใจของผู้ใช้งานในข้อที่ 7) ระบบช่วยผู้ใช้งานให้มองเห็นถึงความเชื่อมโยงของความรู้เรื่องที่ต้องการศึกษา ที่บันทึกไว้ในสื่อสิ่งพิมพ์แต่ละเล่มได้เป็นอย่างดี และข้อที่ 9) ระบบช่วยประหยัดเวลาในการ

พิจารณาคัดเลือกสื่อสิ่งพิมพ์ที่ต้องศึกษาได้เป็นอย่างดี โดยมีผลการประเมินความพึงพอใจในระดับมากที่สุด ค่าเฉลี่ย 4.00 สอดคล้องกับงานวิจัยของ Keller, Gerjets, Scheiter and Garsoffky (2004) และ Tergan and Keller (2005) ที่ระบุว่า การนำเสนอข้อมูลโดยใช้แผนภาพเป็นเครื่องมือช่วยในการแสดงข้อมูล โครงสร้างข้อมูลหรือแหล่งความรู้ที่มีขนาดใหญ่และซับซ้อน จะช่วยลดภาระในการศึกษาเรียนรู้ภายในระยะเวลาอันจำกัดได้เป็นอย่างดี



(3-ก) เชื่อมโยงระหว่างคำค้นกับรายการสื่อสิ่งพิมพ์



(3-ข) เชื่อมโยงระหว่างคำค้นกับคำสำคัญที่พบภายในสื่อสิ่งพิมพ์

ภาพที่ 3 การเชื่อมโยงคำค้นกับรายการสิ่งพิมพ์

เอกสารอ้างอิง

- รินฤดี ไชยวิชิตกุล. (2550). ความสามารถในการกำหนดคำค้นเพื่อการค้นหาสารสนเทศของนักศึกษาระดับปริญญาตรี สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. **วารสารรังสิตสารสนเทศ**, 13(1), 25-35.
- ศุภรินทร์ วิมุกตายน. (2548). ความสำเร็จและความล้มเหลวในการสืบค้นสารสนเทศจากฐานข้อมูลเครือข่ายสารสนเทศห้องสมุดในจุฬาลงกรณ์มหาวิทยาลัยของนิสิตปริญญาโท (ภาคนอกเวลาราชการ) ภาควิชารัฐประศาสนศาสตร์ คณะรัฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. **ข่าวสารห้องสมุดจุฬาลงกรณ์มหาวิทยาลัย**, 21(2).
- ศุภชัย ตั้งวงศ์สานต์. (2551). ระบบการจัดเก็บและการสืบค้นสารสนเทศด้วยคอมพิวเตอร์. กรุงเทพฯ : โรงพิมพ์พิทักษ์การพิมพ์.
- สุทธรม อูมาแสงทองกุล. (2541). การศึกษาผลการสืบค้นสารสนเทศที่ใช้เทคนิคตรรกะนี้แบบกำหนดค่าน้ำหนักที่จัดทำโดยมนุษย์. วิทยานิพนธ์ศิลปศาสตรมหาบัณฑิต สาขาวิชาบรรณารักษศาสตร์และสารนิเทศศาสตร์ มหาวิทยาลัยธรรมศาสตร์.
- สิทธิโชค ปัญญาฤกษ์ชัย, & ศิพानी นุชิตประสิทธิ์ชัย. (2552). ระบบการค้นคืนสารสนเทศโดยใช้เทคนิค N-gram. **Proceedings of The 5th National Conference on Computing and Information Technology**, 307-312.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999). **Modern Information Retrieval**. New York : Addison-Wesley.
- Barrón-Cedeño, A. and Rosso, P. (2009). On Automatic Plagiarism Detection Based on N-grams Comparison. **Lecture Notes In Computer Science**, 5478, 696-700.
- Beitzel, S. M., Jensen, E. C., Chowdhury, A., Grossman, D., and Frieder, O. (2004). Hourly Analysis of a Very Large Topically Categorized Web Query Log. **Proceeding of the 27th Annual International Conference on Research and Development in Information Retrieval**.
- Bollegala, D., Matsuo, Y. and Ishizuka, M. (2007). Measuring Semantic Similarity Between Words Using Web Search Engines. **Proceeding of the 16th International World Wide Web Conference**. 757-766.
- Borgman, C. L. (1996). Why are Online Catalogs Still Hard to Use?. **Journal of the American Society for Information Science**, 47(7), 493-503.
- Caviglia-Harris, J. L. and Hatley, J. (2004). Interdisciplinary Teaching Analyzing Consensus and Conflict in Environmental Studies. **International Journal of Sustainability in Higher Education**, 5(4), 395-403.
- Chu, H. (2003). **Information Representation and Retrieval in the Digital Age**. Medford, New Jersey : Information Today Incorporation.
- Choi, S., Cha, S. and Charles, C.T. (2010). A Survey of Binary Similarity and Distance Measures. **Journal of Systemics, Cybernetics and Informatics**, 8(1), 43-48.
- iProspect. (2006). **iProspect Search Engine User Behavior Study**. [On-line]. Available : http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf

- Jansen, B. J. and Spink, A. (2003). An Analysis of Web Information Seeking and Use : Documents Retrieved Versus Documents Viewed. Proceeding of the 4th International Conference on Internet Computing. ใน วิไลพร เลิศมหากะยริติ ภูริวัตร คัมภีร์ภาพพัฒน และ อนิราช มิ่งขวัญ (2551). รูปแบบการแสดงผล การค้นคืนของเครื่องมือการสืบค้นสารนิเทศบน อินเทอร์เน็ต. วารสารวิชาการพระจอมเกล้า พระนครเหนือ, 18(1), 89-98.
- Keller, T., Gerjets, P., Scheiter, K. and Grasoffky, B. (2004). Information Visualizations as Learning Tools. **Special Interest Meeting 2004 of EARLI SIG 6 and SIG 7**, Tuebingen, Germany. 269-280.
- Lancaster, F. W. (1968). Information Retrieval Systems: Characteristics, Testing, and Evaluation. New York: John Wiley. ใน สุธรรม อูมาแสงทองกุล (2541). การศึกษาผลการสืบค้นสารสนเทศที่ใช้เทคนิคดัชนีแบบกำหนดค่าน้ำหนัก ที่จัดทำโดยมนุษย์. วิทยานิพนธ์ศิลปศาสตร มหาบัณฑิต สาขาวิชาบรรณารักษศาสตร์และ สารนิเทศศาสตร์ มหาวิทยาลัยธรรมศาสตร์.
- Le-Khac, N., Aouad, L. M. and Kechadi, M. (2007). Knowledge Map: Toward a New Approach Supporting the Knowledge Management in Distributed Data Mining. **Proceeding of Third International Conference on Autonomic and Autonomous Systems (ICAS'07)**, 67-72.
- Lourenco, F., Lobo, V. and Bacao F. (2004). Binary-based Similarity Measures for Categorical Data and Their Application in Self-organizing Maps. **Proceeding JOCLAD 2004 - XI Jornadas de Classificacao e Anlise de Dados**, Lisbon.
- Macgregor, G. and McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organization and Resource Discovery Tool. **Library Review**, 55(5), 291-300.
- Maltby, A. (1975). Sayers' Manual of Classification for Librarians (5th ed). Andre Deutch : London. Quoted in Macgregor, G. and McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organization and Resource Discovery Tool. **Library Review**, 55(5), 291-300.
- Matsumura, N., Ohsawa, Y. and Ishizuka, M. (2001). Knowledge Navigation on Visualizing Complementary Documents. **Lecture Notes in Computer Science**, 2226, 258-270.
- Monev, V. (2004). Introduction to Similarity Searching in Chemistry. **Match-Communications in Mathematical and in Computer Chemistry**, 51, 7-38.
- Mustafa, S. and Al-Radaideh, Q. (2004). Using N-grams for Arabic Text Searching. **Journal of the american society for information science and technology**, 55(11), 1002-1007.
- Qiu, J., Yao, Y., Wang, Y. and Wang, X. (2006). Research of E-Government Knowledge Navigation System Based on XTM. **Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IATW'06)**.
- Salton, G., and Buckley, C. (1988). Term-weighting Approaches in Automatic Text Retrieval. **Information Processing & Management**, 24(5), 513-523.

- Salton, G. (1988). Automatic Text Transformations. Quoted in Salton, G. (1989). **Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer**. Addison-Wesley. 425–470.
- Sridhar, M. S. (2004). Subject Searching in the OPAC of a Special Library : Problems and Issues. **OCLC Systems & Services: International Digital Library Perspectives**, 20(4), 183-191.
- Tergan, S. O. and Keller, T. (2005). Visualizing Knowledge and Information : An Introduction. Knowledge and Information Visualization. **Lecture Notes in Computer Science**, 3426, 1-23.
- Yamamoto, H., Ohmi, S. and Tsuji, H. (2003). Entropy-based Indexing Term Selection For N-gram Text Search System. **IEEE International Conference on Systems, Man & Cybernetics (IEEE/SMC 2003)**.
- Zhang, B. and Sargur, N. S. (n.d.). **Properties of Binary Vector Dissimilarity Measures**. [On-line]. Available : http://www.cedar.buffalo.edu/papers/articles/CVPRIP03_propbina.pdf